

CERN 90-04  
24 April 1990

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

**CAS** CERN ACCELERATOR SCHOOL  
THIRD ADVANCED ACCELERATOR PHYSICS COURSE

Uppsala University, Sweden  
18–29 September 1989

PROCEEDINGS  
Editor: S. Turner

GENEVA  
1990

© Copyright CERN, Genève, 1990

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 registrable 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 92-9083-023-9

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

**CAS** CERN ACCELERATOR SCHOOL  
THIRD ADVANCED ACCELERATOR PHYSICS COURSE

Uppsala University, Sweden  
18-29 September 1989

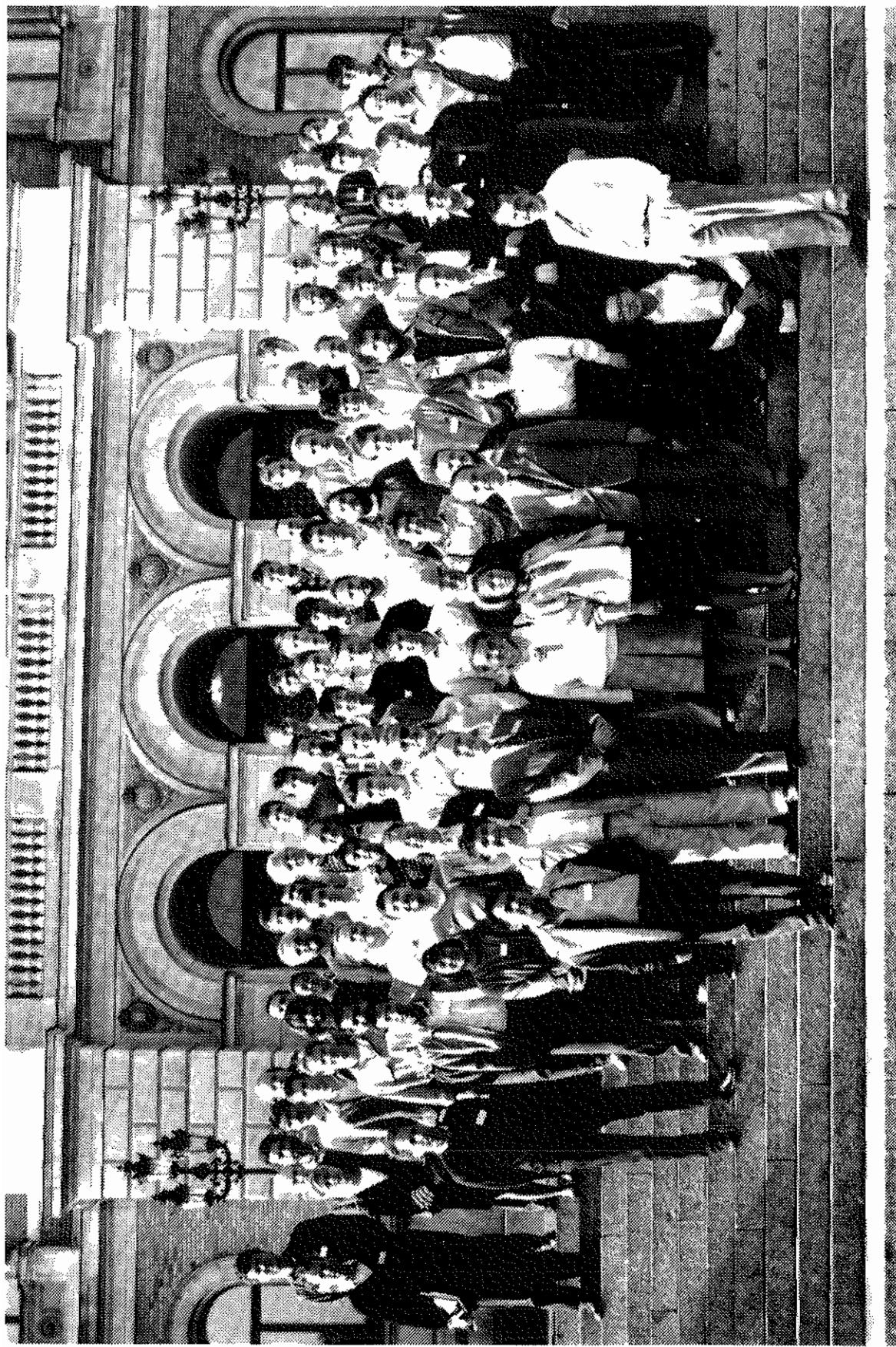
PROCEEDINGS  
Editor: S. Turner

GENEVA  
1990



## **ABSTRACT**

The third version of the CERN Accelerator School's (CAS) advanced course on General Accelerator Physics was given at Uppsala University from 18-29 September, 1989. Its syllabus was based on the previous courses held in Oxford, 1985 and Berlin, 1987 whose proceedings were published as CERN Yellow Reports 87-03 and 89-01 respectively. However, the opportunity was taken to emphasize the physics of small accelerators and storage rings, to present some topics in new ways, and to introduce new seminars. Thus the lectures contained in the present volume include chromaticity, dynamic aperture, kinetic theory, Landau damping, ion-trapping, Schottky noise, laser cooling and small ring lattice problems while the seminars include interpretation of numerical tracking, internal targets and living with radiation.

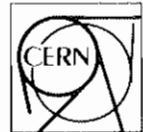




CERN ACCELERATOR SCHOOL

UPPSALA UNIVERSITY

will jointly organize a course on



# ADVANCED ACCELERATOR PHYSICS

18-29 SEPTEMBER 1989  
at Hotel Linné, Uppsala, Sweden

## Lectures:

Hamiltonian mechanics  
Liouville theorem  
Non-linear resonances  
Chromaticity  
Dynamic aperture  
Kinetic theory  
Landau damping  
Lattice and lens design  
Longitudinal motion  
Bunched-beam coherent instabilities  
Wake fields  
Introduction to polarisation  
Intra-beam scattering  
Interactions with residual gas  
Ion trapping and clearing  
Beam transfer function  
Schottky noise on very cold beams

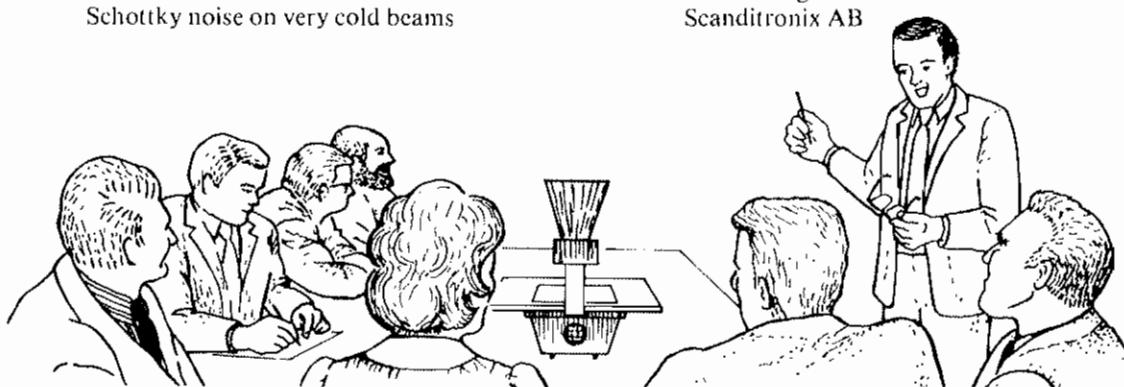
Stochastic cooling  
Electron cooling  
Laser cooling

## Seminars

Living with radiation  
Radio-therapy  
Compact synchrotron-light sources  
Interpretation of numerical tracking  
Internal targets  
Intensity limits in storage rings  
Crystalline ion beams

## Visits

The Svedberg Laboratory  
Manne Siegbahn Institute  
Scanditronix AB



**General information** Students should have a basic knowledge of accelerators. The course is designed to highlight the problems of small rings where appropriate. Persons wishing to attend this school can obtain further information and application forms from the CERN Accelerator School, Mrs. S. von Wartburg, LEP Division, CH 1211 Geneva 23. Application forms must be returned no later than 1 June 1989. The registration fee is SEK 8000 per person if sharing a double room and includes full board and lodging. For rates for single rooms, local students, and for non-participating accompanying persons (space permitting), please see the application form. All participants will receive a copy of the proceedings of our previous and present advanced course. The number of participants is limited to 120.

**Administration** Head of school: P.J. Bryant; Programme Committee: N. Angert, P.J. Bryant, J. Buon, C.J. Herrlander, D. Mohl, K.O. Nielsen, D. Reistad, S. Tazzari; Local Organizing Committee: C. Asbjörk, I. Ericson, D. Reistad, S. Turner, S. von Wartburg.

**Sponsored by:**

**SCANDITRONIX**  
S-75590 UPPSALA

**BALZERS Ltd.**

High vacuum and thin-film technology  
FL-9496 Balzers (Principality of Lichtenstein)

PROGRAMME FOR "ADVANCED ACCELERATOR PHYSICS COURSE", UPPSALA 18-29 SEPTEMBER 1989

## FOREWORD

The CERN 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 and Workshops 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 shown in the table below.

## 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 Present volume	-	Synchrotron radiation and free-electron lasers CERN 90-03 (1990)
1990	Basic (to be published)	Frontiers of particle beams (to be published)	Power converters for particle accelerators (to be published)
1991	Advanced (to be published)	-	RF engineering aspects of particle accelerators (to be published)

The present proceedings are concerned with the third of the advanced courses on general accelerator physics held in Uppsala in September, 1989. This course closely follows the earlier ones at Oxford and Berlin, but with some of the topics improved, extended or presented in a different way. In addition, some completely new subjects were added, especially in the case of the seminars. In this way the chapters included in this present volume are complementary to those published for the first and second advanced courses.

Organization of this school and the publication of its proceedings would not have been possible without the help of many people. The support of the CERN Management and the School's Advisory, Programme and Local Organizing Committees is gratefully acknowledged. For this particular course, the financial and organizational support given by the University of Uppsala and the The Svedberg Laboratory and their staff, especially D. Reistad, A. Johansson, S. Kullander and C. Asbjörk, was most important. The help and enthusiasm of the Manne Siegbahn Institute, particularly C.J. Herrlander, the Statens Vattenfallsverk Forsmarksverket, and the sponsorship of Scanditronix and Balzers Ltd played an important part in ensuring the success of this course. Very special thanks, however, must go to the lecturers and authors who gave so freely of their time and effort to present the topics at the school and in these proceedings. The help of the many CERN service sections who contributed in one way or another to the organization of the school and the production of this present volume was most invaluable. Finally, the attendance, enthusiasm and friendliness of the participants at the school was the best reward for the effort made on their behalf.

S. Turner, Editor

## CONTENTS

	Page no.
<b>Foreword</b>	<b>vii</b>
<b>B. Autin</b>	
<b>Non-linear betatron oscillations</b>	1
Introduction	1
Linear theory	3
Scalar potential of the magnetic field	4
Equations of motion in the action-angle variables	5
Action-angle perturbations in a periodic lattice	9
Compensation schemes for amplitude dependent distortions	16
Conclusion	22
<b>A. Ropert</b>	
<b>Dynamic aperture</b>	26
Introduction	26
Dynamic aperture	27
Analytical approach of the dynamic aperture	30
Compensation procedures	41
Numerical approach	44
Experimental results	49
<b>F. Ruggiero</b>	
<b>Kinetic theory of charged particle beams</b>	52
Introduction	52
Liouville equation	54
The BBGKY hierarchy	55
Vlasov equation	56
Debye screening in a charged plasma: an example	57
Cluster expansion	63
<b>R.D. Kohaupt</b>	
<b>Landau damping</b>	67
Fundamental considerations	67
The framework of stable and unstable loops	67
<b>A. Poncelet</b>	
<b>Ion trapping and clearing</b>	74
Introduction	74
Ion production	75
Ion dynamics (DC beam)	81
Ion dynamics (bunched beams)	86
Clearing means	89
Resonant transverse shaking of the beam	93
Conclusion	96
<b>S. Baird</b>	
<b>Schottky noise from very cold beams</b>	99
Introduction	99
How do we define "hot" and "cold" beams?	99
Longitudinal density fluctuations (a simple picture)	101
Longitudinal stability and coherent motion in a cold beam	105
Comparison with measured spectra	110

<b>E. Bonderup</b>	
<b>Laser cooling</b>	114
Introduction	114
Origin of forces	115
Atom in radiation field	119
Cooling in standing wave	124
Summary	132
<b>E.J.N. Wilson</b>	
<b>Small ring lattice problems</b>	136
Introduction	136
The example - An antiproton accumulator	137
Magnet design, measurement and correction	142
Multipoles in the magnet design	143
Two-dimensional field integrals	148
Curvilinear coordinates	150
Excursions within the end field	151
Conclusions	154
<b>F. Willeke</b>	
<b>Analysis of particle tracking data</b>	156
Introduction	156
Tools for particle tracking	157
Some suggestions how to perform particle tracking	161
Interpretation of tracking data	161
Closing remarks	180
<b>C. Ekström</b>	
<b>Internal targets</b>	184
Introduction	184
Physics experiments	184
Comparison between conventional and internal-target experiments	186
Requirements on internal targets	187
Choice of internal targets	187
Gas-jet targets	188
Polarized targets	192
Solid targets	196
Conclusion	199
<b>S. Kullander</b>	
<b>Living with radiation</b>	203
Introduction	203
Radiation and photography	204
Radiation and the atmosphere	207
Radiation serving medicine	211
Radiation emitted by particles accelerated in the sky	212
Stimulated cyclotron radiation	213
Conclusion	215
List of Participants	217

CERN 90-04  
24 April 1990

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

**CAS** CERN ACCELERATOR SCHOOL  
THIRD ADVANCED ACCELERATOR PHYSICS COURSE

Uppsala University, Sweden  
18–29 September 1989

PROCEEDINGS  
Editor: S. Turner

GENEVA  
1990

© Copyright CERN, Genève, 1990

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 registrable 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 92-9083-023-9

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

**CAS** CERN ACCELERATOR SCHOOL  
THIRD ADVANCED ACCELERATOR PHYSICS COURSE

Uppsala University, Sweden  
18-29 September 1989

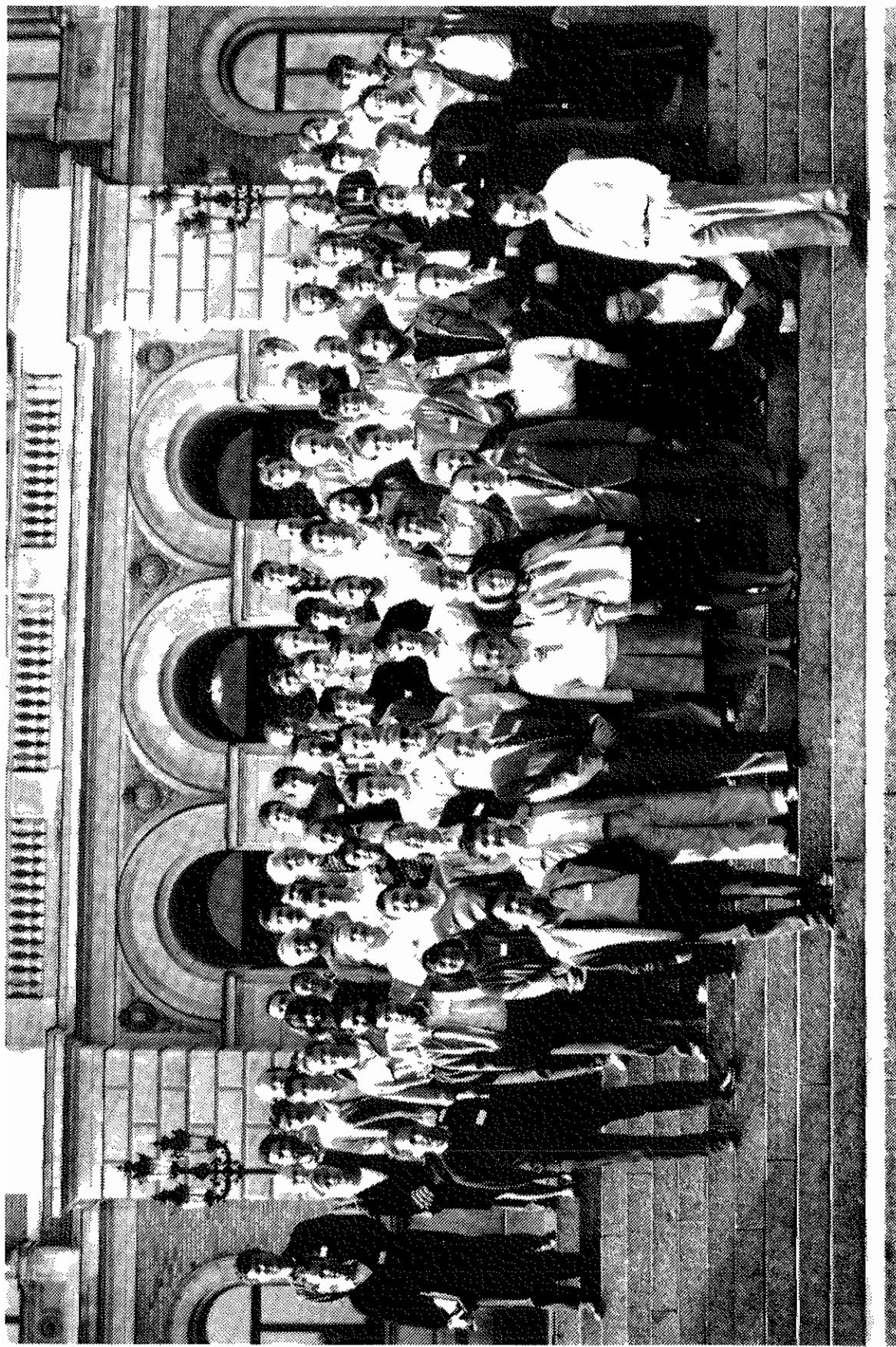
PROCEEDINGS  
Editor: S. Turner

GENEVA  
1990



## **ABSTRACT**

The third version of the CERN Accelerator School's (CAS) advanced course on General Accelerator Physics was given at Uppsala University from 18-29 September, 1989. Its syllabus was based on the previous courses held in Oxford, 1985 and Berlin, 1987 whose proceedings were published as CERN Yellow Reports 87-03 and 89-01 respectively. However, the opportunity was taken to emphasize the physics of small accelerators and storage rings, to present some topics in new ways, and to introduce new seminars. Thus the lectures contained in the present volume include chromaticity, dynamic aperture, kinetic theory, Landau damping, ion-trapping, Schottky noise, laser cooling and small ring lattice problems while the seminars include interpretation of numerical tracking, internal targets and living with radiation.

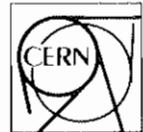




CERN ACCELERATOR SCHOOL

UPPSALA UNIVERSITY

will jointly organize a course on



# ADVANCED ACCELERATOR PHYSICS

18-29 SEPTEMBER 1989  
at Hotel Linné, Uppsala, Sweden

## Lectures:

Hamiltonian mechanics  
Liouville theorem  
Non-linear resonances  
Chromaticity  
Dynamic aperture  
Kinetic theory  
Landau damping  
Lattice and lens design  
Longitudinal motion  
Bunched-beam coherent instabilities  
Wake fields  
Introduction to polarisation  
Intra-beam scattering  
Interactions with residual gas  
Ion trapping and clearing  
Beam transfer function  
Schottky noise on very cold beams

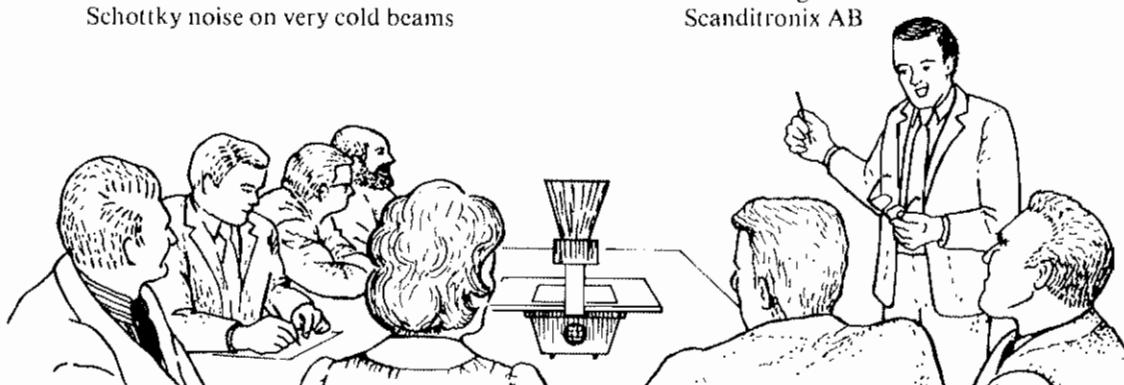
Stochastic cooling  
Electron cooling  
Laser cooling

## Seminars

Living with radiation  
Radio-therapy  
Compact synchrotron-light sources  
Interpretation of numerical tracking  
Internal targets  
Intensity limits in storage rings  
Crystalline ion beams

## Visits

The Svedberg Laboratory  
Manne Siegbahn Institute  
Scanditronix AB



**General information** Students should have a basic knowledge of accelerators. The course is designed to highlight the problems of small rings where appropriate. Persons wishing to attend this school can obtain further information and application forms from the CERN Accelerator School, Mrs. S. von Wartburg, LEP Division, CH 1211 Geneva 23. Application forms must be returned no later than 1 June 1989. The registration fee is SEK 8000 per person if sharing a double room and includes full board and lodging. For rates for single rooms, local students, and for non-participating accompanying persons (space permitting), please see the application form. All participants will receive a copy of the proceedings of our previous and present advanced course. The number of participants is limited to 120.

**Administration** Head of school: P.J. Bryant; Programme Committee: N. Angert, P.J. Bryant, J. Buon, C.J. Herrlander, D. Mohl, K.O. Nielsen, D. Reistad, S. Tazzari; Local Organizing Committee: C. Asbjörk, I. Ericson, D. Reistad, S. Turner, S. von Wartburg.

**Sponsored by:**

**SCANDITRONIX**  
S-75590 UPPSALA

**BALZERS Ltd.**

High vacuum and thin-film technology  
FL-9496 Balzers (Principality of Lichtenstein)

PROGRAMME FOR "ADVANCED ACCELERATOR PHYSICS COURSE", UPPSALA 18-29 SEPTEMBER 1989

## FOREWORD

The CERN 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 and Workshops 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 shown in the table below.

## 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 Present volume	-	Synchrotron radiation and free-electron lasers CERN 90-03 (1990)
1990	Basic (to be published)	Frontiers of particle beams (to be published)	Power converters for particle accelerators (to be published)
1991	Advanced (to be published)	-	RF engineering aspects of particle accelerators (to be published)

The present proceedings are concerned with the third of the advanced courses on general accelerator physics held in Uppsala in September, 1989. This course closely follows the earlier ones at Oxford and Berlin, but with some of the topics improved, extended or presented in a different way. In addition, some completely new subjects were added, especially in the case of the seminars. In this way the chapters included in this present volume are complementary to those published for the first and second advanced courses.

Organization of this school and the publication of its proceedings would not have been possible without the help of many people. The support of the CERN Management and the School's Advisory, Programme and Local Organizing Committees is gratefully acknowledged. For this particular course, the financial and organizational support given by the University of Uppsala and the The Svedberg Laboratory and their staff, especially D. Reistad, A. Johansson, S. Kullander and C. Asbjörk, was most important. The help and enthusiasm of the Manne Siegbahn Institute, particularly C.J. Herrlander, the Statens Vattenfallsverk Forsmarksverket, and the sponsorship of Scanditronix and Balzers Ltd played an important part in ensuring the success of this course. Very special thanks, however, must go to the lecturers and authors who gave so freely of their time and effort to present the topics at the school and in these proceedings. The help of the many CERN service sections who contributed in one way or another to the organization of the school and the production of this present volume was most invaluable. Finally, the attendance, enthusiasm and friendliness of the participants at the school was the best reward for the effort made on their behalf.

S. Turner, Editor

## CONTENTS

	Page no.
Foreword	vii
<b>B. Autin</b>	
Non-linear betatron oscillations	1
Introduction	1
Linear theory	3
Scalar potential of the magnetic field	4
Equations of motion in the action-angle variables	5
Action-angle perturbations in a periodic lattice	9
Compensation schemes for amplitude dependent distortions	16
Conclusion	22
<b>A. Ropert</b>	
Dynamic aperture	26
Introduction	26
Dynamic aperture	27
Analytical approach of the dynamic aperture	30
Compensation procedures	41
Numerical approach	44
Experimental results	49
<b>F. Ruggiero</b>	
Kinetic theory of charged particle beams	52
Introduction	52
Liouville equation	54
The BBGKY hierarchy	55
Vlasov equation	56
Debye screening in a charged plasma: an example	57
Cluster expansion	63
<b>R.D. Kohaupt</b>	
Landau damping	67
Fundamental considerations	67
The framework of stable and unstable loops	67
<b>A. Poncelet</b>	
Ion trapping and clearing	74
Introduction	74
Ion production	75
Ion dynamics (DC beam)	81
Ion dynamics (bunched beams)	86
Clearing means	89
Resonant transverse shaking of the beam	93
Conclusion	96
<b>S. Baird</b>	
Schottky noise from very cold beams	99
Introduction	99
How do we define "hot" and "cold" beams?	99
Longitudinal density fluctuations (a simple picture)	101
Longitudinal stability and coherent motion in a cold beam	105
Comparison with measured spectra	110

<b>E. Bonderup</b>	
<b>Laser cooling</b>	114
Introduction	114
Origin of forces	115
Atom in radiation field	119
Cooling in standing wave	124
Summary	132
<b>E.J.N. Wilson</b>	
<b>Small ring lattice problems</b>	136
Introduction	136
The example - An antiproton accumulator	137
Magnet design, measurement and correction	142
Multipoles in the magnet design	143
Two-dimensional field integrals	148
Curvilinear coordinates	150
Excursions within the end field	151
Conclusions	154
<b>F. Willeke</b>	
<b>Analysis of particle tracking data</b>	156
Introduction	156
Tools for particle tracking	157
Some suggestions how to perform particle tracking	161
Interpretation of tracking data	161
Closing remarks	180
<b>C. Ekström</b>	
<b>Internal targets</b>	184
Introduction	184
Physics experiments	184
Comparison between conventional and internal-target experiments	186
Requirements on internal targets	187
Choice of internal targets	187
Gas-jet targets	188
Polarized targets	192
Solid targets	196
Conclusion	199
<b>S. Kullander</b>	
<b>Living with radiation</b>	203
Introduction	203
Radiation and photography	204
Radiation and the atmosphere	207
Radiation serving medicine	211
Radiation emitted by particles accelerated in the sky	212
Stimulated cyclotron radiation	213
Conclusion	215
List of Participants	217

## NON-LINEAR BETATRON OSCILLATIONS

Bruno Autin

CERN, Geneva, Switzerland

### ABSTRACT

The equations of motion of a particle oscillating in a non-linear magnetic field are solved using a time-dependent perturbation theory extended to the second order. The theory is applied to the design of correction fields which limit the amplitude growth of the oscillation. The correction schemes are verified by numerical integration of the equations of motion ("particle tracking") and discussed on two special machines: the Berkeley Advanced Light Source and the CERN Antiproton Collector for which experimental data have been collected.

### 1. INTRODUCTION

The optical properties of an accelerator or of a storage ring are usually defined in a hierarchical way. There is first the determination of the ideal trajectory which passes through the center of the quadrupoles where there is no magnetic field and which closes onto itself after one turn, this is the *central orbit* which is a consequence of the bending field distribution around the ring and gives its general shape to the machine: a circle, an oval, a racetrack, etc. Then come the focusing properties about the central orbit; they concern the oscillatory motion of the particles in a monochromatic beam whose energy corresponds to the central orbit and determine the number of betatron oscillations per turn, the *tune*, and the beam envelope via the  *$\beta$  function*. In a real beam, the particles have a certain energy distribution; the orbits and the focusing or *chromatic* properties of the *off-momentum* particles such as they result from the dipole and quadrupole fields are usually not acceptable; they have to be controlled using non-linear fields and especially *sextupolar* fields which vary quadratically with the position of the particle with respect to the central orbit. A secondary effect of the non-linear fields is the alteration of the motion for the particles with a large oscillation amplitude. It reduces the number of particles which can be injected or stored in the machine. It is especially severe for the superconducting machines, the large acceptance storage rings and the new generation of synchrotron light sources. A traditional cure has consisted of analyzing the motion into its Fourier components and suppressing the most harmful component yet keeping unchanged the chromatic properties (*resonance* method) [1]. However, modern machines are built in such a way that the *operation point*, the couple of the horizontal and vertical tunes, is far from any known dangerous resonance. The analysis and the compensation of the non-linear distortions must then be treated in the frame of a more general theory.

We use here a first and second-order theory of the general equations of motion. The order 0 is the linear theory (section 2) in which we introduce the main notations and we recall that the phase space particle trajectories are circles for the horizontal and vertical motions when the phase space is normalized. One transforms then the

cartesian coordinates ( $x, y, p_x, p_y$ ) into the polar coordinates ( $J_x, J_y, \mu_x, \mu_y$ ) where  $J$  and  $\mu$  are the action and the angle variables respectively. The reason for this change of coordinates lies in the fact that the integrated value of  $\mu/2\pi$  over one turn is the betatron tune and that the action angle is directly related to the beam size. The magnetic field (section 3) is described by a scalar potential which is directly analyzed in terms of the action-angle variables. The first-order differential system of the four coupled equations of motion (section 4) is derived from the Hamiltonian of the motion. This system is solved by successive approximations using the method of *variation of the constants* [2]. The constants are the initial values of the action and angle variables. At the first order, action and angle are replaced in the general system by their linear expressions and the integration is straightforward. At the second order, the first-order expressions of  $J$  and  $\mu$  are injected into the differential system; the result is written in the form of double integrals which express the fact that the total effect of the non-linear fields is no longer a superposition of the individual fields as is the case in the first-order theory but that the non-linear fields are now correlated between them.

For closed machines (section 5), the condition of periodicity of the magnetic structure must be manifest. As this case is of paramount importance in practice, the integrals are written explicitly using the simplest models of integration. At the first order (section 5.1), the *thin-lens* approximation which consists of concentrating the field into a magnet of zero length using  $\delta$ -functions is very convenient. At the second order (section 5.2), the thin-lens approximation cannot be used and the simplest model consists of assuming a constant value for the  $\beta$  function and a linear variation of the betatron phase with respect to the longitudinal coordinate inside a magnet. The various types of correlation are analyzed in order to give as much insight as possible into the physical problem. They apply to the coupling of a field with itself (section 5.2.1), to the correlation between a given field and the same field at the next turn (section 5.2.2) and to the correlation between two different fields (section 5.2.3). A technical difficulty appears in the evaluation of each type of correlation in the form of *secular* terms, those terms which repeat themselves identically at each turn and lead to an unphysical growth of the oscillation; fortunately, when all the correlations are summed up, these terms cancel out. In the formalism which is developed here, the motion is completely described as the superposition of a periodic motion and of an oscillation about the periodic motion; this feature contrasts with other treatments [3] which are deliberately restricted to the periodic solution of the equations of motion. Our purpose being to design a machine with the maximum aperture, there is no reason to neglect the oscillatory part of the motion. In order to test the validity of the theory, one can compare scatter plots of a section of the phase space based on the analytical formulae with purely numerical non-perturbative techniques. However, the main interest of analytical formulae lies in the prescription of correction schemes (section 6).

The compensation method applies to the correction of the systematic non-linear distortions created by the chromaticity correction elements. It is illustrated in the case of a hadron machine, the CERN Antiproton Collector [4], and of a lepton machine, the Berkeley Advanced Light Source [5] and it is shown that a substantial improvement in machine aperture can be obtained at the cost of a rather modest correction scheme implemented in regions where the orbit position is independent of the particle momentum.

## 2. LINEAR THEORY

In a particle accelerator, the shape of the beam is determined by the focusing system [6] which is composed of linear and non-linear magnetic fields. The linear elements are the *quadrupoles* whose field components are linear functions of the transverse coordinates ( $x, y$ ). In these fields, the particles are submitted to restoring forces proportional to the distance of the particle from the magnetic axis where the field is zero and the transverse coordinates have their origin. The equations of the transverse motion of a particle are then:

$$\begin{aligned} \frac{d^2x}{ds^2} + K_x(s)x &= 0 \\ \frac{d^2y}{ds^2} + K_y(s)y &= 0 \end{aligned} \quad (1)$$

The curvilinear coordinate  $s$  is taken along the reference trajectory which passes through the centre of the quadrupoles. In contrast with the harmonic oscillator equation, the focusing strength  $K$  is not constant but a function of  $s$ . One can still nevertheless write the solution to the differential equations in a phase-amplitude form [1]:

$$\begin{aligned} x &= \sqrt{2\beta_x(s) J_x} \cos[\mu_x(s) + \phi_x] \\ y &= \sqrt{2\beta_y(s) J_y} \cos[\mu_y(s) + \phi_y] \end{aligned} \quad (2)$$

By back-substitution of  $x$  (or  $y$ ) into the differential equations and identification with respect to the sine and cosine terms, it turns out that  $\beta$  and  $\mu$  are related by

$$\mu(s) = \int_0^s \frac{d\sigma}{\beta(\sigma)} \quad (3)$$

and that the derivative of  $\beta$  which is traditionally defined via the coefficient

$$\alpha = -\frac{1}{2} \frac{d\beta}{ds} \quad (4)$$

satisfies the equation

$$\frac{d\alpha}{ds} = K\beta \cdot \frac{1+\alpha^2}{\beta} \quad (5)$$

For brevity, the subscripts are omitted when the relations between physical parameters like  $\beta, \mu, J, \phi, \dots$  or the statements about them are valid for the horizontal and the vertical plane as well. The four constants of the motion are the actions  $J$  and the angles  $\phi$ . The Hamiltonian of the motion which has the form

$$H = \frac{1}{2} [ p_x^2 + p_y^2 + K_x x^2 + K_y y^2 ] \quad (6)$$

with  $(x, y)$  and  $(p_x = dx/ds, p_y = dy/ds)$  as conjugate variables becomes

$$H_1 = \frac{J_x}{\beta_x} + \frac{J_y}{\beta_y} \quad (7)$$

with  $(\mu_x + \phi_x, \mu_y + \phi_y)$  and  $(J_x, J_y)$  as new position and momentum variables. As our purpose is essentially to determine the evolution of the beam envelope in the presence of non-linear fields, the equations of motion will be derived from a Hamiltonian of the same type as  $H_1$ . The non-linear Hamiltonian is deduced from  $H_1$  by addition of a potential term  $V$  which has to be written in the angle-action variables.

### 3. SCALAR POTENTIAL OF THE MAGNETIC FIELD

The scalar potential of a magnetic multipole of order  $m$  is

$$V = \frac{g^{(m-2)}}{m!} Re(x + iy)^m \quad (8)$$

where  $g^{(m-2)}$  is the  $(m-2)$ - derivative of the vertical component of the magnetic field gradient with respect to  $x$ . The order  $m$  may be defined from 1,  $g^{(-1)}$  is then the *dipole* component and  $g^{(0)}$  the quadrupole component, but we shall only consider fields corresponding to values of  $m$  larger than 2:  $g^{(1)}, g^{(2)}, \dots$  are then the *sextupolar*, *octupolar*, ... components. In particle dynamics, the potential comes in the Hamiltonian of the motion in the form

$$V = \frac{e}{p} v \quad (9)$$

where  $e$  is the charge and  $p$  the momentum of the particle. It is therefore convenient to use the derivatives of the focusing strength

$$K^{(m-2)} = \frac{e}{p} g^{(m-2)} \quad (10)$$

By substituting the expressions of the betatron motion into (8), the expression of the potential  $V$  becomes

$$V = \sum_{i=0}^{I(m/2)} \sum_{j=0}^{I(m/2)-i} \sum_{k=0}^i C_{ijk} [\cos \mu_{ijk}^+ + \cos \mu_{ijk}^-] \quad (11)$$

where  $I$  is the integer part function and

$$C_{ijk} = K^{(m-2)} \frac{(J_x \beta_x)^{\frac{m}{2}-i} (-J_y \beta_y)^i}{(2)_2^{\frac{m}{2}-1} j! k! (m-2i-j)! (2i-k)!} \quad (12)$$

$$\mu_{ijk}^\pm = \overline{m_{ijk}^\pm} \cdot \bar{\mu} \quad (13)$$

with

$$\bar{\mu} = (\mu_x + \phi_x, \mu_y + \phi_y) \quad (14)$$

$$\bar{m}_{ijk}^{\pm} = (m_{xij}, \pm m_{yik}) \quad (15)$$

$$m_{xij} = m - 2(i+j) \quad (16)$$

$$m_{yik} = 2(i-k) \quad (17)$$

Special cases occur when either  $i$  or  $m-2i$  are zero, then  $m_y$  or  $m_x$  are zero and the two cosines are the same in the expression of  $V$  which must be re-written

$$V = \sum_{i=0}^{I(m/2)} \sum_{j=0}^{I(m/2)-i} \sum_{k=0}^i C_{ijk} \cos \mu_{ijk}^+ \quad (18)$$

These cases correspond to the combinations

$$\begin{aligned} (i, j, k) &= (0, j < \frac{m}{2}, 0) \\ (i, j, k) &= (\frac{m}{2}, 0, k < \frac{m}{2}) \end{aligned} \quad (19)$$

When  $m$  is even,  $m_x$  and  $m_y$  can be zero simultaneously, a constant term then appears at the right hand side of  $V$  which is always defined within an additive constant anyhow; the combinations of the type

$$(i, j, k) = (i, \frac{m}{2} - i, i) \quad (20)$$

have thus simply to be eliminated.

#### 4. EQUATIONS OF MOTION IN THE ACTION-ANGLE VARIABLES

When a non-linear field is introduced,  $J$  and  $\phi$  become functions of  $s$  and the equations of motion are solved using a method of *variation of constants* also called sometimes *time dependent perturbation* method. The non-linear Hamiltonian is

$$H_1 = \frac{J_x}{\beta_x} + \frac{J_y}{\beta_y} + V \quad (21)$$

The equations of motion are then

$$\begin{aligned}\frac{dJ}{ds} &= -\frac{\partial H_1}{\partial(\mu+\phi)} = -\frac{\partial V}{\partial(\mu+\phi)} \\ \frac{d(\mu+\phi)}{ds} &= \frac{\partial H_1}{\partial J} = \frac{1}{\beta} + \frac{\partial V}{\partial J}\end{aligned}\tag{22}$$

Because of the relation (3) between  $\mu$  and  $\beta$ , the second equation can be simplified and the system re-written

$$\begin{aligned}\frac{dJ}{ds} &= -\frac{\partial V}{\partial \phi} \\ \frac{d\phi}{ds} &= \frac{\partial V}{\partial J}\end{aligned}\tag{23}$$

or, more explicitly,

$$\begin{aligned}\frac{dJ_x}{ds} &= \sum_{i=0}^{I_2^{(m)}} \sum_{j=0}^{I_2^{(m)}} \sum_{k=0}^i C_{ijk} m_{xij} [\sin \mu_{ijk}^+ + \sin \mu_{ijk}^-] \\ \frac{dJ_y}{ds} &= \sum_{i=0}^{I_2^{(m)}} \sum_{j=0}^{I_2^{(m)}} \sum_{k=0}^i C_{ijk} m_{yik} [\sin \mu_{ijk}^+ - \sin \mu_{ijk}^-]\end{aligned}\tag{24}$$

$$\begin{aligned}\frac{d\phi_x}{ds} &= \sum_{i=0}^{I_2^{(m)}} \sum_{j=0}^{I_2^{(m)}} \sum_{k=0}^i \frac{\partial C_{ijk}}{\partial J_x} [\cos \mu_{ijk}^+ + \cos \mu_{ijk}^-] \\ \frac{d\phi_y}{ds} &= \sum_{i=0}^{I_2^{(m)}} \sum_{j=0}^{I_2^{(m)}} \sum_{k=0}^i \frac{\partial C_{ijk}}{\partial J_y} [\cos \mu_{ijk}^+ + \cos \mu_{ijk}^-]\end{aligned}\tag{25}$$

This set of equations can be solved by successive approximations.

#### 4.1 First iteration

The first-order perturbation of the angle-action variables  $(J, \phi)$  is obtained by assuming  $(J, \phi)$  to be constant and equal to  $(J_0, \phi_0)$  at the right hand side of the equations. In order to have an explicit form which can be generalized at the second iteration, let us define the vectors

$$\begin{aligned}\overline{A_x^{(1)}} &= (C_{i,j,k}, C_{i,j,k}) \\ \overline{A_y^{(1)}} &= (C_{i,j,k}, -C_{i,j,k})\end{aligned}\tag{26}$$

$$\begin{aligned}\overline{\mathbf{F}_x^{(1)}} &= \left( \frac{\partial C_{ijk}}{\partial J_x}, \frac{\partial C_{ijk}}{\partial J_y} \right) \\ \overline{\mathbf{F}_y^{(1)}} &= \left( \frac{\partial C_{ijk}}{\partial J_y}, \frac{\partial C_{ijk}}{\partial J_x} \right)\end{aligned}\tag{27}$$

$$\begin{aligned}\overline{\Psi^{(1)}} &= (\sin \mu_{ijk}^+, \sin \mu_{ijk}^-) \\ \overline{\Phi^{(1)}} &= (\cos \mu_{ijk}^+, \cos \mu_{ijk}^-)\end{aligned}\tag{28}$$

The set of equations integrated over a length  $s$  takes the form

$$\begin{aligned}J_{x1} &= \sum_{i=0}^{I(m/2)} \sum_{j=0}^{I(m/2)-i} \sum_{k=0}^i m_{xij} \int_0^s \overline{A_x^{(1)}} \cdot \overline{\Psi^{(1)}} d\sigma \\ J_{y1} &= \sum_{i=0}^{I(m/2)} \sum_{j=0}^{I(m/2)-i} \sum_{k=0}^i m_{yik} \int_0^s \overline{A_y^{(1)}} \cdot \overline{\Psi^{(1)}} d\sigma\end{aligned}\tag{29}$$

$$\begin{aligned}\phi_{x1} &= \sum_{i=0}^{I(m/2)} \sum_{j=0}^{I(m/2)-i} \sum_{k=0}^i \int_0^s \overline{F_x^{(1)}} \cdot \overline{\Phi^{(1)}} d\sigma \\ \phi_{y1} &= \sum_{i=0}^{I(m/2)} \sum_{j=0}^{I(m/2)-i} \sum_{k=0}^i \int_0^s \overline{F_y^{(1)}} \cdot \overline{\Phi^{(1)}} d\sigma\end{aligned}\tag{30}$$

#### 4.2 Second iteration

At the second order, the coefficients  $C$  and the trigonometric functions that we call  $T$  for brevity have to be expanded as functions of  $J$  and  $\phi$  respectively:

$$\begin{aligned}C_{ijk} &= C_{0ijk} + \frac{\partial C_{0ijk}}{\partial J_x} J_{x1} + \frac{\partial C_{0ijk}}{\partial J_y} J_{y1} \\ T_{ijk} &= T_{0ijk} + \frac{\partial T_{0ijk}}{\partial \phi_x} \phi_{x1} + \frac{\partial T_{0ijk}}{\partial \phi_y} \phi_{y1}\end{aligned}\tag{31}$$

The differential equations for the second-order terms of the action variables are thus

$$\begin{aligned}\frac{dJ_{x2}}{ds} &= \sum_{i=0}^{I(m/2)} \sum_{j=0}^{I(m/2)-i} \sum_{k=0}^i m_{xij} \left[ \left( \frac{\partial C_{0ijk}}{\partial J_x} J_{x1} + \frac{\partial C_{0ijk}}{\partial J_y} J_{y1} \right) T_{x0ijk} + \left( \frac{\partial T_{x0ijk}}{\partial \phi_x} \phi_{x1} + \frac{\partial T_{x0ijk}}{\partial \phi_y} \phi_{y1} \right) C_{0ijk} \right] \\ \frac{dJ_{y2}}{ds} &= \sum_{i=0}^{I(m/2)} \sum_{j=0}^{I(m/2)-i} \sum_{k=0}^i m_{yik} \left[ \left( \frac{\partial C_{0ijk}}{\partial J_y} J_{x1} + \frac{\partial C_{0ijk}}{\partial J_x} J_{y1} \right) T_{y0ijk} + \left( \frac{\partial T_{y0ijk}}{\partial \phi_x} \phi_{x1} + \frac{\partial T_{y0ijk}}{\partial \phi_y} \phi_{y1} \right) C_{0ijk} \right]\end{aligned}\tag{32}$$

Due to the products  $J_{xi} T_{x0ijk}$ ...the multiplication of the trigonometric functions gives rise to a new generation of modes defined by the eight characteristic arguments

$$\begin{aligned}
 \Psi_1 &= \mu_{ijk}^+(s) + \mu_{iljlk1}^+(\sigma) \\
 \Psi_2 &= \mu_{ijk}^+(s) - \mu_{iljlk1}^+(\sigma) \\
 \Psi_3 &= \mu_{ijk}^-(s) + \mu_{iljlk1}^-(\sigma) \\
 \Psi_4 &= \mu_{ijk}^-(s) - \mu_{iljlk1}^-(\sigma) \\
 \Psi_5 &= \mu_{ijk}^+(s) + \mu_{iljlk1}^-(\sigma) \\
 \Psi_6 &= \mu_{ijk}^+(s) - \mu_{iljlk1}^-(\sigma) \\
 \Psi_7 &= \mu_{ijk}^-(s) + \mu_{iljlk1}^+(\sigma) \\
 \Psi_8 &= \mu_{ijk}^-(s) - \mu_{iljlk1}^+(\sigma)
 \end{aligned} \tag{33}$$

and the four types of coefficients

$$\begin{aligned}
 b_1 &= \frac{1}{2} m_{x i l j l} \frac{\partial C_{0 i j k}(s)}{\partial J_x} C_{0 i l j l k 1}(\sigma) \\
 b_2 &= \frac{1}{2} m_{x i j} \frac{\partial C_{0 i l j l k 1}(\sigma)}{\partial J_x} C_{0 i j k}(s) \\
 b_3 &= \frac{1}{2} m_{y i l j l} \frac{\partial C_{0 i j k}(s)}{\partial J_y} C_{0 i l j l k 1}(\sigma) \\
 b_4 &= \frac{1}{2} m_{y i k} \frac{\partial C_{0 i l j l k 1}(\sigma)}{\partial J_y} C_{0 i j k}(s)
 \end{aligned} \tag{34}$$

For a given set of indices  $(i, j, k), (i1, j1, k1)$ , the eight modes are independent but, among all the combinations of indices, a repetition of the modes may occur. As the coefficients  $C$  are the products of independent functions of  $(\beta_x, \beta_y)$  and  $(J_x, J_y)$ , the function of  $(\beta_x, \beta_y)$  in the coefficients  $b$  is the same for a given combination of indices. The arguments  $\psi$  always appear in cosine functions and one can form the 8-vector

$$\overline{\Psi^{(2)}} = [\cos \psi_i] \quad i = 1, 8 \tag{35}$$

The functions of the action variables can also be written as 8-vectors whose components are linear combinations of the  $b$  coefficients

$$\begin{aligned}
 \overline{A_x^{(2)}} &= (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8) \\
 \overline{A_y^{(2)}} &= (a_1, a_2, a_3, a_4, -a_5, -a_6, -a_7, -a_8)
 \end{aligned} \tag{36}$$

$$\begin{aligned}
 a_1 &= -b_1 + b_2 - b_3 + b_4 \\
 a_2 &= b_1 + b_2 + b_3 + b_4 \\
 a_3 &= -b_1 + b_2 + b_3 + b_4 \\
 a_4 &= b_1 + b_2 - b_3 + b_4 \\
 a_5 &= -b_1 + b_2 - b_3 - b_4 \\
 a_6 &= b_1 + b_2 + b_3 - b_4 \\
 a_7 &= -b_1 + b_2 + b_3 - b_4 \\
 a_8 &= b_1 + b_2 - b_3 - b_4
 \end{aligned} \tag{37}$$

The second-order perturbation of the action variables integrated over a length  $L$  have then the compact expression which is the expected generalized form of the expressions derived at the first iteration

$$\begin{aligned} J_{x2} &= \sum_{i=0}^{I(m/2)} \sum_{j=0}^{I(m/2)-i} \sum_{k=0}^i m_{xij} \int_0^L ds \sum_{i1=0}^{I(m/2)} \sum_{jl=0}^{K(m/2)-il} \sum_{kl=0}^{jl} \int_0^s \overline{A_x^{(2)}} \cdot \overline{\Psi^{(2)}} d\sigma \\ J_{y2} &= \sum_{i=0}^{I(m/2)} \sum_{j=0}^{I(m/2)-i} \sum_{k=0}^i m_{yik} \int_0^L ds \sum_{i1=0}^{I(m/2)} \sum_{jl=0}^{I(m/2)-il} \sum_{kl=0}^{jl} \int_0^s \overline{A_y^{(2)}} \cdot \overline{\Psi^{(2)}} d\sigma \end{aligned} \quad (38)$$

## 5. ACTION - ANGLE PERTURBATIONS IN A PERIODIC LATTICE

When the particle performs revolutions inside a ring, the solution to the equations of motion must reflect the periodic structure of the magnetic lattice. A ring has an intrinsic  $n$ -fold symmetry ( $n \geq 1$ ) and is said to be made of  $n$  *superperiods*; a superperiod is an assembly of more elementary repetitive elements: the *periods* are also called the *cells* of the lattice. The  $C$  coefficients are the same for all the superperiods and at the position of a given element the linear phase is  $\mu + 2(n-1)\pi Q$  with  $n=1, \dots, N$  in the lattice unfolded over  $N$  superperiods. The purpose of this section is to show how the full integration over the  $N$  superperiods can be limited to an integration over a single superperiod.

### 5.1 First-order perturbation

As the perturbing fields are zero everywhere but at magnet locations, the following type of substitutions can be performed in the solutions (29)

$$\int_0^s \overline{A^{(1)}} \cdot \overline{\Psi^{(1)}} d\sigma = \sum_{i_p=1}^p \int_0^l \overline{A_{i_p}^{(1)}} \cdot \overline{\Psi_{i_p}^{(1)}} d\sigma \quad (39)$$

where  $p$  is the number of magnets. The length  $l$  may vary from magnet to magnet and the integral can be calculated exactly. The analytical integration is useful for applications which require accurate results. It is, however, very heavy and it is much simpler to describe the essential features of the particle oscillation using the *thin-lens* approximation which consists of concentrating all the field into a  $\delta$ -function (Fig.1). The  $\int$  symbol thus disappears from the formulae provided the rule

$$\int_0^s \overline{A^{(1)}} \cdot \overline{\Psi^{(1)}} d\sigma = \sum_{i_p=1}^p \overline{A_{i_p}^{(1)}} \cdot \overline{\Psi_{i_p}^{(1)}} l \quad (40)$$

is applied.

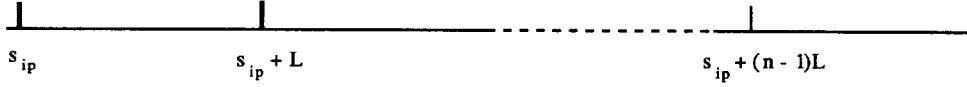


Fig. 1 Unfolded lattice with thin lenses in a first-order perturbation

By inspection of the expression (28), it can be seen that the cumulative effect of a perturbing element  $i_p$  over  $N$  superperiods is found by evaluating geometric series of the type

$$S = \sum_{n=1}^N \exp i \overline{m_{ijk}^\pm} (\bar{\mu} + (n-1)2\pi\bar{Q}) \quad (41)$$

where

$$\bar{Q} = (Q_x, Q_y) \quad (42)$$

The perturbation of the action at the end of the  $N$ -th superperiod is thus obtained from the expressions (29) and (30) where the integration is limited to the length  $L$  of a superperiod and the vector  $\Psi^{(1)}$  is given by

$$\overline{\Psi^{(1)}} = \overline{\Psi_0^{(1)}} + \overline{\Psi_N^{(1)}} \quad (43)$$

with

$$\overline{\Psi_0^{(1)}} = \left( \frac{\cos \overline{m_{ijk}^+}(\bar{\mu} - \pi\bar{Q})}{2 \sin \overline{m_{ijk}^+} \cdot \pi\bar{Q}}, \frac{\cos \overline{m_{ijk}^-}(\bar{\mu} - \pi\bar{Q})}{2 \sin \overline{m_{ijk}^-} \cdot \pi\bar{Q}} \right) \quad (44)$$

$$\overline{\Psi_N^{(1)}} = \left( \frac{\cos \overline{m_{ijk}^+}(\bar{\mu} + (2N-1)\pi\bar{Q})}{2 \sin \overline{m_{ijk}^+} \cdot \pi\bar{Q}}, \frac{\cos \overline{m_{ijk}^-}(\bar{\mu} + (2N-1)\pi\bar{Q})}{2 \sin \overline{m_{ijk}^-} \cdot \pi\bar{Q}} \right) \quad (45)$$

The phase components contain a *stationary* term, independent of  $N$ , which can also be obtained as the periodic solution of the equations of motion and a *time-dependent* term superimposed on the previous one. Moreover, as one could expect, when the frequency of one of the oscillation modes becomes an integer, a *resonance* condition

$$\overline{m_{ijk}^\pm} \cdot \bar{Q} = r \quad (46)$$

is fulfilled and the perturbation theory is no longer valid. This is the famous problem of the *small denominators* which is at the heart of KAM theory [7].

## 5.2 Second-order perturbation

In the previous section, the action distortion has been calculated at the end of a superperiod by a straightforward addition of the elementary perturbations. At the second order, it is necessary to evaluate the distortion at every element inside a superperiod (Fig. 2).

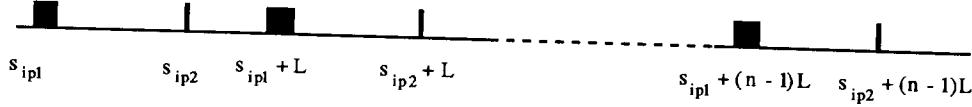


Fig. 2 Unfolded lattice in a second-order perturbation

Let us consider two elements located at  $s_1$  and  $s_2$  with  $s_2$  downstream of  $s_1$ . In the double integration of (38), the perturbation at  $s_1$  is calculated first and  $s_2$  plays the role of  $\sigma$ , then the perturbation is calculated at  $s_2$  and  $s_1$  plays the role of  $\sigma$ . It is clear that these two correlations must have a similar structure since the vectors  $A$  are the same and the relative phases can only differ by their sign. To calculate the double integral, we shall group the two correlations between elements located at  $s_1$  and  $s_2$  in a single term, the integral over  $\sigma$  is thus extended to the elements downstream to  $s$  only so that its lower and upper limits are  $s$  and  $s+L$  respectively. The thin-lens model can be used in *almost* all cases. For correlations between different magnets (*cross-correlation terms*) or between images of a magnet in all the superperiods, one can indeed write

$$\int_0^L \delta(s-s_{ip}) ds \int_0^s \overline{A^{(2)}} \cdot \overline{\Psi^{(2)}} \delta(\sigma-\sigma_{ip}) d\sigma = \sum_{i_p=1}^p \sum_{i_p=1}^{i_p-1} \overline{A^{(2)}} \cdot \overline{\Psi^{(2)}} j^2 \quad (47)$$

but, for the correlation of a magnet with itself (*self-correlation terms*), the double integral must be calculated differently using an explicit form for  $\Psi^{(2)}$ .

### 5.2.1 *Self correlation*

- By inspection of the structure of the arguments  $\psi_i$ , it turns out that the components of  $\Psi^{(2)}$  are

$$\begin{aligned} \Psi_i^{(2)} &= \operatorname{Re} e^{i [\mu_{ijk}^{\epsilon_1}(s) + \epsilon_2 \mu_{ijlkl}^{\epsilon_1}(\sigma)]} \quad i = 1, 2, 3, 4 \\ \Psi_i^{(2)} &= \operatorname{Re} e^{i [\mu_{ijk}^{\epsilon_1}(s) + \epsilon_2 \mu_{ijlkl}^{\epsilon_1}(\sigma)]} \quad i = 5, 6, 7, 8 \end{aligned} \quad (48)$$

where  $\epsilon_1$  and  $\epsilon_2$  are either +1 or -1. For a single magnet, the simplest approximation which can be used beyond the thin-lens model assumes that the length is small enough that  $\beta$  can be considered as constant and therefore  $\mu$  as a linear function of  $s$  according to (3)

$$\begin{aligned} \mu(s) &= \mu(s_{ip}) + \frac{s}{\beta_{ip}} \\ \mu(\sigma) &= \mu(s_{ip}) + \frac{\sigma}{\beta_{ip}} \end{aligned} \quad (49)$$

It is then elementary to show that the contribution of the self-correlation term to the second-order perturbation of the action in one superperiod is

$$\frac{1}{2} \overline{A_1^{(2)}} \cdot \overline{\Psi_1^{(2)}} l^2$$

whatever  $\epsilon_1$  or  $\epsilon_2$  may be; the vectors are evaluated at the beginning of the element. When the perturbation is summed over N superperiods (Fig. 3), it is useful to define the quantities

$$Q_{ijk}^{\pm} = \overline{m_{ijk}^{\pm}} \cdot \bar{Q} \quad (50)$$

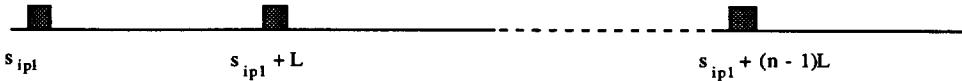


Fig. 3 Cumulative effect of single elements in a second-order perturbation

The contribution from the self-correlation terms can then be written

$$I_1 = \frac{1}{2} \overline{A_1^{(2)}} \cdot \overline{\Psi_1^{(2)}} l^2 \sum_{n=1}^N \operatorname{Re} e^{i 2\pi (Q_{ijk}^{\epsilon_1} + \epsilon_2 Q_{iljkl}^{\pm \epsilon_1}) (n-1)} \quad (51)$$

For some combinations of modes that we call *subtractive* and note with the superscript "-", the exponent vanishes if the betatron tunes fulfill the condition

$$Q_{ijk}^{\epsilon_1} + \epsilon_2 Q_{iljkl}^{\pm \epsilon_1} = 0 \quad (52)$$

which can also be applied to the betatron phases

$$\mu_{ijk}^{\epsilon_1}(s_{ip}) + \epsilon_2 \mu_{iljkl}^{\pm \epsilon_1}(s_{ip}) = 0 \quad (53)$$

$I_1$  takes the value

$$I_1 = \frac{1}{2} \overline{A_1^{(2)-}} \cdot \overline{\Psi_1^{(2)}} l^2 \quad (54)$$

the components of  $\Psi^{(2)}$  – being simply

$$\Psi_{1i}^{(2)-} = N \quad (55)$$

The other combinations are said to be *additive*, they are noted with the superscript "+" and their contribution is

$$I_1^t = \frac{1}{2} \overline{A_1^{(2)+}} \cdot \overline{\Psi_1^{(2)t}} I^2 \quad (56)$$

the components of  $\Psi^{(2)+}$  being given by

$$\Psi_{li}^{(2)+} = \Psi_{0,li}^{(2)+} + \Psi_{N,li}^{(2)+} \quad (57)$$

$$\Psi_{0,li}^{(2)+} = \frac{\sin(\pi Q_{ijk}^{e_i} - \mu_{ijk}^{e_i}(s_{ip}) + \varepsilon_2 (\pi Q_{iljlk1}^{\pm e_i} - \mu_{iljlk1}^{\pm e_i}(s_{ip})))}{2 \sin \pi(Q_{ijk}^{e_i} + \varepsilon_2 Q_{iljlk1}^{\pm e_i})} \quad (58)$$

$$\Psi_{N,li}^{(2)+} = -\frac{\sin(\pi(2N-1)Q_{ijk}^{e_i} + \mu_{ijk}^{e_i}(s_{ip}) + \varepsilon_2 (\pi(2N-1)Q_{iljlk1}^{\pm e_i} + \mu_{iljlk1}^{\pm e_i}(s_{ip})))}{2 \sin \pi(Q_{ijk}^{e_i} + \varepsilon_2 Q_{iljlk1}^{\pm e_i})} \quad (59)$$

The expression of  $I_1$  is puzzling because it contains terms which increase with  $N$ , the *secular* terms, and this is in contradiction with the principle of energy conservation. We shall see that the correlations between images of a same element in the various superperiods have also secular terms which, fortunately, cancel the previous ones.

### 5.2.2 Correlations between an element and its images in the superperiods

The distribution of the images of a single element is shown in Fig. 4.

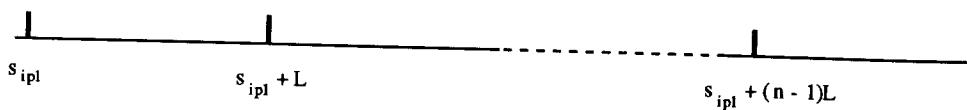


Fig. 4 Images of a single element in the superperiods

We have to consider the new contribution to the action distortion

$$I_2 = \overline{A_1^{(2)}} \cdot \overline{\Psi_2^{(2)}} I^2 \quad (60)$$

where the components of  $\Psi_2$  are

$$\Psi_2^{(2)} = \sum_{n=1}^N \operatorname{Re} e^{i[\mu_{ijk}^{e_i}(s_{ip}) + (n-1)2\pi Q_{ijk}^{e_i} + \varepsilon_2 \mu_{iljlk1}^{\pm e_i}(s_{ip})]} [1 + \dots + e^{i(n-2)\varepsilon_2 2\pi Q_{iljlk1}^{\pm e_i}}] \quad (61)$$

The decomposition made for  $I_1$  can be repeated for  $I_2$  and it turns out that

$$I_2 = -I_1 + \frac{1}{2} \overline{A_1^{(2)}} \cdot \overline{\Psi_2^{(2)}} I^2 \quad (62)$$

$$I_2^t = \frac{1}{2} \overline{A_1^{(2)+}} \cdot \overline{\Psi_2^{(2)+}} I^2 \quad (63)$$

with

$$\Psi_{2i}^{(2)} = \left( \frac{\sin \pi N Q_{ijk}^{\epsilon_i}}{\sin \pi Q_{ijk}^{\epsilon_i}} \right)^2 \quad (64)$$

$$\Psi_{2i}^{(2)*} = \Psi_{0,2i}^{(2)*} + \Psi_{N,2i}^{(2)*} \quad (65)$$

$$\begin{aligned} \Psi_{0,2i}^{(2)*} = & - \frac{\cos [\mu_{ijk}^{\epsilon_i} + \epsilon_2 \mu_{i1j1k1}^{\pm\epsilon_i} - \pi (Q_{ijk}^{\epsilon_i} + \epsilon_2 Q_{i1j1k1}^{\pm\epsilon_i})]}{2 \sin \pi Q_{i1j1k1}^{\pm\epsilon_i} \sin \pi Q_{ijk}^{\epsilon_i}} \\ & + \frac{\cos [\mu_{ijk}^{\epsilon_i} + \epsilon_2 \mu_{i1j1k1}^{\pm\epsilon_i} - \pi (Q_{ijk}^{\epsilon_i} + 2\epsilon_2 Q_{i1j1k1}^{\pm\epsilon_i})]}{2 \sin \pi Q_{i1j1k1}^{\pm\epsilon_i} \sin \pi (Q_{ijk}^{\epsilon_i} + \epsilon_2 Q_{i1j1k1}^{\pm\epsilon_i})} \end{aligned} \quad (66)$$

$$\begin{aligned} \Psi_{N,2i}^{(2)*} = & \frac{\cos [\mu_{ijk}^{\epsilon_i} + \epsilon_2 \mu_{i1j1k1}^{\pm\epsilon_i} + \pi ((2N-1)Q_{ijk}^{\epsilon_i} + \epsilon_2 Q_{i1j1k1}^{\pm\epsilon_i})]}{2 \sin \pi Q_{i1j1k1}^{\pm\epsilon_i} \sin \pi Q_{ijk}^{\epsilon_i}} \\ & - \frac{\cos [\mu_{ijk}^{\epsilon_i} + \epsilon_2 \mu_{i1j1k1}^{\pm\epsilon_i} - \pi ((2N-1)Q_{ijk}^{\epsilon_i} + 2\epsilon_2 (N-1)Q_{i1j1k1}^{\pm\epsilon_i})]}{2 \sin \pi Q_{i1j1k1}^{\pm\epsilon_i} \sin \pi (Q_{ijk}^{\epsilon_i} + \epsilon_2 Q_{i1j1k1}^{\pm\epsilon_i})} \end{aligned} \quad (67)$$

For the subtractive modes, the secular terms in (62) cancel those found in the previous section and the phase vector (64) is independent of the position of the element.

### 5.2.3 *Cross-correlations*

For the interaction between elements located at two different positions (Fig. 5) in the superperiod, the calculation can be conducted in the same way as in the previous case.

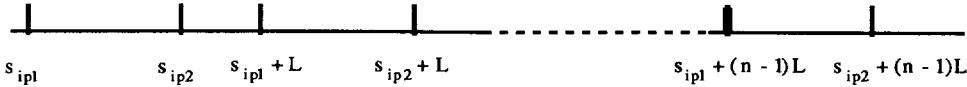


Fig 5 Cross-correlations between distinct elements

The contribution to the action distortion is then

$$I_3 = \overline{A_{12}^{(2)}} \cdot \overline{\Psi_3^{(2)}} l^2 \quad (68)$$

where the phase vector is of the form

$$\overline{\Psi_3^{(2)}} = \overline{\Psi_{12}^{(2)}} + \overline{\Psi_{21}^{(2)}} \quad (69)$$

with the components

$$\Psi_{12i}^{(2)} = \sum_{n=1}^N \operatorname{Re} e^{i[\mu_{ijk}^{\varepsilon_i}(s_{ip1}) + (n-1)2\pi Q_{ijk}^{\varepsilon_i} + \varepsilon_2 \mu_{i1j1k1}^{\pm\varepsilon_i}(s_{ip2})]} [1 + \dots + e^{i(n-2)\varepsilon_2 2\pi Q_{i1j1k1}^{\pm\varepsilon_i}}] \quad (70)$$

$$\Psi_{21i}^{(2)} = \sum_{n=1}^N \operatorname{Re} e^{i[\mu_{ijk}^{\varepsilon_i}(s_{ip2}) + (n-1)2\pi Q_{ijk}^{\varepsilon_i} + \varepsilon_2 \mu_{i1j1k1}^{\pm\varepsilon_i}(s_{ip1})]} [1 + \dots + e^{i(n-1)\varepsilon_2 2\pi Q_{i1j1k1}^{\pm\varepsilon_i}}] \quad (71)$$

After summation and addition of the two types of correlations, one finds that the secular terms cancel out and that the residual contribution of the subtractive combinations is

$$I_3 = \overline{A_{12}^{(2)}} \cdot \overline{\Psi_3^{(2)}} l^2 \quad (72)$$

with

$$\Psi_3^{(2)} = \left( \frac{\sin N\pi Q_{ijk}^{\varepsilon_i}}{\sin \pi Q_{ijk}^{\varepsilon_i}} \right)^2 \cos(\mu_{ijk}^{\varepsilon_i}(s_{ip2}) - \mu_{ijk}^{\varepsilon_i}(s_{ip1})) \quad (73)$$

It is interesting to note that, thanks to the relation (53), a single set of modes characterized by the indices  $(i, j, k)$  is sufficient to calculate this type of cross-correlation modes. For the additive modes,  $I_3$  becomes

$$I_3^+ = \overline{A_{12}^{(2)+}} \cdot \overline{\Psi_3^{(2)+}} l^2 \quad (74)$$

with

$$\Psi_{3i}^{(2)+} = \Psi_{0,3i}^{(2)+} + \Psi_{N,3i}^{(2)+} \quad (75)$$

$$\begin{aligned} \Psi_{0,3i}^{(2)+} &= \frac{1}{4 \sin \pi \varepsilon_2 Q_{i1j1k1}^{\pm\varepsilon_i} \sin \pi (Q_{ijk}^{\varepsilon_i} + \varepsilon_2 Q_{i1j1k1}^{\pm\varepsilon_i})} \\ &\quad \left\{ \cos(\mu_{ijk}^{\varepsilon_i}(s_{ip1}) + \varepsilon_2 \mu_{i1j1k1}^{\pm\varepsilon_i}(s_{ip2}) - \pi(Q_{ijk}^{\varepsilon_i} + 2\varepsilon_2 Q_{i1j1k1}^{\pm\varepsilon_i})) \right. \\ &\quad \left. + \cos(\mu_{ijk}^{\varepsilon_i}(s_{ip2}) + \varepsilon_2 \mu_{i1j1k1}^{\pm\varepsilon_i}(s_{ip1}) - \pi Q_{ijk}^{\varepsilon_i}) \right\} \\ &- \frac{1}{4 \sin \pi \varepsilon_2 Q_{i1j1k1}^{\pm\varepsilon_i} \sin \pi Q_{ijk}^{\varepsilon_i}} \\ &\quad \left\{ \cos(\mu_{ijk}^{\varepsilon_i}(s_{ip1}) + \varepsilon_2 \mu_{i1j1k1}^{\pm\varepsilon_i}(s_{ip2}) - \pi(Q_{ijk}^{\varepsilon_i} + \varepsilon_2 Q_{i1j1k1}^{\pm\varepsilon_i})) \right. \\ &\quad \left. + \cos(\mu_{ijk}^{\varepsilon_i}(s_{ip2}) + \varepsilon_2 \mu_{i1j1k1}^{\pm\varepsilon_i}(s_{ip1}) - \pi(Q_{ijk}^{\varepsilon_i} + \varepsilon_2 Q_{i1j1k1}^{\pm\varepsilon_i})) \right\} \quad (76) \end{aligned}$$

$$\begin{aligned}
 \Psi_{N,3i}^{(2)+} = & -\frac{1}{4 \sin \pi \epsilon_2 Q_{i1j1k1}^{\pm \epsilon_1} \sin \pi (Q_{ijk}^{\epsilon_1} + \epsilon_2 Q_{i1j1k1}^{\pm \epsilon_1})} \\
 & \left\{ \cos (\mu_{ijk}^{\epsilon_1}(s_{ip1}) + \epsilon_2 \mu_{i1j1k1}^{\pm \epsilon_1}(s_{ip2}) + \pi ((2N-1)Q_{ijk}^{\epsilon_1} + 2(N-1)\epsilon_2 Q_{i1j1k1}^{\pm \epsilon_1})) \right. \\
 & + \cos (\mu_{ijk}^{\epsilon_1}(s_{ip2}) + \epsilon_2 \mu_{i1j1k1}^{\pm \epsilon_1}(s_{ip1}) + \pi ((2N-1)Q_{ijk}^{\epsilon_1} + 2N\epsilon_2 Q_{i1j1k1}^{\pm \epsilon_1})) \\
 & + \frac{1}{4 \sin \pi \epsilon_2 Q_{i1j1k1}^{\pm \epsilon_1} \sin \pi Q_{ijk}^{\epsilon_1}} \\
 & \left. \left\{ \cos (\mu_{ijk}^{\epsilon_1}(s_{ip1}) + \epsilon_2 \mu_{i1j1k1}^{\pm \epsilon_1}(s_{ip2}) + \pi ((2N-1)Q_{ijk}^{\epsilon_1} - \epsilon_2 Q_{i1j1k1}^{\pm \epsilon_1})) \right. \right. \\
 & \left. \left. + \cos (\mu_{ijk}^{\epsilon_1}(s_{ip2}) + \epsilon_2 \mu_{i1j1k1}^{\pm \epsilon_1}(s_{ip1}) + \pi ((2N-1)Q_{ijk}^{\epsilon_1} - \epsilon_2 Q_{i1j1k1}^{\pm \epsilon_1})) \right\} \right\} \quad (77)
 \end{aligned}$$

These formulae complete the treatment of second-order perturbations. They are not easy to read but the complication inherent to high-order perturbations is less and less untractable with symbolic codes such as Macsyma, Reduce or *Mathematica* [12] and it is left as an exercise to tabulate the resonances generated by a sextupolar field and their driving terms.

## 6. COMPENSATION SCHEMES FOR AMPLITUDE DEPENDENT DISTORTIONS

As already mentioned in the introduction, systematic non-linear fields are introduced to act on the chromatic properties of the machine and it is the amplitude dependent distortions they produce that we want to correct. Their coupling with the off-momentum particles depends on the distance between the off-momentum closed orbit and the central orbit via a quantity called the *orbit dispersion*. If one cannot design a chromaticity correction system self compensated [8] with respect to the non-linear distortions, and this is often the case save in very large storage rings, the correcting fields are to be placed in dispersionless straight sections. The method which will be described [9] can be applied in any circumstance but, for clarity, we shall assume that momentum and amplitude dependent effects are decoupled.

### 6.1 Method

There are a few preliminary remarks which are very important in the design of a correction scheme. First, all the distortions depend not only on the initial amplitude but also on the initial phase of the particle oscillation via the betatron phase advance  $\mu$  (Eq. 14). A correction scheme must obviously be valid for all the particles in the beam. In order for it to be phase independent, the cosine and sine components of the oscillation have to be corrected simultaneously; in other terms, a *vector* correction has to be performed. A simplification occurs when the machine superperiod has a specular symmetry; then, for two symmetric elements, the driving term has the same amplitude and the associated phases are  $m(-\pi Q + \mu)$  and  $m(-\pi Q + (2\pi Q - \mu))$  and thus opposite so that the sine components vanish. Amplitude independence is obtained by using a correcting field of the same nature as the chromaticity field. Let us note that a correction independent of the initial phase is also independent of the number

of turns so that it is sufficient to take the periodic terms into consideration; this means that, when a scatter plot of the action or angle variables is observed at some position around the machine, both the average value and the oscillation of the variable about its average value are reduced.

Another aspect of the correction concerns the observation of the distortion. If the scheme respects the periodicity of the ring, it seems that a single observation point at the end of the superperiod is sufficient provided it is not a node of the oscillation. For a given chromatic field, the indices  $m_{xij}$  (Eq. 16) and  $m_{yik}$  (Eq. 17) of the oscillation modes are determined. For instance, the modes of a sextupolar field are characterized by the four couples (1,0), (3,0), (1,2), (1,-2). The cosine and sine components of each mode are then calculated to the first order at the observation point for the action distortion using the relations (26), (29), (39) or (40) and (44), they form a known vector  $\mathbf{b}$ . In the case of a sextupole,  $\mathbf{b}$  has 8 or 4 independent components, depending on the superperiod symmetry:

$$\mathbf{b} = \sum_{\substack{\text{chromaticity} \\ \text{sextupoles}}} K' l \sqrt{2 J_x \beta_x} \begin{pmatrix} (J_x \beta_x - J_y \beta_y) \cos(\mu_x - \mu_{x0}) \\ J_x \beta_x \cos 3(\mu_x - \mu_{x0}) \\ J_y \beta_y \cos(\mu_x - \mu_{x0} + 2(\mu_y - \mu_{y0})) \\ J_y \beta_y \cos(\mu_x - \mu_{x0} - 2(\mu_y - \mu_{y0})) \end{pmatrix} \quad (78)$$

The four other components are obtained by replacing cos by sin. The constant factors have been omitted. All the variables are characteristic of the sextupoles with the exception of the phase terms subscripted with 0 which are referred to the observation point.

The correction multipoles, in number  $n$ , are located at every possible place. Each corrector is characterized by a vector calculated with the same expressions as those used for the chromaticity multipoles but with a unit strength ( $K'l=1$ ). Each correction vector is a column in the correction matrix  $\mathbf{A}$ . The correction vector  $\mathbf{x}$  is made of the  $n$  unknown correction strengths and the residual vector to be minimized is

$$\mathbf{r} = \mathbf{A} \mathbf{x} + \mathbf{b} \quad (79)$$

A good minimization is obtained with the MICADO program [10] which selects the most efficient correctors in an iterative process, the number of correctors being equal to the order of the iteration. It is theoretically possible to get  $\mathbf{r}$  zero but an exact cancellation of the first-order terms usually leads to a catastrophic over compensation of the non-linear distortions.

If a second-order calculation turns out to be necessary, the relation (79) is still valid if  $\mathbf{A}$  is considered as a function of  $\mathbf{x}$ . The correction multipoles selected in the first iteration are maintained but their strengths are recalculated by adjoining to the first-order expressions the second-order expressions calculated in section 5.2 and minimizing the norm of  $\mathbf{r}$  with respect to its components.

The last step consists of testing the correction scheme using a numerical tracking program which provides scatter plots of the motion and permits the quality of the correction to be assessed. In these programs, the equations of motion are integrated numerically [11] in the non-linear fields and the solution is not limited by the order of a perturbative treatment.

## 6.2 Applications

The method which has just been outlined was applied to the CERN Antiproton Collector (ACOL) and to the Berkeley Advanced Light Source (ALS). An antiproton beam does not radiate and the goal is to make the beam envelope respect the size defined by the linear emittance as close as possible. In a synchrotron light source, the electrons or the positrons are ultra relativistic and radiate significantly; in the process of photon emission, a particle can reach a very large amplitude, typically more than ten times the standard beam width, and the criterion is rather to get the maximum *dynamic aperture*. The dynamic aperture is defined as the set of all the couples of initial transverse coordinates (x,y) for which the particle oscillation is stable.

### 6.2.1 Antiproton Collector

The ring has a twofold superperiodicity and each superperiod has a specular symmetry. A quadrant is represented in Fig. 6. The horizontal and vertical variations of the linear betatron tune with the momentum and the quadratic momentum dependence of the orbit in the long straight sections are controlled by three families of sextupoles (SF1, SF2, SD) in the arcs . The sextupolar field is superimposed on the linear field of the quadrupoles and obtained with a special design of the pole profiles.

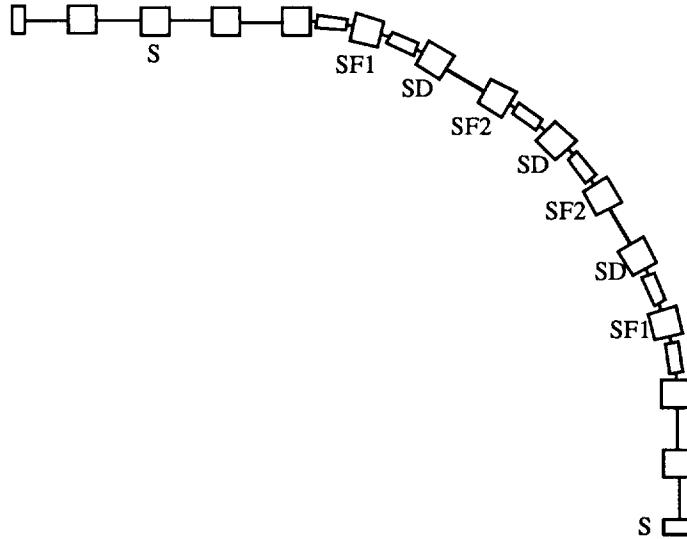
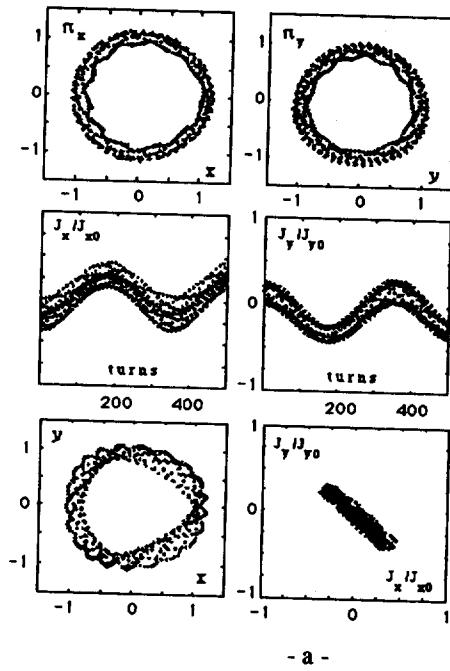
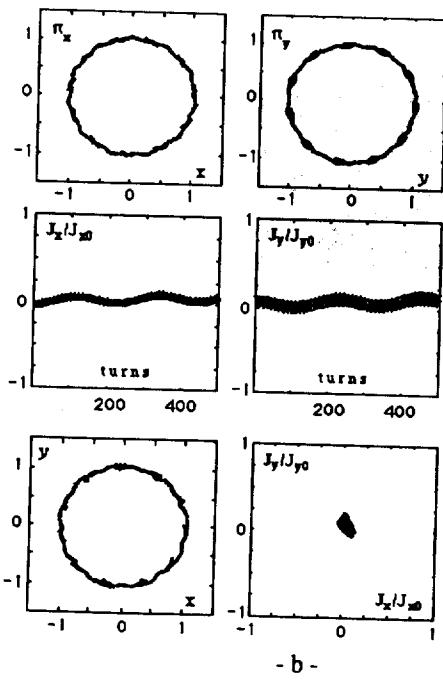


Fig. 6 Quadrant of ACOL

The correction sextupolar fields were first distributed all along the straight sections and computations made with MICADO and tested with a numerical tracking program showed that a sufficient correction (Fig. 7) could be obtained with the sextupoles S produced by pole face windings (Fig. 8 ).



- a -



- b -

Fig. 7 Transverse dynamics in ACOL before (a) and after (b) correction

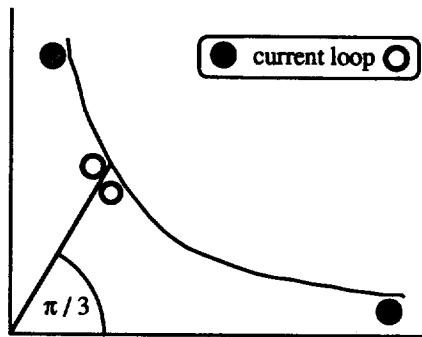


Fig. 8 Configuration of the sextupolar conductors on the pole of a quadrupole

The experimental evidence of the importance of this type of correction should be given by comparing the number of antiprotons injected into the ring without and with sextupolar correction but this has not been possible until now because the antiproton beam had insufficient emittance. However, two measurements prove the validity of the method:

- i) the non-linear coupling observed when the horizontal and vertical betatron tunes are equal almost disappears after correction (Fig. 9);

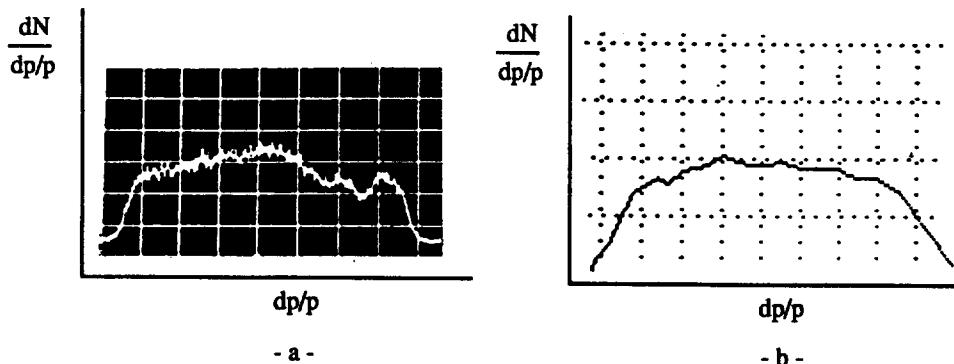


Fig. 9 Beam density distribution before (a) and after (b) sextupolar correction

- ii) the motion of the bunch center of gravity resulting from a large injection error both in the horizontal and vertical planes (see F. Willeke, these proceedings) is much more regular after correction (Fig. 10).

#### 6.2.2 Advanced Light Source

The Berkeley Advanced Light Source (ALS) has a 12-fold superperiodicity and each superperiod (Fig. 11) has a mirror symmetry. The computational techniques used for ACOL have been resumed for this machine.

The main difference between the two lattices lies in the *triple bend achromat* which makes tiny orbit dispersion and leads to outstandingly high strengths for the chromaticity sextupoles:  $K'l_F = 11.4 \text{ m}^{-2}$ ,  $K'l_D = -8.6 \text{ m}^{-2}$ .

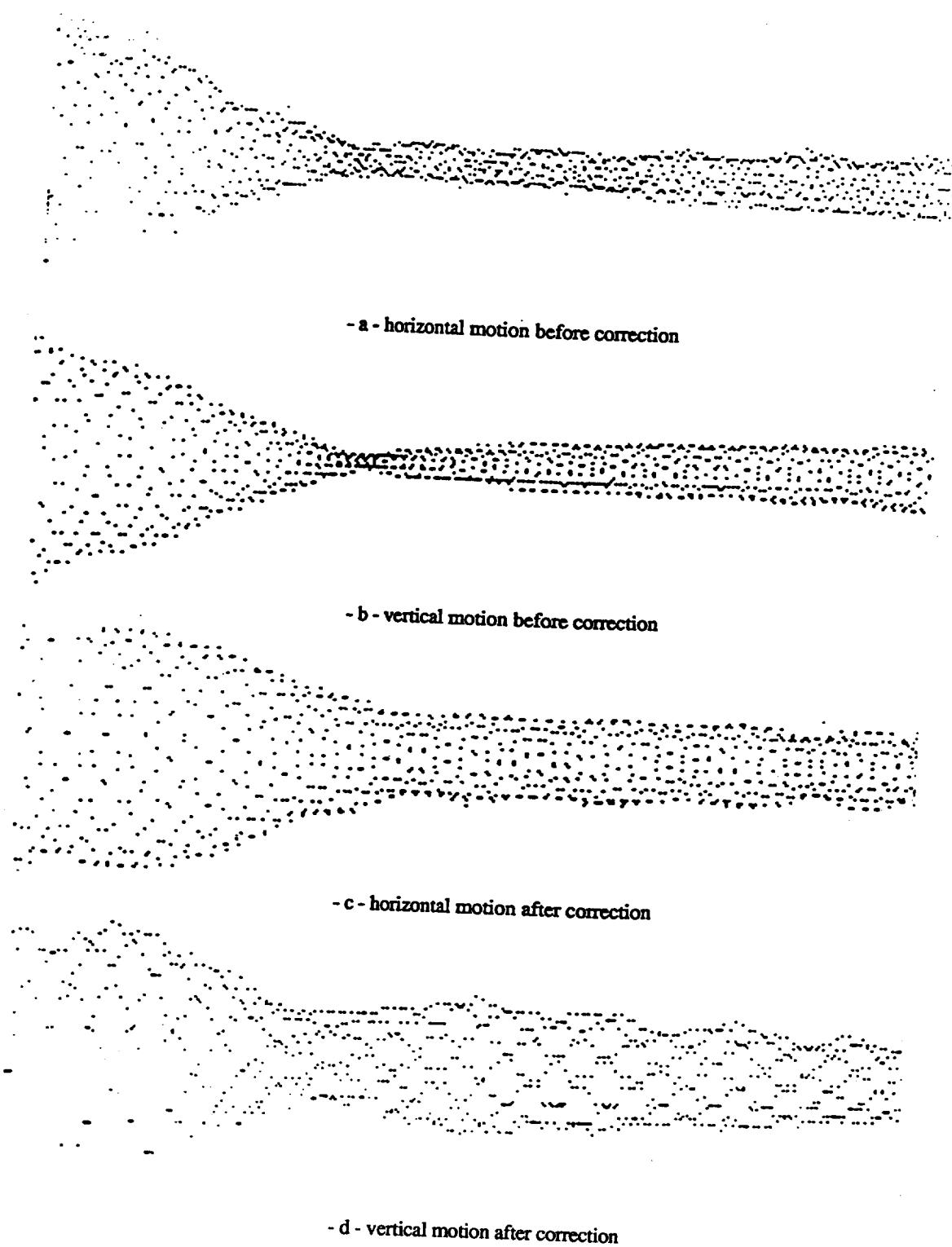


Fig. 10 Motion of the bunch center of gravity

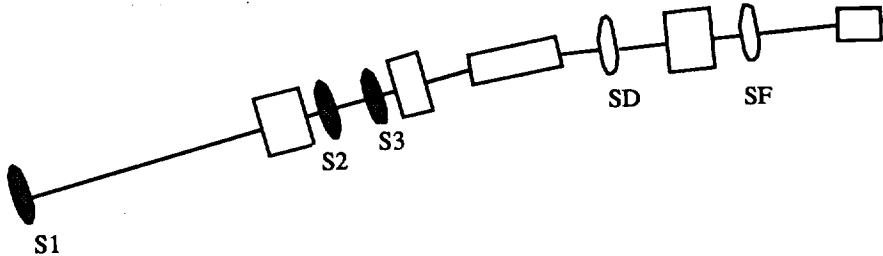


Fig. 11 ALS half superperiod

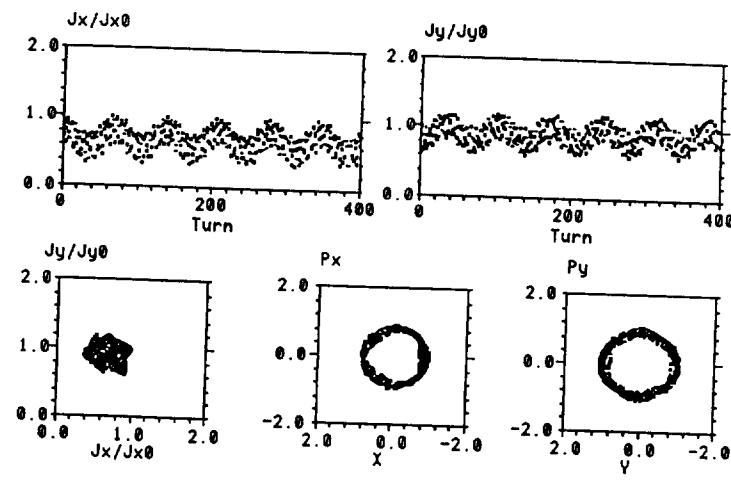
Two schemes were defined: one with a single weak sextupole ( $K'l_2 = .87 \text{ m}^{-2}$ ), and another one with three stronger sextupoles ( $K'l_1 = -4.49 \text{ m}^{-2}$ ,  $K'l_2 = 5.96 \text{ m}^{-2}$ ,  $K'l_3 = -13.2 \text{ m}^{-2}$ ). From the inspection of the scatter plots (Fig. 12), it seems that the last scheme (case c) should give the best results. As a matter of fact, the improvement of the dynamic aperture (Fig. 13), if there is any, with respect to case b does not justify the complication of the three sextupole scheme. Each couple  $(x, y)$  which defines the dynamic aperture corresponds to the first unstable particle of a sample of 25 particles, all with the same initial amplitude but with different initial phases. To illustrate the importance of that definition, Fig. 14 shows the dynamic aperture which would be deduced from a single particle (scattered points) as compared to the one which results from a multiple particle tracking (connected points). As soon as a particle is near its limit of stability, the first-order perturbation theory is no longer sufficient and future schemes will be based on formulae expanded to the second order.

## 7. CONCLUSION

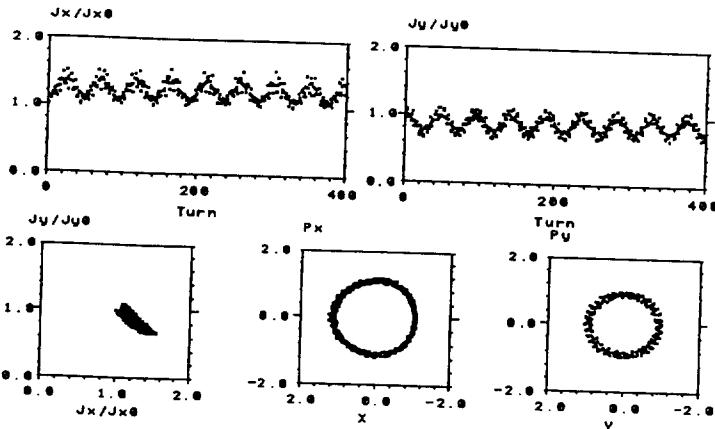
A theory of the non-linear betatron oscillation has been presented based on a solution of the action-angle equations of motion expanded to the second order. Emphasis has been put on the action variable because it is directly related to the beam size. The intricacy of this type of calculation is greatly alleviated by the use of symbolic programs. The insight given by analytical calculations is applied to the design of correction schemes. Until now, these schemes have been derived using the first-order formulae only and they provide substantial improvements of the particle transverse dynamics but, when the criterion of the dynamical aperture is used, it becomes necessary to resort to second-order expressions. A correction scheme has been tested experimentally and it turns out that solutions made of a single family of correctors, and therefore very easy to implement in a real machine, may lead to substantial improvements in the use of the machine aperture.

## ACKNOWLEDGEMENTS

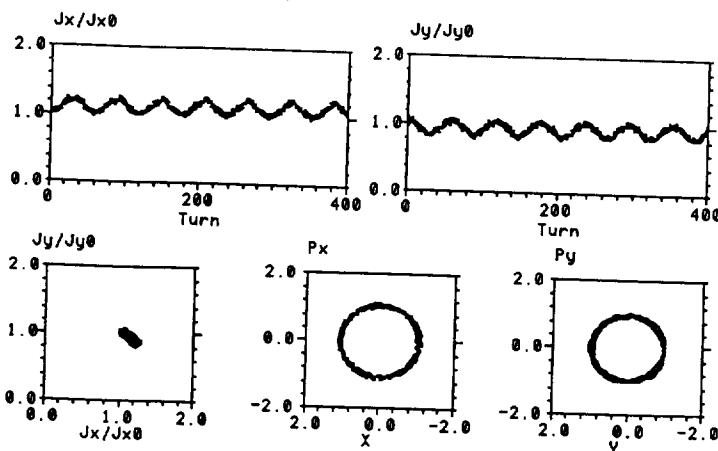
The theory has been elaborated with J. Bengtsson. Numerical tracking for ALS is due to H. Nishimura. H.H. Umstatter has designed the ACOL sextupolar pole face windings. ACOL experimental results are due to P. Krejcek. Discussions with M. Chanel and J. Bengtsson have been very stimulating thanks to the experience they have developed in the operation of the LEAR machine at CERN.



- a -



- b -



- c -

Fig. 12 Transverse dynamics in ALS for the bare machine (a), a one- (b) and a three- (c) sextupole correction scheme

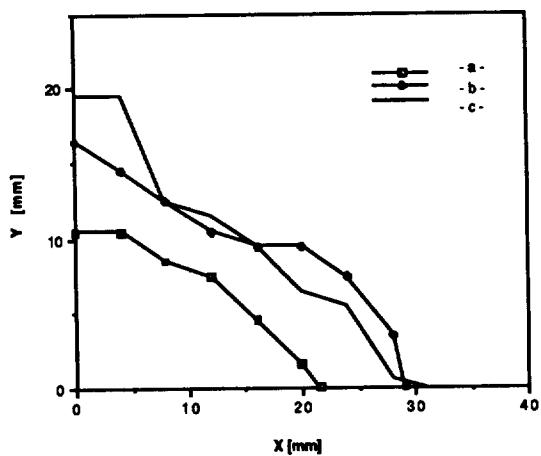


Fig. 13 Dynamic apertures for the bare machine (a) and two correction schemes (b,c).

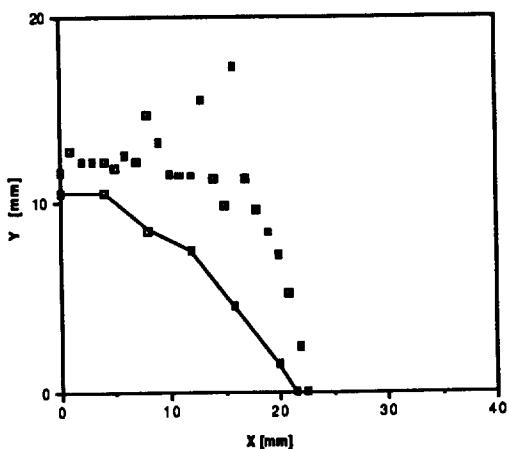


Fig. 14 Comparison of the dynamic apertures deduced from a single (scattered points ) and a multiple tracking (solid line )

#### REFERENCES

- [1] G. Guignard, *The General Theory of all Sum and Difference Resonances in a 3D Magnetic Field in a Synchrotron*. CERN 76-06 (1976).
- [2] H. Goldstein, *Classical Mechanics*. Addison Wesley (1980).
- [3] R. Ruth, *Single Particle Dynamics in Circular Accelerators*. AIP Conf. Proc. 153 (1985).
- [4] B. Autin, *The CERN Antiproton Collector*, Proc. of the CAS on Antiprotons for Colliding Beams Facilities, CERN 84-15 (1984).

- [5] *Advanced Light Source Design Report*, Lawrence Berkeley Laboratory (1988).
- [6] E. Courant and H. Snyder, *Theory of Alternating Gradient Synchrotron*, Annals of Physics 3 (1958).
- [7] V.I. Arnold, *Mathematical Methods of Classical Mechanics*, Springer Verlag (1978).
- [8] A. Verdier, *Non-Interleaved Sextupole Schemes for LEP and Related Effects of Systematic Multipole Components in the Lattice Magnets*, CERN-LEP-MA/83 (1983).
- [9] A. Ando, *Emittance Distortions with Sextupoles*, Proc.of the 12th Int. Conf. on High Energy Accelerators, Fermilab (1983).  
B. Autin, *Non-linear Betatron Oscillations*, Physics of Particle Accelerators, AIP Conf. Proc. 153 (1985).  
T. Collins, *Distortion Functions*, Fermilab 84/114 (1984).
- [10] B. Autin, *Lattice Perturbations*, Physics of Particle Accelerators, AIP Conf.Proc. 127 (1983).
- [11] A. Dragt, *Lectures on Non-linear Orbit Dynamics*, Physics of Particle Accelerators, AIP Conf. Proc. 87 (1981).
- [12] S. Wolfram, *Mathematica*, Addison Wesley (1988).



## DYNAMIC APERTURE

*A. Ropert*

ESRF, Grenoble, France

### ABSTRACT

The concept of dynamic aperture provides an overall description of the non-linear effects arising from chromaticity correcting sextupoles and field imperfections of magnets in circular accelerators. These effects tend to limit stable amplitudes of motion. The importance of this problem in recent accelerators has motivated numerous analytical and numerical developments in order to better understand the particle behaviour. Several of the analytical approaches based on Hamiltonian formalism will be described. Particle tracking is widely used to calculate the stability limit. Tracking codes will be discussed and numerical simulations compared with experimental results.

### 1. INTRODUCTION

The determination of dynamic aperture was not a critical issue in the early days of circular accelerators in which the dynamic aperture was larger than the physical aperture. It has become a matter of serious concern with the advent of a new generation of specialized rings, such as dedicated synchrotron radiation sources, large  $e^+ e^-$  colliders or superconducting accelerators. In these machines, complicated non-linear forces act on the particles and may lead to unstable motion. The main sources of non-linearities are the sextupoles which are used to compensate the natural chromaticity and the transverse multipole fields resulting from imperfections in magnets. The threshold amplitude beyond which the betatron motion becomes unbounded is called the dynamic aperture. In order to achieve design performance, it is of prime importance to understand what is the dynamic aperture of such machines and to find strategies to improve it.

Considerable effort is being pursued in many laboratories to study the problem of dynamic aperture. Most of the investigations are based on analytical developments, numerical simulations or combinations of both. The purpose of this lecture is to review the different approaches dealing with dynamic aperture and discuss their main features.

Analytical methods aim at finding criteria to estimate amplitude limitations of stable motion. Some of these methods will be sketched in Section 3. Most of them are based on perturbation techniques of some kind. By calculating to as high an order as possible non-linear distortions of the phase space as a function of initial amplitude of motion, these methods

approach the true dynamic aperture.

Another procedure consists in deriving some physical quantities which are representative of the perturbation and can be minimized in order to optimize compensation schemes and enlarge the dynamic aperture. This will be treated in Section 4.

Numerical calculations based on single particle tracking are a very powerful tool for determining the dynamic aperture. The available computer codes will be discussed in Section 5 and an example of tracking presented. Unfortunately the limitations of both analytical and numerical methods are highlighted by the difficulties encountered when trying to simulate a real machine, because many complicated phenomena are added to the non-linear dynamics. In order to illustrate this point, some experimental results will be presented in Section 6.

In the next Section, the concept of dynamic aperture is introduced and the basic design requirements discussed.

## 2 DYNAMIC APERTURE

### 2.1 Definition of the dynamic aperture

One of the major issues, when designing circular accelerators, is to ensure that the motion will remain bounded for a time long enough with respect to time scales of interest (damping times for electrons, synchrotron periods or diffusion times for protons and heavy particles). Given initial conditions in amplitude and phase ( $J_0, \phi_0$ ), the criterion is that the amplitude at any time  $t$  satisfies the inequality  $|J(t)| < A$ , where  $A$  is a finite boundary. The problem is then to find the largest value  $J_0$  for any phase  $\phi_0$  such that  $J(t)$  remains finite for a time long enough. The largest initial amplitude  $J_0$  up to which the betatron motion is stable is called the dynamic aperture.

More generally, the stable region is a volume in the six-dimensional phase space defined by the transverse emittances  $E_x$  and  $E_y$  and the longitudinal motion. One of the usual ways to represent the dynamic aperture is to make cuts in the stable volume and to plot the stable horizontal amplitudes as a function of the stable vertical amplitudes (sometimes defined according to the cut  $E_y = 1/2 E_x$ , which corresponds to full coupling in electron machines) for a given value of the momentum deviation with respect to the synchronous particle  $\delta$ .

The ideal situation corresponds to the case where the dynamic aperture is almost the same as the physical aperture of the ring which is defined mainly by the size of the vacuum chamber. If the dynamic aperture proves to be too small, some optimization work must be carried out to reduce the non-linear effects which are responsible for the reduction of dynamic aperture. This addresses the problem of calculating or predicting the dynamic aperture of a machine.

## 2.2 Determination of dynamic aperture

Hamiltonian formalism and equations of motion derived from it are the natural tools to describe the stability of motion and determine the limit of stability. The evolution of the motion of a particle as a function of time is specified by a vector  $\{x(t), p_x(t), y(t), p_y(t), z(t), p_z(t)\}$  describing a path in the six-dimensional phase space. The transverse displacements are  $x$  and  $y$ , while  $z$  is measured along the particle's trajectory. The functions  $p_x, p_y, p_z$  are the canonical momenta conjugate respectively to  $x, y, z$ . In these coordinates, the Hamiltonian describing the betatron motion in the presence of non-linear fields is given by:

$$H = \frac{1}{2} (p_x^2 + k_x x^2 + p_y^2 + k_y y^2) + H_1 \quad (1)$$

$k_x$  and  $k_y$  describe the linear focusing and  $H_1$  represents the perturbation.

$$H_1 = \sum_{n,m} A_{mn} x^m y^n \quad (2)$$

Following the general treatment of [1], one introduces action-angle variables  $(I, \phi)$  which are linked to the original variables through a generating function and obtain the new Hamiltonian of the system:

$$H = v_x I_x + v_y I_y + H_1 \quad (3)$$

This is the starting point of the formalism using the standard perturbation theory in order to investigate the effects of non-linear fields.

An alternative way to deal with the Hamiltonian-generated canonical transformation describing the evolution of the particle motion is to represent it through a symplectic transfer map  $\mathcal{M}$ . If we launch a particle into the machine with initial conditions  $(x, p_x, \dots)$ , it will come back after a single turn with transformed coordinates  $(x_f, p_{xf}, \dots)$  under the action of the map  $\mathcal{M}$ . If we iteratively apply the transformation  $\mathcal{M}$  many times and observe the particle behaviour, it is possible to make statements about the stability of the particle orbit. The symplectic mapping that carries particles around the machine can be numerically represented. This procedure of iterating the transfer map to determine if the motion is stable is called tracking.

The concept of dynamic aperture comes from the fact that in a real machine the transfer map  $\mathcal{M}$  is a non-linear map. This implies that the motion could become unstable. This behaviour is illustrated in Fig. 1 for an ideal machine when non-linear elements such as sextupoles or octupoles are included. With adequate settings of these multipoles, one can empirically increase the maximum stable amplitudes.

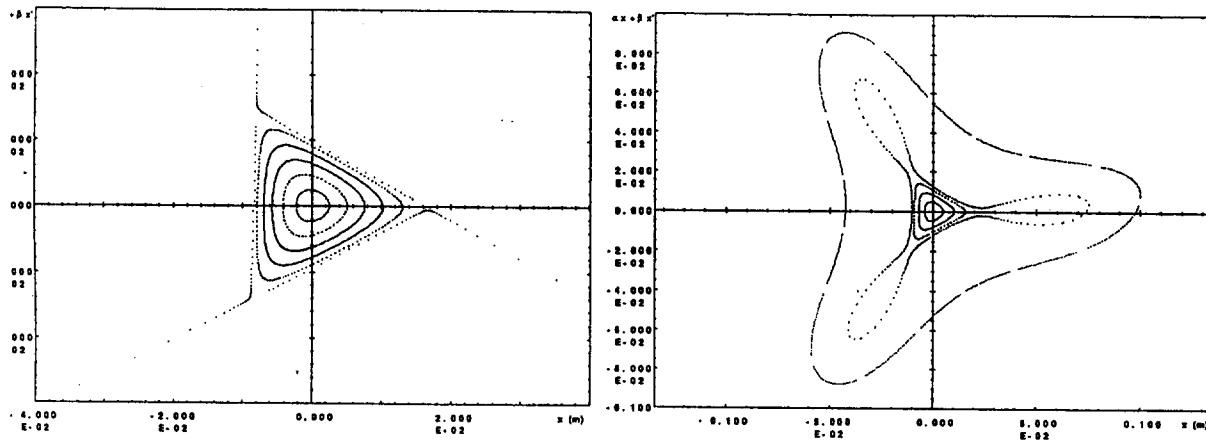


Fig. 1 Example of tracking with sextupoles and octupoles

### 2.3 Specific applications of the concept of dynamic aperture

The determination of adequate dynamic aperture has been a critical process for several types of machines in recent years. Basically there are two main categories of machines concerned by this problem:

i) chromaticity dominated machines where the main contribution to the non-linear fields comes from the chromaticity correcting sextupoles. This is the case of the new generation of synchrotron light sources which are characterized by an extremely high focusing in order to achieve low emittances and high brilliance. This implies strong gradients and also large values of betatron functions so that the chromaticity tends to be large. Consequently, the strengths of the sextupoles used in these rings are very large and introduce various kinds of geometric and chromatic aberrations, thus limiting the maximum stable amplitudes. A large dynamic aperture is needed to accommodate the oscillations of scattered particles and achieve a good beam lifetime. Various schemes are used to improve the dynamic aperture of these machines.

Dynamic aperture is also a major issue for large  $e^- e^+$  colliders. Achieving optimum beam-beam interaction and high luminosity requires very small transverse dimensions at the interaction point and the design of low- $\beta$  insertions. The sextupoles correcting the chromaticity generated by the low- $\beta$  quadrupoles are the limiting factor. Again the optimization of correction schemes allows to enlarge the dynamic aperture.

Dynamic aperture is also important for smaller machines like low energy proton machines (antiproton accumulators for instance), where the beam has a large cross section and occupies a large area in the tune diagram. In contrast with the large colliders with regular FODO arcs, these machines are generally characterized by more irregular lattices. Consequently intrinsic resonance compensation is not possible and improvement of the dynamic aperture implies dedicated correction schemes.

ii) error-dominated machines. The other type of accelerators for which the dynamic aperture is an important figure of merit is large superconducting accelerators or colliders. In these machines, the field errors in the magnets are the most important source of non-linearities, because the field quality of superconducting magnets is relatively poor. Also the magnet aperture is generally small, so that the beam is close to the coils and perturbed field regions. The dynamic aperture is firstly limited by systematic field imperfections which are mainly generated by persistent current field distortions at low excitation. Besides systematic field errors, the dynamic aperture is also reduced by random errors resulting from manufacturing tolerances. Adequate strategies have been considered to compensate these errors.

### 3 ANALYTICAL APPROACH OF THE DYNAMIC APERTURE

#### 3.1 Generalities

The numerical modelling of a machine is a powerful tool for determining the dynamic aperture. However, in order to improve the understanding of the relevant features driving the particle motion in the presence of non-linear forces and also to reduce the tracking time, there is a strong motivation to investigate different analytical methods dealing with amplitude limitations of stable motion.

The goal of such studies is to obtain closed expressions either for invariant distortions or amplitude limits up to which the motion remains bounded. These methods thus do not give a direct estimation of the dynamic aperture. But they do provide information on the amplitude at which the phase space topology gets highly complicated, going to chaotic regime i.e. sensitively dependent on the initial conditions, or eventually to unbounded motion.

The intention of this lecture is not to present an overview of all the analytical methods recently developed but to give an insight of a few of them. Most of the methods described below are based on Hamiltonian formalism and the related equations of motion. They approach the limit of stability from below, i.e. from the side where the motion is still stable because the initial amplitudes are not yet too large. They use perturbation techniques of some kind in order to analyse the effects of non-linearities. Arbitrarily we will not discuss non-perturbative methods or attempt to approach the limit of stability from above.

These methods are generally restricted to studying low orders of the perturbing field strength, because the analytical developments (series, linearizations, iterations) become tedious after a few order calculations. This is one of the major limitations of these theories because, in most present accelerator designs, it is likely that such terms are large. This is the case in superconducting colliders with strong non-linear behaviour of magnets. In electron colliders or synchrotron light sources, this is due to strong chromaticity sextupoles used to compensate the chromatic aberrations generated by high gradient quadrupoles. Therefore the study of particle motion may require retaining higher-order terms.

Another limiting factor comes from the fact that, in many cases, the generalization to magnetic non-linear elements other than sextupoles does not yet exist, or the extension to two-dimensional systems remains to be done. Also it has to be mentioned that the convergence of the perturbation procedures might be doubtful.

### 3.2 Resonance approach

This analysis corresponds to the situation in which a single resonance can be isolated. In that case, the particle tune is near a rational number and a driving term acts on the particle with a frequency near the natural frequency of the particle motion. This driving term increases the amplitude of the particle and possibly leads to unstable motion.

The description of this resonant behaviour is based on Hamilton's equations. The general treatment is performed for a two-dimensional motion. However, in order to illustrate the approach used to estimate the limit of particle stability, let us take a practical example of one-dimensional point sextupole driven resonance. In action-angle variables, the Hamiltonian of the motion can be written as:

$$H(I, \phi, \theta) = v I + \frac{1}{3} a_2 v^{3/2} \beta \delta(\theta) \left( \frac{2I}{v} \right)^{3/2} \cos^3 \phi \quad (4)$$

The next step consists in Fourier transforming the driving term and ignoring the rapidly varying terms which do not drive a third-order resonance. One gets:

$$H(I, \phi, \theta) = v I + \frac{1}{24\pi} a_2 v \beta^{3/2} \left( \frac{2I}{v} \right)^{3/2} \cos(3\phi - m\theta) \quad (5)$$

Then, in order to make the Hamiltonian independent of  $\theta$  and obtain an integral of the motion, we make the following canonical transformation:

$$\begin{aligned} \phi &\rightarrow \psi = \phi - \frac{m}{3}\theta \\ I &\rightarrow J = I \end{aligned}$$

The Hamiltonian in these variables is explicitly  $\theta$  independent:

$$H(J, \psi) = (v - \frac{m}{3})J + \frac{1}{24\pi} a_2 v \beta^{3/2} \left( \frac{2J}{v} \right)^{3/2} \cos 3\psi \quad (6)$$

Constant  $H$  contours in  $(J, \psi)$  space will characterize particle motion. In order to have a more concrete representation of particle trajectories, it is convenient to return to the  $(x, p_x)$  phase space by a series of inverse transformations and to plot level curves of the function  $H$ . An example of a phase space plot is given in Fig. 2. One can easily identify the phase space distortions growing with amplitude or increased driving strength, as well as the stability

limits. The curves are identical to the ones observed through numerical analysis.

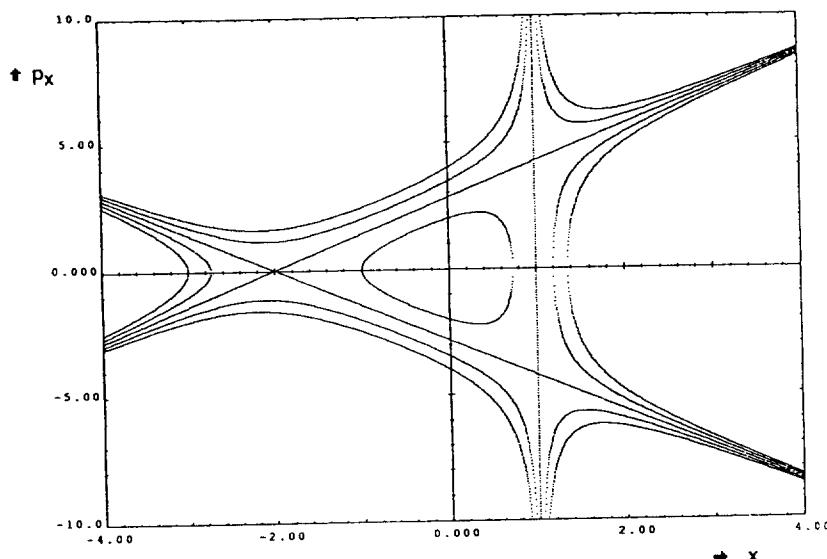


Fig. 2 Level curves of invariant in  $x, p_x$  phase space

Additional calculations allow to derive information on some figures of merit such as tune shifts with amplitude or stop bands.

Opposite to the situation of single resonances, one can consider the case where the tunes  $v_x, v_y$  are such that no strong resonance dominates and that the working point can only be close to weak resonances. Therefore all non-linear driving terms in the equation of motion contribute to the motion. The non-linear behaviour can be studied using various perturbation theory tools in order to evaluate the distortions of the invariant curves as well as the amplitude modulations within certain limits of validity. Some of them are described below.

### 3.3 Phase space distortion

#### 3.3.1 Hamiltonian calculations

This method accounts for the distortion of the phase space due to non-linear field effects. It is based on a standard canonical perturbation technique and Moser transformations [2] in order to develop an expression for a phase space invariant. The treatment starts with the Hamiltonian describing the betatron oscillations with non-linear magnetic fields given by Eq. (3). For simplicity we shall drop the  $x$  and  $y$  index in the following developments, whenever possible.  $\mathbf{l}$  and  $\phi$  are therefore two-dimensional vectors.

The Hamiltonian is not an invariant because it is a function of  $\theta$ . The method consists in making a canonical transformation which makes the new Hamiltonian, called  $G$ , independent of  $\theta$  to first-order. By iterating the process with successive canonical transformations, one can push the  $\theta$ -dependence to sufficiently high order that it may be neglected. The treatment is

given in detail in [3].

The first step canonical transformation from  $(I, \phi)$  to  $(J, \psi)$  can be defined by a generating function of the form :

$$F(\phi, J, \theta) = \phi J + F_1(\phi, J, \theta) \quad (7)$$

with the following relations between the old and new variables:

$$I = \frac{\partial F}{\partial \phi} = J + \frac{\partial F_1}{\partial \phi} \quad ; \quad \psi = \frac{\partial F}{\partial J} = \phi + \frac{\partial F_1}{\partial J} \quad (8)$$

We can now write the new Hamiltonian as:

$$G(\psi, J, \theta) = H(I, \phi, \theta) + \frac{\partial F(\phi, J, \theta)}{\partial \theta} \quad (9)$$

In order to express both  $H$  and  $G$  as functions of the  $F$  variables, we may use Taylor's expansions and obtain (putting aside the time dependence):

$$\begin{aligned} G(\psi, J, \theta) &= G(\phi, J, \theta) + \frac{\partial G}{\partial \psi} \frac{\partial F_1}{\partial J} + \dots \\ &= H_0(J) + v \frac{\partial F_1}{\partial \phi} + H_1(J, \phi) + \frac{\partial F_1}{\partial \theta} + \frac{\partial H_1}{\partial I} \frac{\partial F_1}{\partial \phi} + \dots \end{aligned} \quad (10)$$

Since the perturbation appears in the functions  $G$ ,  $F_1$ ,  $H_1$ , the product of two partial derivatives are of second-order in the perturbation and only the four first terms in Eq. (10) are first-order terms. Taking advantage of the periodicity of all the functions in  $\phi$  and  $\psi$ , it is possible to analyse them in Fourier series with respect to these variables. Therefore Eq. (10) can be replaced by a series of relations between the harmonic coefficients noted  $g_{np}$ ,  $f_{np}$ ,  $h_{np}$ . To first-order in the perturbation, one gets:

$$g_{np} = h_{np} + i(nv - p)f_{np} \quad (11)$$

and the first-order solution for the function  $F_1$  is:

$$f_{np}^{(1)} = i \frac{h_{np} - g_{np}}{nv - p} \quad (12)$$

If the tune is rational, the method will reduce to the usual single resonance treatment. We consider here only the case where  $v$  is irrational, which means that the denominator of Eq. (12) is different from 0. To make the new Hamiltonian  $G$  independent of  $\theta$  to first-order, we can

choose  $g_{np} = 0$  and push the  $\theta$  dependence to higher order, since  $G$  contains only higher order in the perturbation. If we add now the second-order terms in the  $G$  equation, we obtain the second-order solution of  $F_1$ :

$$f_{np}^{(2)} = f_{np}^{(1)} + \sum_{n',p'} (n - n') \frac{\partial h_{n'p'}}{\partial I} \frac{f_{n-n', p-p'}}{n v - p} \quad (13)$$

This equation can be solved by iteration, putting the first-order  $f_{np}$  on the right hand side. The generating function  $F_1$  can then be written to second-order as:

$$F_1 = \sum n f_{np}^{(2)} J_x^{(l+k)/2} J_y^{(l+m)/2} \exp [i n (\phi + v \theta) - i p \theta] \quad (14)$$

The relations between  $(I, \phi)$  and  $(J, \psi)$  via  $F_1$  give the distortions of the phase space trajectories. In particular, the distortions of the invariants  $I$  are expressed as:

$$I = J - \sum n f_{np}^{(2)} J_x^{(l+k)/2} J_y^{(l+m)/2} \exp [i n (\phi + v \theta) - i p \theta] \quad (15)$$

At this stage, the new Hamiltonian  $G$  which contains only second-order terms is still  $\theta$ -dependent, but this residual dependence can be suppressed by averaging over  $\theta$ , i.e. keeping only low frequency terms. The Hamiltonian then becomes:

$$\bar{G} = v_x J_x + v_y J_y + \sum_{jk} D_{jk} J_x^{j/2} J_y^{k/2} \quad (16)$$

This Hamiltonian is now a constant of the motion where the two first terms correspond to the ellipses associated with the linear motion, while the sum gives the ellipse distortion due to non-linearities.

The validity of the description implies that  $I_{x,y} \geq 0$ , i.e.  $\Delta I \leq J$  by virtue of Eq. (15). This means that the method fails because the action variable becomes negative, which is not compatible with the definition of the action. The treatment requires iterations in order to improve the accuracy of the determination of the distorted invariants and the amplitude modulation. Since third and higher-order terms of the Hamiltonian have been neglected, this implies that the perturbation is small enough. Consequently, this approach has difficulties in describing the motion close to the stability limit where the perturbation is strong. The procedure has to be applied to higher orders. In practice, however, it is difficult to go above the second-order since the analytical treatment becomes more and more tedious.

This method has been applied to study the emittance distortions in the presence of sextupoles and compared with direct tracking calculations. In the KEK case [4], a very good agreement was found between both methods for rather weak sextupoles but large deviations appear when dealing with strong sextupoles. A similar analysis has been performed for the Tevatron lattice with one sextupole [5]. The distortions of the invariants  $I$  have been calculated,

using the expression (15) and compared with tracking. As shown in Fig. 3, there is a reasonable agreement between analytical and numerical results, provided the distortions are not too large.

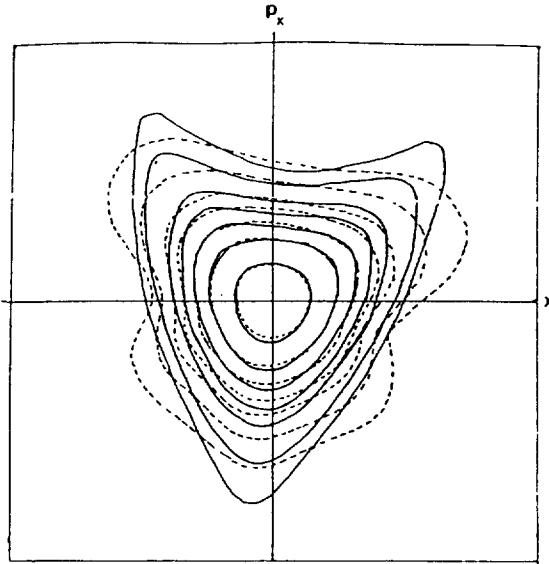


Fig. 3 Invariant distortions with sextupoles for the Tevatron lattice

### 3.3.2 Calculations from the equations of motion

Another method dealing with phase space distortions due to sextupoles and calculations from the equations of motion has been developed [6]. It starts from the statement that the phase space ellipses corresponding to the Courant-Snyder invariants will be distorted by nonlinearities. The method therefore proceeds by transporting the linear invariant through the lattice and kicking each point on the invariant at each sextupole. The non-linear kicks are calculated under the assumption that the motion is purely linear and that the propagation from the other sextupoles does not affect these kicks. This means that only terms linear in sextupole strengths are considered. At the end of one turn, the distorted invariant is compared to the linear one. The distortion is characterized by a set of "distortion functions" which can be expressed in terms of the sextupole strengths and the linear lattice functions.

These distortion functions can be interpreted by analogy with the linear lattice functions and one can construct an invariant phase space which remains unchanged after one turn if only first-order terms in sextupole strengths are used in the transfer. If second-order terms are used, the additional distortion can be attributed to a set of "second-order distortion functions". In this way, the distortions may be calculated order by order in the sextupole strengths. This method does not provide an absolute calculation of dynamic aperture. However, the analytical expressions of the distortion functions provide information on the sources of distortions and can be used in minimization procedures.

### 3.4 Use of Lie transforms

Several procedures employing Lie transformations have been developed [7], [8]. We shall sketch one of them which is based on the Deprit's algorithm [9]. Let us assume that the full Hamiltonian depends on a control parameter  $\epsilon$  and admits a power series expansion about  $\epsilon = 0$ , as well as the generating function  $F$  and the new Hamiltonian  $G$ .

$$H = \sum_{n=0}^{\infty} \frac{\epsilon^n}{n!} H_n \quad F = \sum_{n=0}^{\infty} \frac{\epsilon^n}{n!} F_{n+1} \quad G = \sum_{n=0}^{\infty} \frac{\epsilon^n}{n!} G_n \quad (17)$$

The functions  $H_n$  are known; it is required to find all  $G_n$  and  $F_n$  order by order. This is done by using a recursive algorithm based on an operator  $D$  which is defined as the total derivative along the direction of the local  $H_0$ -orbits.  $D$  acts on a function  $f$  in the following way:

$$D f = \frac{\partial f}{\partial \theta} + \{f, H_0\} \quad (18)$$

where  $\{ , \}$  is the Poisson bracket.

The recursion gives the following equations for the lower orders:

0 <sup>th</sup> order	$G_0 = H_0$
1 <sup>st</sup> order	$D F_1 + G_1 = H_1$
2 <sup>nd</sup> order	$D F_2 + G_2 = H_2 + \{H_1, F_1\} + \{G_1, F_1\}$

(19)

At each order, there are two unknown functions  $G_n$  and  $F_n$ .  $G_n$  must be chosen to cancel the average term on the r.h.s. of Eq. (19), because the presence of such a term would make  $F_n$  unbounded. Now, to find the functions  $F_n$ , we must solve partial differential equations of the form:

$$D F_n = (\text{r.h.s.})_n - G_n \quad (20)$$

Solutions of Eq. (20) can be obtained by integrating the inhomogeneous term along  $H_0$ -orbits. Another way is to use the eigenfunctions of the operator  $D$  as a basis to expand and solve the differential equations (20) algebraically.

The procedure can be applied conveniently to high orders. For instance the Lie transform approach has been used to study the effects of sextupoles and octupoles with arbitrary harmonics. The following example of zeroth harmonic sextupoles and octupoles [9] illustrates the procedure in a concrete way. In that case, the Hamiltonian can be expressed as:

$$H = v l + \epsilon l^{3/2} \sin^3 \phi + \frac{1}{2} \epsilon^2 \kappa l^2 \sin^4 \phi \quad (21)$$

The first-order equation to be solved is:

$$\left( \frac{\partial}{\partial \theta} + v \frac{\partial}{\partial \phi} \right) F_1 + G_1 = l^{3/2} \sin^3 \phi \quad (22)$$

First of all, since the r.h.s has a non-zero average, we must choose  $G_1 = 0$ . Then, by using trajectory integration, we write the solution for  $F_1$  as:

$$F_1 = \left( \frac{l^{3/2}}{v} \right) \left( -\frac{3}{4} \cos \phi + \frac{1}{12} \cos 3\phi \right) \quad (23)$$

Because  $G_1 = 0$ , the second-order equation is simplified to:

$$\left( \frac{\partial}{\partial \theta} + v \frac{\partial}{\partial \phi} \right) F_2 + G_2 = H_2 + \{ H_1, F_1 \} \quad (24)$$

The Poisson bracket is easily evaluated and  $G_2$  is chosen to cancel the average term on the r.h.s. of Eq. (24). The solution for  $F_2$  is then deduced:

$$F_2 = l^2 \left[ \frac{1}{4v} \left( \frac{3}{2v} - \kappa \right) \cos 2\phi + \frac{1}{32v} \left( \frac{3}{2v} + \kappa \right) \sin 4\phi \right] \quad (25)$$

Pushing to higher orders is tedious. However, the calculations are straightforward and one can derive the averaged Hamiltonian to fourth-order. The interest of the method is to give explicit expressions for the generating function and the orbit. The question of convergence is the main drawback of this method.

### 3. 5 Successive linearization method

Instead of using the Hamiltonian and canonical transformations in order to obtain a new Hamiltonian independent of  $\theta$ , the perturbation treatment can be performed by iterations on the equations of motion directly and successive linearizations [10]. Up to now, the treatment is restricted to the case of structures containing only sextupoles as non-linear elements. Also the extension to two-dimensional systems remains to be done. Under these restrictions, the method is interesting since it converges rapidly at each iteration which contains contributions of all orders in the perturbation.

Let us write the non-linear horizontal betatron motion in the presence of sextupoles:

$$x'' + k(s)x = \frac{1}{2} m(s)x^2 \quad (26)$$

Practical applications imply the sextupole strength  $m(s)$  small enough such that the r.h.s. perturbation term is small compared to the linear part. An approximate solution is obtained by a first linearization:

$$x''^{(0)} + k(s) x^{(0)} = 0 \quad (27)$$

The solution of Eq. (27) which describes the betatron motion in the presence of linear fields is well known. Then, one writes the complete solution as  $x(s) = x^{(0)}(s) + u(s)$ . The corresponding equation for  $u(s)$  can be expressed as :

$$u'' + [k(s) - m(s) x^{(0)}] u = \frac{1}{2} m(s) [x^{(0)}^2 + u^2] \quad (28)$$

Dropping again the quadratic term in  $u$  yields the second linearization:

$$u''^{(0)} + [k(s) - m(s) x^{(0)}] u^{(0)} = \frac{1}{2} m(s) x^{(0)}^2 \quad (29)$$

This linear equation contains the sextupole strength and the approximate solution depends on it. After two linearizations, the approximation of  $x$  then becomes  $x(s) = x^{(0)}(s) + u^{(0)}(s)$  and the stability of motion implies that  $u^{(0)}$  is bounded. The linearization process can be continued to higher level, giving a more and more precise solution of the initial equation, even if the analytical developments become cumbersome.

Let us consider Eq. (29). The term proportional to  $m(s) x^{(0)}$  on the l.h.s. corresponds to a focusing force which depends on the amplitude and may induce parametric resonances. Above a certain value of  $x^{(0)}$ , the non-linear term of Eq. (29) dominates and a rapid self-amplification will take place, leading to unbounded motion. In real systems, however, the instability is observed at small amplitudes already, before the non-linear term becomes dominant, so that the perturbation treatment is applicable. Hence the complete information about the stability limit is, in general, contained in the homogeneous equation:

$$u''^{(0)} + [k(s) - m(s) x^{(0)}] u^{(0)} = 0 \quad (30)$$

In general  $x^{(0)}$  is not periodic over one period. Equation (30) can only be reduced to a vector recurrence with a non-constant transfer matrix.

$$\begin{pmatrix} u^{(0)} \\ u'^{(0)} \end{pmatrix}_{n+1} = M_n \begin{pmatrix} u^{(0)} \\ u'^{(0)} \end{pmatrix}_n \quad (31)$$

However, if the tune  $\nu$  of the structure is rational, i.e.  $\nu$  equal to  $p/q$ ,  $p$  and  $q$  being integers,  $x^{(0)}$  becomes periodic over  $q$  periods and Eq. (30) becomes of Hill's type again with an associated transfer matrix  $R$  over the new period. One can then apply the linear theory which tells us that the solution  $u^{(0)}$  will be bounded if the condition  $| \text{Tr}(R) | \leq 2$  is fulfilled. Since the coefficients of  $R$  are polynomials in the initial values  $x_0$  and  $x'_0$ , the condition can be used for a direct estimation of the stability limit.

However, for complicated structures and large  $q$ , it becomes tedious to get closed expressions for the matrix product  $\mathcal{R}$ . Nevertheless, evaluating the product of a reasonable number of matrices offers some advantage with respect to tracking of many particles over a large number of periods.

This method has been applied to the LEP lattice [10]. Applying the condition  $|\text{Tr}(\mathcal{R})| \leq 2$  leads to a curve  $x_0 \text{ lim} = f(v)$ , with  $x_0 \text{ lim}$  being the maximum initial value for which  $u^{(0)}$  is still bounded. Fig. 4 shows the comparison between the complete resulting curve and tracking results. The agreement is good in a large range of  $v$  values.

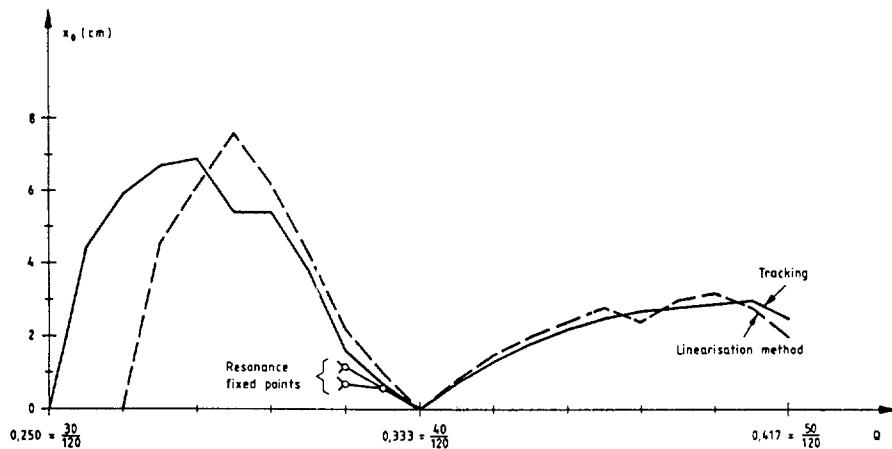


Fig. 4 Stability limit for the LEP lattice after two linearizations

### 3.6 Secular perturbation

This method [11] aims at deriving an analytical approximate expression for the dynamic aperture, using a classical secular perturbation theory (or Lindsted-Poincaré technique). It is based on the expansion of the solution of motion as a power series of the non-linear strength parameter and removing secular terms in each order of perturbation by selecting frequencies appropriately.

In this lecture, the discussion will be restricted to non-linear effects generated by sextupoles. However the treatment may easily be extended to any order multipole. In the case of pure horizontal motion, the equation of motion is given by:

$$\frac{d^2x}{d\theta^2} + v_0^2 x + \epsilon f(\theta) x^2 = 0 \quad , \quad f(\theta) = -m(\theta) v_0^2 \beta_x^{5/2} \quad (32)$$

where the parameter  $\epsilon$ , which is equal to unity in the above equation, scales the strength of the perturbation.

The fundamental idea of secular perturbation theory is based on the fact that one of the effects of non-linear terms is to shift the frequency of the system from the linear value  $\nu_0$  to  $\nu$ . This leads to expand both  $x$  and the tune  $\nu_0$  as power series in the control parameter  $\epsilon$ :

$$\begin{aligned} x &= x_0 + \epsilon x_1 + \epsilon^2 x_2 + \dots + \epsilon^n x_n \\ \omega &= 1 + \epsilon \omega_1 + \epsilon^2 \omega_2 + \dots + \epsilon^n \omega_n \\ \nu &= \nu_0 \omega \end{aligned} \tag{33}$$

The first step consists in transforming the independent variable as  $\Theta = \theta \omega$ , so that Eq. (32) can be rewritten:

$$\omega^2 x'' + \nu_0^2 x + \epsilon f(\Theta) x^2 = 0 \tag{34}$$

where the " denotes the derivative with respect to  $\Theta$ .

Now, we insert the perturbation expansion (33) into Eq. (34) and find a general expression for the  $n$ -th order perturbation contribution:

$$x''_n + \nu_0^2 x_n = -f(\Theta) \sum_{j=1}^{n-1} x_j x_{n-1-j} - \sum_{j=0}^{n-1} x''_j \Omega_{n-j} \tag{35}$$

where the  $\Omega_n$  are defined as:  $\Omega_n = \sum_{i=0}^n \omega_i \omega_{n-i}$

We obtain  $n$  linear equations which can be solved successively. Each of these equations contains one free parameter  $\omega_n$  for the non-linear frequency correction. These corrections are chosen such as to eliminate the secular terms in any of the  $n$  equations. Secular terms are defined as trigonometric contributions to the r.h.s. of Eq. (35) proportional to  $\cos \nu \theta$  and  $\sin \nu \theta$ . Generally, one can demonstrate that  $x_n$  may be represented as:

$$x_n = A \cos \nu \Theta + \sum_{n=1}^{\infty} \sum_{l=0}^{\infty} c_l^n q_l(\Theta) \tag{36}$$

where the  $c_l^n$  are constant coefficients and the  $q_l$  are a set of basis functions on which we can expand the particular solution of Eq. (35). Using Eq. (33) and (36), we then find the exact solution for  $x(\Theta)$ :

$$x(\Theta) = A \cos \nu \Theta + q_0(\Theta) \sum_{n=1}^{\infty} c_0^n + q_1(\Theta) \sum_{n=1}^{\infty} c_1^n + q_2(\Theta) \sum_{n=1}^{\infty} c_2^n + \dots \tag{37}$$

Since the basis functions are linearly independent, divergence of one of the series is a sufficient condition for the secular perturbation theory to break down. We determine the

approximate convergence limit of the series (37) by using d'Alembert's criterion:

$$\lim_{n \rightarrow \infty} \left| \frac{c_i^{n+1}}{c_i^n} \right| \leq 1 \quad (38)$$

This model has been applied to the LEP lattice with sextupoles [11], restricting the treatment to second-order theory and comparing the results of the analytical approach with tracking. In Fig. 5, the maximum stable initial amplitudes are plotted as a function of horizontal tune in both cases. The agreement with tracking is relatively good. Clearly the effects of resonances up to fourth-order are visible. The fifth-order resonance indicated by tracking is not covered by second-order perturbation theory.

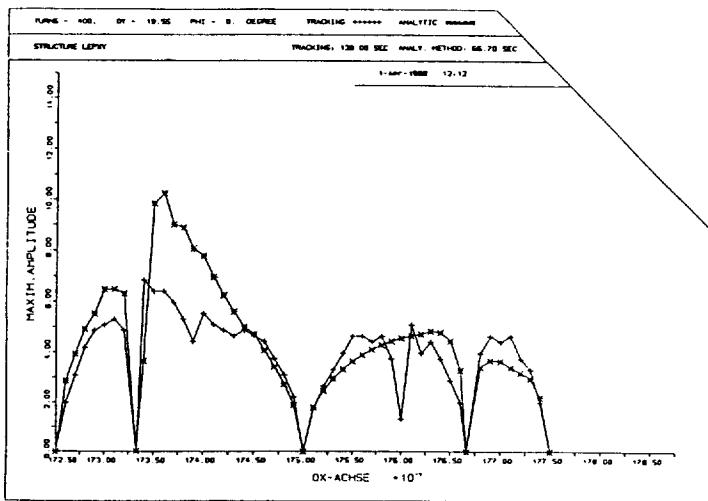


Fig. 5 Maximum stable initial horizontal amplitude for LEP

#### 4. COMPENSATION PROCEDURES

Analytical methods bring a better understanding of the mechanisms which govern the motion of particles in the presence of non-linear fields. However the practical problem addressed to accelerator designers is to design correction schemes capable of limiting the perturbation of motion. It is therefore important to define procedures in order to compensate non-linear effects and to estimate the quality of the correction by looking at the dynamic aperture. The object of this Section is to describe these intermediate procedures between the pure analytical or numerical approaches.

The strategy consists in calculating some physical quantities which characterize the perturbation induced by the non-linear fields and to minimize them in order to enlarge the dynamic aperture. Generally, the relevant figures of merit are resonance strengths, non-linear variations of tune and betatron functions with momentum, tune shifts with amplitude driven by sextupoles in second-order, amplitude distortions. For instance the program HARMON [12]

exploits such calculations to optimize the distribution of sextupoles.

Depending on the type of machine under consideration, different compensation schemes can be envisaged (self-compensation in large electron colliders, single resonance compensation, broad-band compensation).

Large colliders are generally built up from regular FODO cells. In addition to the chromaticity correction, the compensation of the modulation of the  $\beta$  functions for off-momentum particles is a concern because the  $\beta$ -beat is responsible for a non-linear tune shift with momentum. For these reasons, additional sextupoles must be introduced. The goal of the sextupole compensation scheme is to compensate the higher-order chromatic effects without creating large additional non-linear resonances.

The solution generally adopted is to use interleaved sextupole schemes providing an intrinsic cancellation of the driving terms of non-linear resonances. This is achieved if the phase advance per cell is rational and if the lattice is built of supercells composed of two identical subsections, each having a phase advance of  $n\pi$ . Figure 6 gives an illustration of this strategy in the case of the LEP lattice [10].

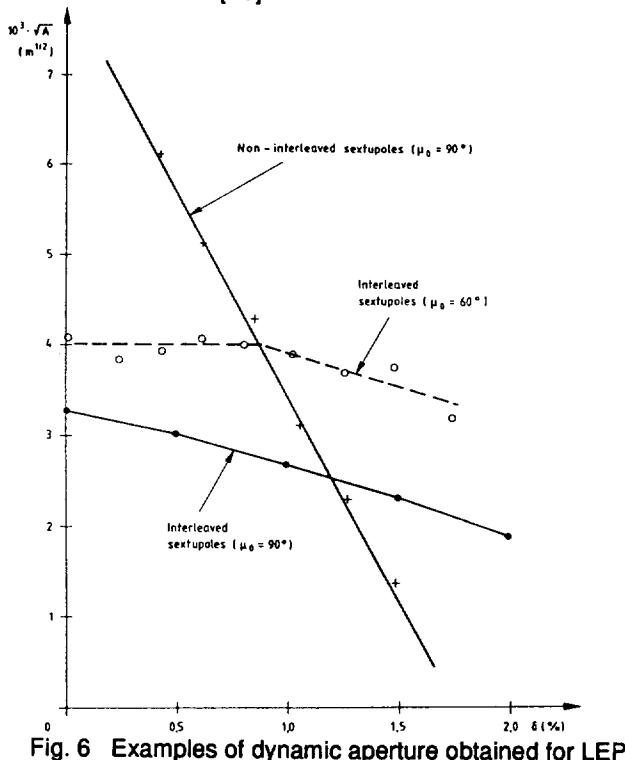


Fig. 6 Examples of dynamic aperture obtained for LEP

The concept of intrinsically resonance-compensated schemes cannot generally be applied to smaller machines which have more complex and irregular lattices than colliders. Nevertheless, correction procedures capable of compensating a large band of harmonics driving non-linear resonances have been optimised. The compensation of distortion functions is discussed in detail in another lecture [13].

An alternative procedure [14] consists in compensating simultaneously the driving terms of the most harmful harmonics and the non-linear tune shifts with amplitude which can be derived from the standard perturbation theory:

$$\begin{aligned}\Delta v_x &= A J_x + B J_y \\ \Delta v_y &= B J_x + C J_y\end{aligned}\quad (39)$$

The coefficients A, B, C are expressed in the harmonic expansion:

$$\begin{aligned}A &= -\frac{3}{4} \left\{ \sum_m \frac{|D_m|^2}{v_x - m} + 3 \sum_n \frac{|F_n|^2}{3v_x - n} \right\} \\ B &= \frac{1}{2} \left\{ \sum_q \frac{|H_q|^2}{q v_x - 2v_y - q} - \sum_r \frac{|L_r|^2}{r v_x + 2v_y - r} - \sum_m \frac{|D_m| |G_m|}{v_x - m} \cos(d_m - g_m) \right\} \\ C &= -\frac{1}{4} \left\{ \sum_q \frac{|H_q|^2}{q v_x - 2v_y - q} + \sum_r \frac{|L_r|^2}{r v_x + 2v_y - r} + \sum_p \frac{|G_p|^2}{v_x - p} \right\}\end{aligned}\quad (40)$$

where the resonant harmonic coefficients  $D_m$ ,  $F_n$ ,  $G_p$ ,  $H_q$ ,  $L_r$  driving the sextupole resonances are the summation of products of sextupole strengths and linear lattice functions

This strategy has been successfully applied to the ESRF lattice, using the minimization procedure implemented in the code BETA [15], to compensate simultaneously the driving terms  $D_m$ ,  $F_n$ ,  $G_p$ ,  $H_q$ ,  $L_r$  and the tune shifts with amplitude. This is demonstrated in Fig. 7 which shows a comparison between the dynamic aperture of the original lattice and the present design.

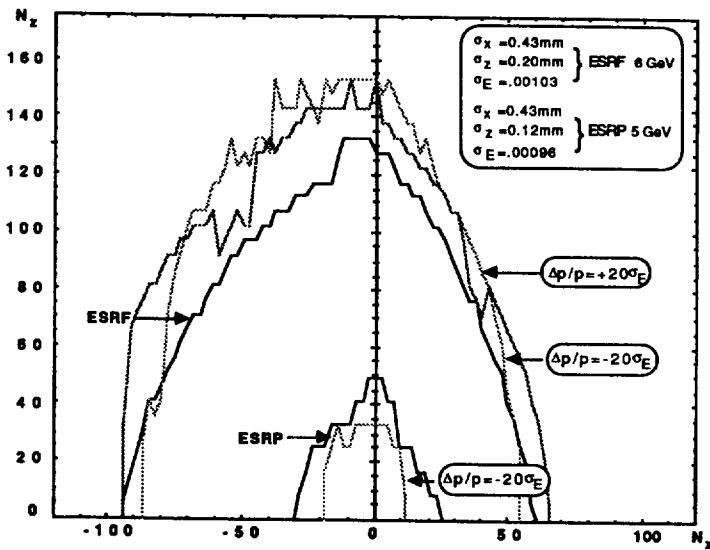


Fig. 7 Dynamic aperture of the ESRF

## 5. NUMERICAL APPROACH

### 5.1 Particle tracking and codes

Generally, analytical tools are restricted to relatively simplified cases. Therefore, in parallel with analytical results, we need numerical modelling of the machine that will include high orders of all types of perturbations and errors of all kinds. This modelling is performed through the use of numerical methods known as tracking. The goal of tracking is to simulate directly single-particle motion with all relevant effects included in the model and to determine the regions of phase space that are stable.

The process used in tracking simulations is as follows. A test particle is launched at a convenient location along the ring with selected initial conditions and subject to the transfer map which describes the particle motion. After one turn, the position of the particle is recorded and used as initial conditions for a new application of the transfer map. By iterating the process for a variety of test initial conditions over many turns, one can empirically determine the stability of various regions of phase space. The analysis of the results leads to defining a maximum stable amplitude which is taken to be the dynamic aperture.

Various codes have been developed to perform the numerical iterations required for tracking simulations. They mainly differ by the methods used for representing the non-linear transfer map and by the dimension of the mapping which can vary from two to six according to the effects which are included (pure x-y motion, coupled betatron x-y motion, coupled synchro-betatron motion). Some of the methods used in the most popular codes will be briefly discussed.

a) The simplest method employed to describe the transfer map is the thin lens model or "kick" approximation. Linear transformation matrices are used to represent the linear elements such as dipoles, quadrupoles or drifts. Magnetic elements with non-linear fields are treated in the impulse approximation. It is assumed that the transverse coordinates of the particle remain unchanged and the transformation is represented as a function of the initial conditions by the equations:

$$\begin{aligned} q_f &= q_i \\ p_f &= p_i + f(q_i) \end{aligned} \tag{41}$$

where  $q$  stands for  $x$  or  $y$  and  $p$  is the conjugate momentum of  $q$ .

The entire accelerator is then modelled by a sequence of interleaved linear matrices and non-linear impulse transformations. Energy oscillations can also be treated by using a six-dimensional phase space and simulating the effect of the r.f. system with transformations of the same form as Eq. (41).

The thin lens approximation is valid, provided the non-linear elements under consideration are short and have high fields. The model can describe all orders of perturbations. The transformations are easily evaluated. However, a full turn may require large computer time if

many non-linear elements are present in the lattice. However this model is ideal for treating many examples of accelerator designs. It is employed by a large number of programs such as BETA [15], PATRICIA [16], RACETRACK [17], TEAPOT [18].

b) Higher order matrix methods are used by the programs MAD [19] and DIMAD [20]. In these methods, each component of the six-dimensional vector specifying the phase space is considered as a function of all components, when entering an element. This function is written as a Taylor's expansion and then truncated at some order (generally second-order). Thus the results may be written for any  $i = 1, 2, \dots, 6$ :

$$u_i = \sum_k M_{ij} u_j + \sum_{j,k} T_{ijk} u_j u_k \quad (42)$$

The matrix elements  $M_{ij}$  and  $T_{ijk}$  are obtained directly from the equations of motion. However Eq. (42) is a truncated series and so it is not a canonical transformation. This drawback may be overcome by using the generating function technique described below. The method is limited to second or third-order but there is no restriction on the length or strength of the non-linear elements. Since a specific series is provided for each element, the method has the advantage that several transformations can be combined so that a group of elements can be described by a single transformation ("concatenation"), thus leading to an increase in tracking speed.

c) Lie transformation techniques represent the transfer map in terms of Lie transformations which are operators directly generated from Hamilton's equations. The formalism makes use of the mathematic tools of the Lie algebra. It starts with the definition of the Lie operator denoted by  $f$ . Hamilton's equations can be written in terms of the Lie operator generated by the Hamiltonian in the following way:

$$\begin{aligned} \frac{dq_i}{dt} &= \frac{\partial H}{\partial p_i} = -\{H, q_i\} = -:H:q_i \\ \frac{dp_i}{dt} &= -\frac{\partial H}{\partial q_i} = -\{H, p_i\} = -:H:p_i \end{aligned} \quad (43)$$

These equations can be integrated to obtain the solution for the motion of the particle:

$$u_i(t) = e^{-t:H} u_i(0) \quad (44)$$

where  $u_i$  (with  $i = 1, \dots, 6$ ) is any of the phase space coordinates and the exponential operator is a "Lie transformation". The relationship (44) represents the exact canonical transformation describing the motion in a single element. By using the properties of Lie algebra, it is possible to combine several Lie transformations to produce a single transformation describing the transfer map through a collection of elements. One of the advantages of the method is that one can construct a canonical transformation which represents the transfer through a finite length

element accurately to high order. However, in practice, the description is limited to third or fourth-order. One of the tracking codes making use of the Lie algebra formalism is the program MARILYE [21]. This formalism has also been implemented in MAD.

d) Generating function methods are based on the fact that the transfer map in an accelerator can be represented in terms of generating functions  $F$ . The transformation has the general following form:

$$q_f = \frac{\partial F (q_i, p_i)}{\partial p_f} ; \quad p_i = \frac{\partial F (q_i, p_i)}{\partial q_f} \quad (45)$$

Eq. (45) is canonical in all orders. However the method is restricted to the description of low order non-linearities (3 or 4) at present. Unfortunately the equations are implicit;  $q_f$  and  $p_f$  can only be solved numerically in most cases. This method is being used as an auxiliary calculation tool in programs based on other techniques, such as MAD, MARILYE or DIMAD.

e) The equations of motion can be solved by numerical integration using procedures like Runge-Kutta algorithms. In most of the procedures the integration algorithm is not canonical, which is a major drawback for the tracking process since it may lead to unphysical damping or growth of the phase space and overestimation or underestimation of the dynamic aperture. A symplectic algorithm has been proposed [22] which preserves Poisson brackets and makes each canonical step a canonical transformation. During the transformation in one element, both the position and the momentum are modified by terms of high order in the element length. To improve the accuracy, the integration step can be subdivided in each element. This is, of course, at the cost of increased computer time.

In addition to modelling the effects of chromaticity sextupoles, a general tracking program must incorporate all non-linear effects which potentially limit the dynamic aperture. Most of the codes listed above provide these additional features. This means:

i) any other multipole term. This concerns mainly field errors in the magnets. They can be either systematic errors which are the same from magnet to magnet and are due to the design of coils, finite shape of the magnet..., or random errors due to construction errors. The procedure is the same as for simulating the effects of sextupoles.

ii) closed orbit errors. The investigation of their effects and also their correction is important because generally the small apertures of the magnets impose severe limitations on the allowable orbit distortions and thereby the available dynamic aperture. Also in machines like the new generation of synchrotron light sources, the sensitivity of the lattices to closed orbit errors is such that zero (or too small) dynamic apertures may be generated and that the optimization of adequate corrections is a critical process.

## 5.2 Limitations on tracking codes

Although tracking simulations can provide very useful information, attention must be paid to several problems. One of the major limitations is due to the computing power and the time needed to execute tracking simulations. This constraint might be more or less severe, depending

on the type of accelerator under consideration. For instance electron storage rings are short term machines because of the radiation damping effect which introduces the time scale over which the motion is to be studied. Tracking over  $10^{2-4}$  turns is therefore relevant for these machines. On the other hand, proton storage rings or large colliders are long term machines which involve stable operation for minutes (at injection) or hours (in the collision mode). The time scale of interest is then  $10^{8-9}$  turns, which means that in these machines tracking studies are cpu time limited. In addition, the results of long term tracking and their extrapolation to longer time scales might be questionable.

The other limitation comes from the requirement that the motion should remain symplectic during simulations. Otherwise unphysical damping or growth of the phase space may be obtained, leading to a doubtful estimation of the dynamic aperture. This constraint implies, of course that the transformation used in the tracking code must be canonical even though the ultimate limitation is given by the number system used by the compiler and the rounding errors which make the solutions no longer symplectic.

### 5.3 An example of particle tracking

The example of a synchrotron radiation electron storage has been chosen to illustrate particle tracking features and the usefulness of tracking during machine design phase. The tracking studies have been performed for the ESRF lattice [23], using the program BETA. The machine is optimized to provide an extremely low emittance. It is characterized by a very strong focusing which results in high values of the chromaticity and strong correcting sextupoles. Consequently the motion is highly non-linear and yields a small dynamic aperture. This behaviour is illustrated in Fig. 8 which shows the horizontal phase space for particles launched with various initial conditions and tracked over 500 turns. Distortions from the linear motion are quite apparent, even at small amplitudes. The presence of islands close to the stability limit indicates the proximity of a high order resonance.

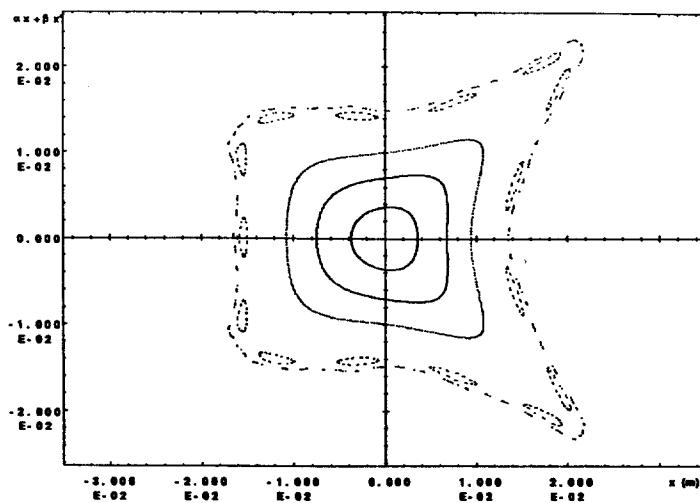


Fig. 8 Horizontal phase space

These aberrations can be compensated by using additional sextupoles located in the dispersion-free straight sections which therefore do not affect the chromatic properties of the lattice. These sextupoles were optimized according to the strategy discussed in Section 4. Although correcting sextupoles are the dominant source of dynamic aperture reduction, systematic and random field errors have also adverse effects on the dynamic aperture. This is shown in Fig. 9 where all realistic multipoles in dipoles have been simulated, using the kick approximation.

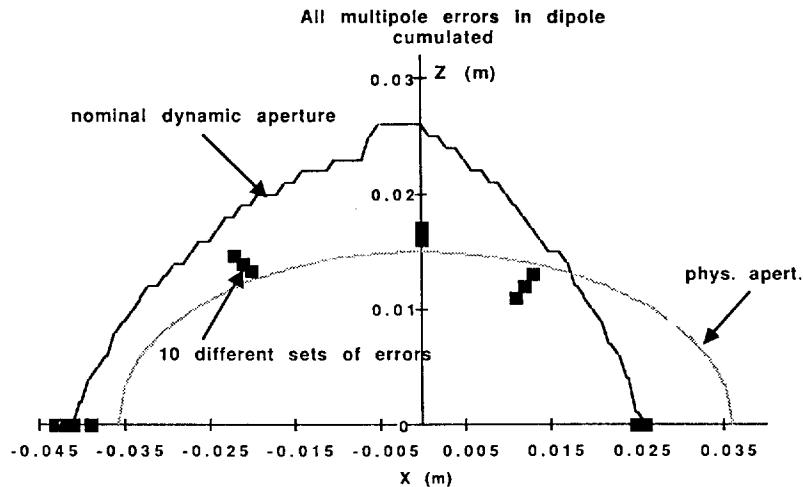


Fig. 9 Dynamic aperture with field errors in dipoles

As already outlined, closed orbit errors have a dramatic impact upon the dynamic aperture. Due to the strong focusing, the amplification by the lattice of quadrupole mispositioning is very large and results in a very severe degradation of the dynamic aperture as shown in Fig. 10.

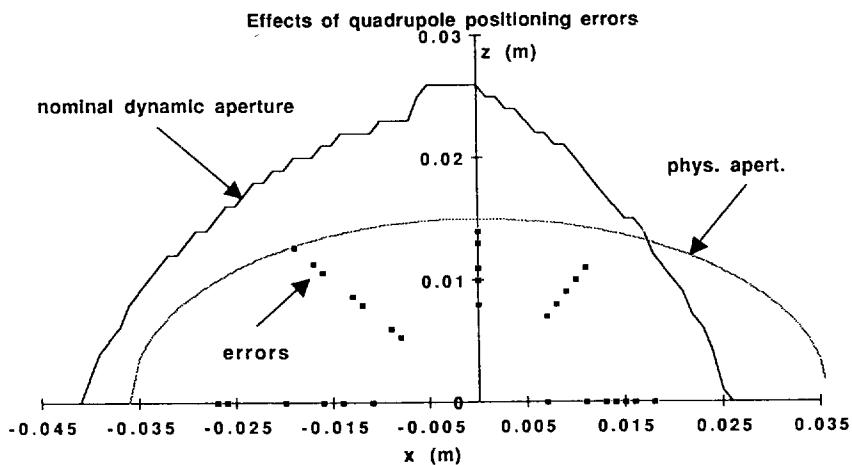


Fig. 10 Effects of quadrupole positioning errors on the dynamic aperture

## 6. EXPERIMENTAL RESULTS

Tracking studies are used extensively during the design phase of a machine and are generally complemented by theoretical developments. However, it is well known that a real machine is much more complicated than a mathematical model and that other phenomena like collective effects, beam-beam interaction, ion trapping, .... can affect performance. Therefore direct measurements of the dynamic aperture are needed to check the validity of tracking studies and understand the behaviour of existing machines as far as non-linearities are concerned. Also specific experiments are being devised in order to define criteria useful in designing new machines, like superconducting colliders. In that case, the operating conditions of these future machines are simulated by measuring the dynamic aperture in the presence of non-linearities artificially introduced in the machine.

The usual technique [24] to perform dynamic tracking of the beam in an accelerator consists in exciting coherent oscillations by firing a kicker magnet and recording at each revolution the transverse position of the beam on two monitors separated by  $(2k + 1) \pi/2$  in betatron phase. Phase space plots or Fourier spectra of the position signals can be obtained from these measurements.

For electron machines, a specific method can be used to measure the dynamic aperture. It is based on lifetime measurements as a function of variable apertures which are artificially created by introducing collimators in the physical aperture. As long as this aperture limitation lies outside the stable area of motion, the lifetime should be constant. At the point where it restricts the stable motion, a clear reduction of the lifetime should be recorded. This point corresponds to the dynamic aperture of the ring.

This technique has been applied to measure the dynamic aperture of BESSY [25]. Figure 11 shows the vertical dynamic aperture of the ring as a function of tune, determined from measurements (l.h.s. curve) and from tracking simulations (r.h.s. curve). The experimental results exhibit a strong tune dependent variation of the dynamic aperture. By introducing adequate non-linear fields in the simulations, most of the resonances experimentally identified can be reproduced and the principal behaviour of the measured dynamic aperture explained.

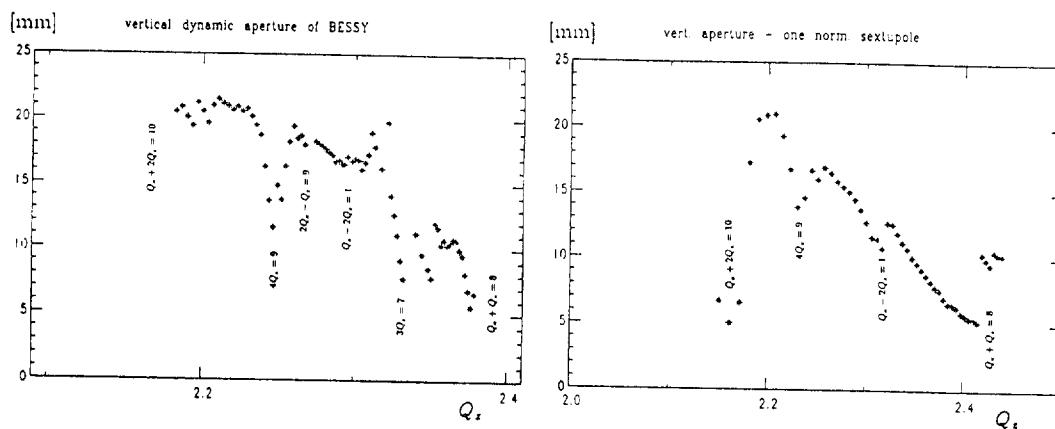


Fig. 11 The vertical dynamic aperture of BESSY

Similar experiments based on the measurement of beam lifetime when moving scrapers towards the beam were performed at the SPS in coast mode in order to evaluate the stability of particles over long time scales [26]. Figure 12 shows the experimental results with the dynamic aperture plotted as a function of sextupole currents. The influence of fifth-order resonances excited by sextupoles is clearly visible.

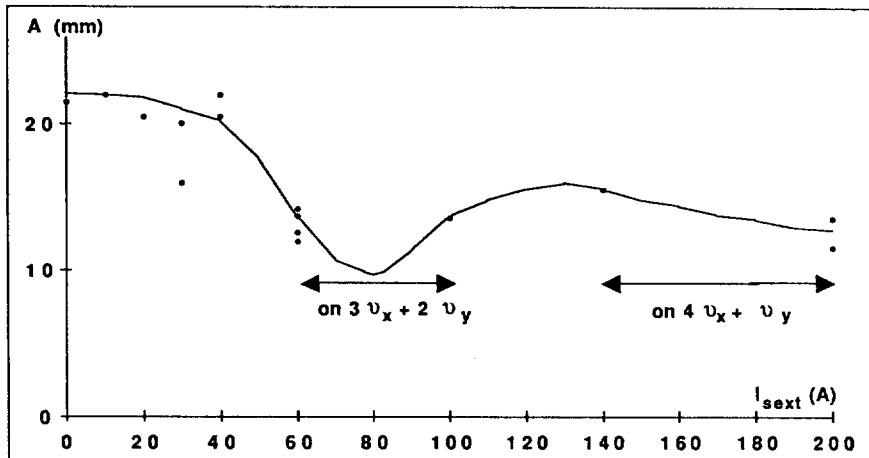


Fig. 12 SPS dynamic aperture in coast mode

In parallel, specific measurements of the dynamic aperture have been devised on existing machines such as the Tevatron [27] or the CERN SPS [28]. In both cases additional sextupoles are used to produce an adjustable third-order resonance driving term and study the non-linear dynamics in these machines in the presence of well established non-linear conditions. An example of measured and computed dynamic aperture in the presence of an array of sextupoles at various excitations [27] is shown in Fig. 13. Results are in reasonable agreement with computer simulations, when considering the uncertainties involved in simulations which are mainly due to the assumed closed orbit.

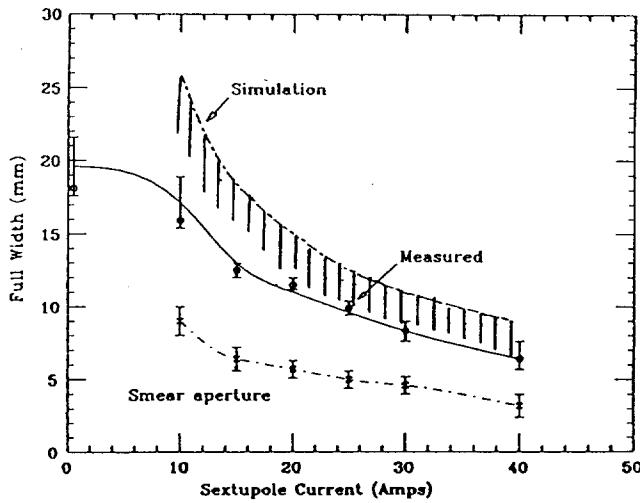


Fig. 13 Measured and computed dynamic aperture at the Tevatron

REFERENCES

- [1] J.S. Bell, Hamiltonian Mechanics, Proc. of the CERN Accelerator School on Advanced Accelerator Physics, Oxford, 1985, CERN 87-03 (1987)
- [2] J. Moser, Stabilitätverhalten kanonischer Differentialgleichungssysteme, Nach. Akad. Wiss. Göttingen IIA (6) (1955)
- [3] G. Guignard, Chromaticity: Nonlinear Aberrations, Proc. of the CERN Accelerator School on Advanced Accelerator Physics, Oxford, 1985, CERN 87-03 (1987)
- [4] A. Ando, Distortion of Beam Emittance with Nonlinear Magnetic Fields, Particle Accelerators, Vol 15, p 177 (1984)
- [5] F. Willeke, Analytical Study of the Tevatron Nonlinear Dynamics, FNAL, FN-422 (1985)
- [6] T. Collins, Distortion Functions, FNAL Tech. Note Fermilab-84/114 (1984)
- [7] A.J. Dragt, Lectures on Nonlinear Orbit Dynamics, AIP Conf. Proc., No. 87 (1981)
- [8] E. Forest, Lie Algebraic Methods for Charged Particle Beams and Light Optics, Ph.D. Thesis, Univ. Maryland (1984)
- [9] L. Michelotti, Moser-like Transformations Using the Lie Transform, Particle Accelerators, Vol 16, p 233 (1985)
- [10] G. Guignard and H. Hagel, Sextupole Correction and Dynamic Aperture. Numerical and Analytical Tools, CERN-LEP-TH/85-3 (1985)
- [11] H. Hagel and H. Moshammer, Analytic Calculation of the Dynamic Aperture for the Two Dimensional Betatron Motion in Storage Rings, EPAC, 1988 (World Scientific, Singapore, 1989) p696
- [12] M. Donald, Chromaticity Correction in Circular Accelerators and Storage Rings, PEP Note 311 (1979)
- [13] B. Autin, Chromaticity, these Proceedings
- [14] A. Ropert, Sextupole Correction Scheme for the ESRF, Proc. of the 2nd Advanced ICFA Beam Dynamics Workshop, Lugano, CERN 88-04 (1988)
- [15] L. Farvacque, J.L. Laclare, A. Ropert, BETA Users' Guide, ESRF-SR/LAT/88-08 (1987)
- [16] H. Wiedemann, Chromaticity Correction in Large Storage Rings, PEP Note 220 (1976)
- [17] A. Wrölich, RACETRACK, a Computer Code for the Simulation of Nonlinear Particle Motion in Accelerators, DESY 84-026 (1984)
- [18] L. Schachinger and R. Talman, TEAPOT, a Thin Lens Accelerator Program for Optics and Tracking, SSC Central Design Group Note SSC 52 (1985)
- [19] F.C. Iselin, the MAD Program, Proc. 12th International Conference on High Energy Accelerators, FNAL (1983)
- [20] R.V. Serfranckx et al., User's Guide to the Program DIMAD, SLAC Report 285, UC-28(A) (1985)
- [21] A.J. Dragt et al., MARYLIE 3:0, a Program for Nonlinear Analysis of Accelerators and Beam Line Lattices, IEEE Trans. Nuc. Sci., NS-32, 2311 (1983)
- [22] R.D. Ruth, A Canonical Integration Technique, IEEE Trans. Nuc. Sci., NS-30, 2669 (1983)
- [23] ESRF Foundation Phase Report (Feb. 1987)
- [24] P. Morton et al., A Diagnostic for Dynamic Aperture, IEEE Trans. Nuc. Sci., NS-32, 2291 (1985)
- [25] B. Simon and P. Kuske, The Dynamic Aperture of BESSY, Proc. of the 2nd Advanced ICFA Beam Dynamics Workshop, Lugano, CERN 88-04 (1988)
- [26] A. Hillaire, Dynamical Aperture at the SPS, Proc. of the 2nd Advanced ICFA Beam Dynamics Workshop, Lugano, CERN 88-04 (1988)
- [27] J.M. Peterson et al., Dynamic Aperture Measurements at the Tevatron, Proc. EPAC, 1988, (World Scientific, Singapore, 1989) p 266
- [28] L. Evans et al., The Non Linear Dynamic Aperture Experiment in the SPS, Proc. EPAC, 1988, (World Scientific, Singapore, 1989) p 619



## KINETIC THEORY OF CHARGED PARTICLE BEAMS

F. Ruggiero

CERN, Geneva, Switzerland

### ABSTRACT

An outline of the kinetic theory for a system of non-relativistic charged particles is presented, with special emphasis on the truncation of the BBGKY hierarchy. In particular, the existence of Debye screening in a non-neutral plasma is addressed and its relevance to the validity of the Vlasov equation is discussed. The analysis of an idealized model, consisting of a gas of charged harmonic oscillators at thermal equilibrium, shows that Debye screening is indeed possible in space-charge dominated systems. On the other hand this is not a necessary condition for the derivation of the Vlasov equation, since most of the particle pairs can be considered as uncorrelated, provided the distance of closest approach between thermal particles be much smaller than the average inter-particle distance.

### 1 INTRODUCTION

The kinetic theory has a long and fascinating history [1] closely related to many crucial scientific events such as the advent of the atomic model of matter, the abandoning of the ether hypothesis or the birth of quantum mechanics. It provides a bridge between the microscopic description of a system, in terms of reversible dynamical laws for its elementary constituents, and the macroscopic description involving experimentally accessible quantities. The latter include fluid variables, such as density or pressure, and thermodynamic variables, such as temperature (associated with thermal equilibrium) or entropy (in connection with the irreversibility of any real thermodynamic process). Equally accessible to experiment, however, are local *fluctuations* from the equilibrium state, as well as *collective* and *relaxation* phenomena that cannot be explained from the purely thermodynamical point of view. In particle accelerators, Schottky noise and the associated cooling techniques are based on local fluctuations of the charge density in the beam, while bunch lengthening, Landau damping or intra-beam scattering are examples of collective and relaxation phenomena; a correct description of these phenomena requires a kinetic approach.

The starting point of the kinetic theory is the set of coupled equations of motion for a many-particle system. The fact that any macroscopic system consists of some  $10^{10}$ – $10^{20}$  particles, interacting among themselves and with their environment, is not a conceptual problem as long as one accepts a microscopic description in terms of a complete set of observables for the system. Even neglecting quantum mechanical effects, a fully relativistic description should treat the electromagnetic field on an equal footing with the coordinates and momenta of each particle. The attempt to eliminate the field variables in favour of the particle variables results into radiation reaction and ‘memory effects’ (wake fields) that complicate the situation and shall not be considered here.

Owing to the lack of detailed information about the microscopic state of the system, one resorts to a statistical description based on the introduction of a probability density in the N-particle phase space. Similarly, to determine the initial conditions for multiparticle tracking in accelerator physics, we

customarily use random number generators with given distribution laws. In the next section, we show that the time evolution of this probability density is similar to that of an incompressible fluid (Liouville theorem).

The statistical description in the N-particle phase space is still very complicated, since it contains all the information about possible correlations between the particles. In Section 3 we discuss a systematic method (the BBGKY hierarchy) to go over from the N-particle to the single-particle phase space, consisting in the introduction of distribution functions of lower and lower order. Under certain conditions, which will be partly clarified in the following, the resulting chain of equations governing the evolution of these reduced distribution functions can be truncated at some level. Then one can express the distributions of higher order in terms of those of lower order (the choice of the initial conditions is not crucial, thanks to Landau damping [2]) and, eventually, it becomes possible to obtain a kinetic equation for the distribution function in the single-particle phase space. In general, as a consequence of the truncation, this equation contains terms that correspond to irreversible phenomena, such as diffusion and friction [3,4,5].

The simplest way to truncate the BBGKY hierarchy is to assume that all particles are independent. Then, as discussed in Section 4, the resulting kinetic equation is the so-called Vlasov equation and the corresponding (reversible) evolution of the distribution function in the single-particle phase space is again similar to that of an incompressible fluid. Although the Vlasov equation is a powerful tool to describe collective effects, such as Landau damping or bunch lengthening [6], it does not apply on the time scale of irreversible phenomena directly related to the granular structure of the system, e.g. intra-beam scattering. This is a consequence of the *independent particle approximation*, equivalent to neglecting any correlation between the particles of the system. (The same assumption is usually done when determining the initial conditions for multiparticle tracking. However, correlations can possibly develop in the subsequent evolution of the system obtained by tracking, since the particles are not replaced by a smoothed-out distribution).

In view of the long range nature of the Coulomb interaction, one might expect that the independent particle approximation can only be valid as a consequence of Debye screening. This phenomenon is usually discussed in connection with neutral plasmas, consisting of a negative electron gas moving in a background of positive ions. Let us assume that a test particle with negative charge is introduced in the plasma and that, after some time, a state of thermodynamic equilibrium characterized by a temperature  $T$  is reached. Then, as a consequence of the Boltzmann factor  $\exp(-e\varphi/T)$  associated with the Coulomb potential  $\varphi$  of the test charge, in the vicinity of this charge the electron density is lower (Coulomb repulsion) and the ion density higher (Coulomb attraction) than the unperturbed density. The net result is a screening effect, leading to an exponential decay of the potential  $\varphi$  over radial distances of the order of the Debye length. This screening effect is usually advocated to justify the independent particle approximation [5], since the Coulomb interaction between two particles at a distance much larger than the Debye length is practically suppressed. In the case of a non-neutral plasma, such as a particle beam in an accelerator, it is legitimate to ask whether Debye screening is still possible and thus whether the use of the Vlasov equation, relying upon the independent particle approximation, has any theoretical foundation.

In Section 5, we discuss an idealized model consisting of a gas of charged harmonic oscillators at thermal equilibrium. There we show that Debye screening is indeed possible, provided the system contains enough particles to be dominated by space-charge effects. Nevertheless, in most high-energy particle accelerators this last condition is not fulfilled and yet the Vlasov equation is often employed, yielding results in good agreement with experience. The relevance of Debye screening to the validity of the independent particle approximation is finally addressed in Section 6, where we compute the pair correlation function

at thermodynamic equilibrium. The result is that the ratio of the correlated and uncorrelated parts of the two-particle distribution function is small over the statistically most populated fraction of phase space, provided the distance of closest approach between thermal particles be much smaller than the average inter-particle distance. This condition is fulfilled in most accelerators and can also be expressed by saying that the ratio between the average Coulomb energy and the average kinetic energy per particle must be much smaller than one. In the opposite situation, i.e. when the beam is so cold that Coulomb energy dominates over thermal energy, new interesting phenomena are expected and in particular beam crystallization [7].

## 2 LIOUVILLE EQUATION

We consider a system of  $N$  identical particles and denote by  $\mathbf{x}_i = (\mathbf{q}_i, \mathbf{p}_i)$  the phase-space variables of the  $i^{\text{th}}$  particle. Then the microscopic state of the system corresponds to one point in the  $6N$ -dimensional phase space with coordinates  $X = (\mathbf{x}_1, \dots, \mathbf{x}_N)$ , known as the  $\Gamma$ -space. The time evolution of this representative point is associated with the  $N$ -particle Hamiltonian  $H_N(X)$  through Hamilton's equations

$$\dot{\mathbf{q}}_i = \frac{\partial H_N}{\partial \mathbf{p}_i}, \quad \dot{\mathbf{p}}_i = -\frac{\partial H_N}{\partial \mathbf{q}_i}. \quad (1)$$

For a system of non-relativistic charged particles (of mass  $m$  and charge  $e$ ), under the effect of external conservative forces, the Hamiltonian  $H_N$  reads

$$H_N(X) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m} + \sum_{i < j=1}^N \phi_{ij} + \sum_{i=1}^N U(\mathbf{q}_i), \quad (2)$$

where  $\phi_{ij} = e^2/|\mathbf{q}_i - \mathbf{q}_j|$  is the Coulomb potential energy and  $U(\mathbf{q})$  is the potential energy associated with the external field.

Since the microscopic state of the system cannot be exactly specified, we introduce a distribution function  $D(X, t)$  giving the probability density of finding the system at time  $t$  in a neighborhood of point  $X$  in the  $\Gamma$ -space. The total probability is normalized to one, i.e.

$$\int dX D(X, t) = 1, \quad (3)$$

and this must remain true for any time  $t$ . Therefore, the distribution  $D(X, t)$  must satisfy a continuity equation in  $\Gamma$ -space

$$\frac{\partial D}{\partial t} + \frac{\partial}{\partial X} (\dot{X} D) = 0, \quad (4)$$

whose meaning is simply that the local change of probability density is due to a flow of probability to or from somewhere else. The keypoint in the derivation of the Liouville equation is that, as a consequence of Hamilton's equations, the probability flow in  $\Gamma$ -space is similar to that of an *incompressible* fluid. Indeed, from Eqs. (1), we obtain

$$\frac{\partial}{\partial X} \dot{X} = \sum_{i=1}^N \frac{\partial}{\partial \mathbf{x}_i} \dot{\mathbf{x}}_i = \sum_{i=1}^N \left[ \frac{\partial}{\partial \mathbf{q}_i} \cdot \left( \frac{\partial H_N}{\partial \mathbf{p}_i} \right) + \frac{\partial}{\partial \mathbf{p}_i} \cdot \left( -\frac{\partial H_N}{\partial \mathbf{q}_i} \right) \right] = 0, \quad (5)$$

i.e. the velocity field  $\dot{X}$  is divergence-free. Therefore, the continuity equation (4) becomes the Liouville equation

$$\frac{dD}{dt} \equiv \frac{\partial D}{\partial t} + \dot{X} \frac{\partial D}{\partial X} = 0, \quad (6)$$

where  $d/dt = \partial/\partial t + \dot{X} \partial/\partial X$  denotes the *convective derivative* in  $\Gamma$ -space and expresses the rate of change along the phase-space trajectory of any representative point.

### 3 THE BBGKY HIERARCHY

A first step from the microscopic description of the system to a description more accessible to experiment consists in assuming that all points in the  $\Gamma$ -space, whose coordinates only differ by the exchange of the variables  $x_i$  and  $x_j$  of the  $i^{\text{th}}$  and the  $j^{\text{th}}$  particles, should be considered as equivalent (this is the classical counterpart of the quantum indistinguishability of particles). Therefore, we introduce an  $N$ -particle distribution function  $f_N(X, t)$  obtained by summing the probability density  $D(X, t)$  over all possible permutations of the  $N$  particles

$$f_N(X, t) = \sum_{\substack{\text{particle} \\ \text{permutations}}} D(X, t). \quad (7)$$

This distribution is thus normalized to  $N!$  and, similarly to the probability density  $D$ , its evolution is governed by the Liouville equation, which can be written

$$\frac{df_N}{dt} = L_N f_N = 0, \quad L_N f \equiv \frac{\partial f}{\partial t} + \{f, H_N\}, \quad (8)$$

where the  $N$ -particle Liouville operator  $L_N$  is defined in terms of Poisson brackets in the  $\Gamma$ -space

$$\{f, H_N\} = \sum_{i=1}^N \left( \frac{\partial f}{\partial \mathbf{q}_i} \cdot \frac{\partial H_N}{\partial \mathbf{p}_i} - \frac{\partial f}{\partial \mathbf{p}_i} \cdot \frac{\partial H_N}{\partial \mathbf{q}_i} \right). \quad (9)$$

Let us remark that, since both the Hamiltonian  $H_N$  and the initial value of  $f_N$  are invariant under the exchange of two particles, the same remains true for  $f_N(X, t)$  at all subsequent times.

As further steps towards a less and less detailed description of the system, we can subsequently integrate the  $N$ -particle distribution over the variables of one, two,  $\dots$ ,  $N - s$  particles, thus defining a hierarchy of reduced distribution functions,  $f_s(x_1, \dots, x_s, t)$ , proportional to the particle density in the  $s$ -particle phase space

$$f_s(x_1, \dots, x_s, t) = \frac{1}{(N-s)!} \int dx_{s+1} \dots dx_N f_N(x_1, \dots, x_N, t). \quad (10)$$

These reduced distributions are normalized to  $N!/(N-s)!$ , i.e. to the number of possible permutations of  $s$  particles out of  $N$ . From the Liouville equation (8) and from the definition of  $f_s$ , we obtain the following chain of equations for the reduced distributions

$$L_s f_s = \sum_{i=1}^s \int dx_{s+1} \frac{\partial \phi_{i,s+1}}{\partial \mathbf{q}_i} \cdot \frac{\partial f_{s+1}}{\partial \mathbf{p}_i}, \quad s = 1, 2, \dots, N-1, \quad (11)$$

where  $L_s$  is the Liouville operator associated with the  $s$ -particle Hamiltonian; it is defined by Eqs. (2), (8) and (9) with  $N$  replaced by  $s$ . The r.h.s. of Eq. (11) represents the Coulomb interaction with the remaining  $N - s$  particles and involves the distribution  $f_{s+1}$ , which is one step above  $f_s$  in the so-called Bogolyubov–Born–Green–Kirkwood–Yvon hierarchy. In particular, for  $s = 1$ , we obtain

$$\frac{\partial f_1}{\partial t} + \frac{\mathbf{p}_1}{m} \cdot \frac{\partial f_1}{\partial \mathbf{q}_1} - \frac{\partial U(\mathbf{q}_1)}{\partial \mathbf{q}_1} \cdot \frac{\partial f_1}{\partial \mathbf{p}_1} = \int d\mathbf{q}_2 d\mathbf{p}_2 \frac{\partial \phi_{12}}{\partial \mathbf{q}_1} \cdot \frac{\partial f_2}{\partial \mathbf{p}_1}, \quad (12)$$

where the l.h.s. is  $L_1 f_1$  and corresponds to the evolution of a free particle under the effect of the external potential  $U$ , while the r.h.s. contains the Coulomb interaction with the remaining  $N - 1$  particles. The distribution function  $f_1(\mathbf{x}, t) = f(\mathbf{q}, \mathbf{p}, t)$  is the particle density in the single-particle phase space, also called  $\mu$ -space, and is normalized to the total number of particles  $N$ .

#### 4 VLASOV EQUATION

The simplest way to truncate the BBGKY hierarchy is to assume that all particles are independent. In particular, for very large  $N$ , this implies that the probability of finding a particle at  $\mathbf{x}_1$  is not influenced by the fact that another particle be at  $\mathbf{x}_2$  and thus that the two-particle distribution  $f_2(\mathbf{x}_1, \mathbf{x}_2, t)$  is the product of the single-particle distributions  $f(\mathbf{x}_1, t)$  and  $f(\mathbf{x}_2, t)$

$$f_2(\mathbf{x}_1, \mathbf{x}_2, t) = f(\mathbf{x}_1, t)f(\mathbf{x}_2, t). \quad (13)$$

When expression (13) can be used, from Eq. (12) we obtain the (non-relativistic) Vlasov equation

$$\frac{\partial f}{\partial t} + \frac{\mathbf{p}}{m} \cdot \frac{\partial f}{\partial \mathbf{q}} - \left( \frac{\partial U}{\partial \mathbf{q}} + e \frac{\partial \Phi}{\partial \mathbf{q}} \right) \cdot \frac{\partial f}{\partial \mathbf{p}} = 0. \quad (14)$$

The first term in parenthesis represents the external force and the second term is the internal electrostatic force associated with the *self-consistent* Coulomb potential  $\Phi(\mathbf{q}, t)$

$$\Phi(\mathbf{q}, t) = \int d^3 q' d^3 p' \frac{e}{|\mathbf{q} - \mathbf{q}'|} f(\mathbf{q}', \mathbf{p}', t). \quad (15)$$

Let us remark that this self-consistent potential can be expressed in terms of the charge density  $\rho(\mathbf{q}, t)$ , proportional to the particle density  $n(\mathbf{q}, t)$

$$n(\mathbf{q}, t) = \int d^3 p f(\mathbf{q}, \mathbf{p}, t), \quad \rho(\mathbf{q}, t) = e n(\mathbf{q}, t). \quad (16)$$

Indeed we have

$$\Phi(\mathbf{q}, t) = \int d^3 q' \frac{\rho(\mathbf{q}', t)}{|\mathbf{q} - \mathbf{q}'|} \quad (17)$$

and the potential  $\Phi$  satisfies the Poisson equation

$$\nabla^2 \Phi = -4\pi\rho. \quad (18)$$

We can say that, in independent particle approximation, the system behaves as a *charged fluid*; this fluid generates its own electric field, through the charge density  $\rho$  and the Poisson equation, and moves according to external forces plus self-consistent electrostatic forces. Therefore, the Vlasov equation can be cast into the form of a Liouville equation in  $\mu$ -space

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \{f, H\} = 0, \quad (19)$$

where the effective single-particle Hamiltonian  $H(\mathbf{q}, \mathbf{p}, t)$  reads

$$H(\mathbf{q}, \mathbf{p}, t) = \frac{\mathbf{p}^2}{2m} + e\Phi(\mathbf{q}, t) + U(\mathbf{q}). \quad (20)$$

Equation (19) governs the time evolution of an incompressible fluid  $f(\mathbf{q}, \mathbf{p}, t)$  in  $\mu$ -space. However, it is important to recall that the effective Hamiltonian depends on  $f$ , through the self-consistent potential  $\Phi$ , and thus that the Vlasov equation is nonlinear in  $f$  and describes the *collective behaviour* of our system. For a particle accelerator, this corresponds to the broad class of current-dependent phenomena that are not related to the granular structure of the particle beams. In view of the fact that the Vlasov equation includes the average Coulomb interaction, it becomes less surprising that the independent particle approximation can be a good approximation. Indeed it amounts to neglecting fluctuations of the Coulomb field from its average value, associated with the granularity of the particle distribution. Even in the absence of Debye screening (that still occurs in space-charge dominated systems, as we shall see in the next section), it is

reasonable to expect that these fluctuations become large only at short inter-particle distances (binary collisions).

When the single-particle distribution can be written as a function  $f = f[H]$  of the effective Hamiltonian  $H$ , the system is in a self-consistent steady state. Indeed  $\{f, H\} = 0$  and therefore  $\partial f / \partial t = 0$ , i.e. the distribution  $f$  does not depend explicitly on time (this result is valid provided the external potential  $U$  is not time dependent). In particular, the state of thermodynamic equilibrium at temperature  $T$  is described by the Boltzmann distribution

$$f = f_o \exp\left(-\frac{H}{T}\right). \quad (21)$$

It should be stressed, however, that the existence of a self-consistent equilibrium corresponding to a given functional dependence  $f[H]$  cannot be taken for granted: for example, the regime of potential well bunch lengthening in electron storage rings [6] is valid only at low beam currents. Let us also remark that the most general self-consistent steady state, characterized by  $\{f, H\} = 0$ , corresponds to a distribution  $f$  which is a generic function of all the invariants of motion associated with the effective Hamiltonian [8].

## 5 DEBYE SCREENING IN A CHARGED PLASMA: AN EXAMPLE

In order to discuss the phenomenon of Debye screening in a non-neutral plasma, in this section we consider the example of a gas of charged harmonic oscillators at thermodynamic equilibrium. For the sake of simplicity, we assume that the harmonic oscillators be isotropic and thus that the external potential energy has spherical symmetry

$$U(\mathbf{q}) = \frac{1}{2} k \mathbf{q} \cdot \mathbf{q} = \frac{1}{2} kr^2, \quad (22)$$

where  $k$  is the elastic constant and  $r$  the radial variable. This is a highly idealized model for a bunch of particles circulating in a storage ring: in smooth approximation, the elastic force corresponds to betatron and synchrotron focusing towards the center of the bunch. On the other hand, this is a relatively good model for an ion trap [7]. In both cases, the physical origin of the elastic force is the electromagnetic field generated by charges and currents located *outside* the system.

Using the Boltzmann distribution Eq. (21) and inserting the elastic potential (22) into the effective Hamiltonian (20), the particle density  $n(r)$  corresponding to a steady-state solution of the Vlasov equation becomes

$$n(r) = n_o \exp\left\{-\frac{e\Phi(r) + \frac{1}{2}kr^2}{T}\right\} \quad (23)$$

and the Poisson equation for the self-consistent Coulomb potential  $\Phi$  reads

$$\nabla^2 \Phi(r) = \frac{1}{r^2} \frac{d}{dr} \left[ r^2 \frac{d\Phi(r)}{dr} \right] = -4\pi e n(r). \quad (24)$$

These two equations must be solved with the following normalization constraint on  $n(r)$ :

$$\int_0^\infty dr 4\pi r^2 n(r) = N. \quad (25)$$

Therefore, we look for a differential equation for  $n(r)$ .

We start from the Poisson equation (24) that, after multiplication by  $r^2$  and integration over  $r$ , becomes

$$-\frac{d\Phi(r)}{dr} = \frac{e}{r^2} \int_0^r dr' 4\pi r'^2 n(r') = \frac{eN(r)}{r^2} = E_r(r), \quad (26)$$

where  $N(r)$  denotes the number of particles within a sphere of radius  $r$  and  $E_r$  is the radial component of the electric field. The last equality shows that the electric field is given by the Coulomb law for a point-like

charge  $eN(r)$ . Computing the gradient of the density  $n(r)$ , given by Eq. (23), we obtain

$$T \frac{dn(r)}{dr} = n(r) [eE_r(r) - kr] = \nabla_r p. \quad (27)$$

Here  $eE_r$  is the electrostatic repulsive force and  $-kr$  the elastic restoring force in the radial direction; their sum, multiplied by the density  $n$ , can be equated to the radial gradient of the pressure  $p$  in our charged fluid. Therefore, as a consequence of the Boltzmann distribution, we recover the ideal gas equation of state  $p = nT$ . Multiplying Eq. (27) by  $r^2/n(r)$ , differentiating with respect to  $r$  and using Eq. (26) we have

$$\frac{d}{dr} \left[ r^2 \frac{T}{n(r)} \frac{dn(r)}{dr} \right] = \frac{d}{dr} [e^2 N(r) - kr^3] = 4\pi e^2 n(r) r^2 - 3kr^2. \quad (28)$$

This is the differential equation for  $n(r)$  we were looking for. It can also be written

$$\nabla^2 (\ln[n(r)]) = \frac{4\pi e^2 n(r) - 3k}{T}. \quad (29)$$

We shall discuss the solution of Eq. (29) in the case of unlimited charge (i.e. for  $N \rightarrow \infty$ ), in the low charge limit and, finally, for a space-charge dominated system with a finite number of particles.

### 5.1 Unlimited charge

In the limit  $N \rightarrow \infty$ , Eq. (29) has the following simple solution, corresponding to a uniform particle density

$$n(r) = \bar{n} = \frac{3k}{4\pi e^2}. \quad (30)$$

For this special value of the particle density, the charge in a sphere of radius  $r$  is

$$eN(r) = e \int_0^r dr' 4\pi r'^2 n(r') = \frac{4}{3}\pi \bar{n} e r^3 = \frac{k}{e} r^3 \quad (31)$$

and the radial electric field depends linearly on  $r$

$$eE_r(r) = \frac{e^2 N(r)}{r^2} = kr. \quad (32)$$

Therefore there is a balance between the electrostatic force  $eE_r$  and the elastic force  $-kr$ , corresponding to a vanishing pressure gradient  $\nabla_r p = 0$ .

Let us remark that the uniform density  $\bar{n}$  is automatically selected by the system under the only condition of unlimited charge. This value  $\bar{n}$  can be used as a natural unit to normalize the density  $n(r)$ , even in the general case of a finite number of particles. Therefore, Eq. (29) can be written

$$\nabla^2 \left[ \ln \left( \frac{n(r)}{\bar{n}} \right) \right] = \frac{3k}{T} \left[ \frac{n(r)}{\bar{n}} - 1 \right] \quad (33)$$

and the quantity  $3k/T$  can be associated to a fundamental length  $\bar{\lambda}_D$ , to be identified with the Debye length of a (non-neutral) plasma with density  $\bar{n}$  and temperature  $T$

$$\bar{\lambda}_D^2 = \frac{T}{3k} = \frac{T}{4\pi \bar{n} e^2}. \quad (34)$$

Using  $\bar{\lambda}_D$  as a natural unit to normalize  $r$  and denoting by  $x = r/\bar{\lambda}_D$  the dimensionless radial coordinate, Eq. (33) can be finally cast into the form

$$\frac{1}{x^2} \frac{d}{dx} \left\{ x^2 \frac{d[\ln(n/\bar{n})]}{dx} \right\} = \frac{n}{\bar{n}} - 1. \quad (35)$$

## 5.2 Low charge limit

We now discuss Eq. (35) in the limit

$$\frac{n}{\bar{n}} \ll 1. \quad (36)$$

Therefore we neglect  $n/\bar{n}$  in the r.h.s. of this equation, multiply it by  $x^2$  and integrate with respect to  $x$ , thus obtaining

$$\frac{d[\ln(n/\bar{n})]}{dx} = -\frac{x}{3}. \quad (37)$$

The solution is

$$\frac{n}{\bar{n}} = A \exp\left(-\frac{x^2}{6}\right) = A \exp\left(-\frac{kr^2}{2T}\right), \quad (38)$$

where  $A$  is an integration constant determined by the normalization condition Eq. (25), i.e. by

$$\frac{N}{\bar{n}} = A \int_0^\infty dr 4\pi r^2 \exp\left(-\frac{kr^2}{2T}\right) = A \left(\frac{2\pi T}{k}\right)^{3/2}. \quad (39)$$

On the other hand, our assumption Eq. (36) is valid only when  $A \ll 1$ . Therefore, we must require that

$$A = \frac{N}{\bar{n}} \left(\frac{k}{2\pi T}\right)^{3/2} = \frac{2}{3} Ne^2 \sqrt{\frac{k}{2\pi T^3}} \ll 1. \quad (40)$$

This is equivalent to

$$N \ll N_D = \frac{3}{2e^2} \sqrt{\frac{2\pi T^3}{k}} \sim \bar{n} \bar{\lambda}_D^3, \quad (41)$$

where  $N_D$  will be shown to be the minimum number of particles required in order to have Debye screening in our charged plasma. Therefore Eq. (36) is valid in the low charge limit, i.e. when the total number of charged particles  $N$  is much smaller than the number of particles  $N_D$  that would be contained in a sphere of radius  $\bar{\lambda}_D$  with uniform density  $\bar{n}$ . Under this condition, the density  $n(r)$  given by Eq. (38) can be written

$$n(r) = N \left(\frac{k}{2\pi T}\right)^{3/2} \exp\left(-\frac{kr^2}{2T}\right) \quad (42)$$

and corresponds to a Boltzmann distribution in the elastic potential alone, i.e. without space-charge effects.

It is interesting to note that, in the low charge limit, the Debye length  $\bar{\lambda}_D$  of Eq. (34) coincides with the standard deviation of the Gaussian density  $n(r)$ , divided by  $\sqrt{3}$ . Therefore, in the case of a low-current electron storage ring,  $\bar{\lambda}_D$  can be roughly identified with the r.m.s. transverse beam size in smooth approximation (this identification should not be taken too seriously, since our model is valid only for a system of isotropic oscillators. Furthermore, in strong focusing machines the beam is *not* in thermodynamic equilibrium [9] and its temperature is not well defined). Let us stress that  $\bar{\lambda}_D$  is the Debye length of a plasma with density  $\bar{n} \gg n$  and thus it is much shorter than the *local* Debye length, usually defined as  $\lambda_D = \sqrt{T/4\pi ne^2}$ .

The ratio  $n/\bar{n} \sim N/N_D$  for a high-energy collider can be expressed in terms of the linear beam-beam tune shift  $\xi$  as follows:

$$\frac{n}{\bar{n}} \sim \frac{1}{\gamma^2} \frac{\xi}{\nu_\beta} \frac{L}{\sigma_x} \frac{\sigma_z}{\sigma_y}, \quad (43)$$

where  $\gamma$  denotes the Lorentz factor corresponding to the beam energy,  $\nu_\beta$  is the betatron tune,  $L/\sigma_z$  the ratio between circumference of the storage ring and bunch length, while  $\sigma_x/\sigma_y$  is the average horizontal to vertical aspect ratio of the beam in the regular lattice (it reduces to unity for a round beam). Indeed the tune shift due to direct space-charge effect should become equal to  $\nu_\beta$  for  $n = \bar{n}$ : it is related to the linear beam-beam tune shift through the relativistic factor  $1/\gamma^2$ , as a consequence of the partial compensation of

electric and magnetic forces, and the bunching factor  $L/\sigma_z$ , that accounts for the localization of the beam-beam interaction in contrast to the space-charge effect which is distributed. Assuming that the beam-beam tune shift is the same in the horizontal and in the vertical plane, for a flat beam the space-charge induced tune shift is larger for vertical oscillations; this is the reason for the aspect ratio appearing in Eq. (43). In the case of LEP at 50 GeV, we have  $\gamma \approx 10^5$ ,  $v_B \approx 77$ ,  $\sigma_x/\sigma_y \approx 5$ ,  $L \approx 27$  km and  $\sigma_z \approx 1$  cm. Therefore, even at the beam-beam limit when  $\xi \approx 0.05$ , we obtain a ratio  $n/\bar{n} \sim 10^{-6}$ , indicating that the plasma is far from being dominated by direct space-charge effects.

### 5.3 Space-charge dominated plasma

When the number of particles is finite, but much larger than  $N_D$ , the results of the analysis relative to the case of unlimited charge suggest that the particle density in a sphere of radius  $R \gg \bar{\lambda}_D$  will be almost uniform and equal to  $\bar{n}$ . The radius  $R$  of this sphere can be estimated by assuming that it contains almost all the particles, i.e. that  $\frac{4}{3}\pi R^3 \bar{n} \sim N$ . Therefore

$$R \sim \left( \frac{N}{\frac{4}{3}\pi \bar{n}} \right)^{1/3} = \left( \frac{Ne^2}{k} \right)^{1/3} \quad (44)$$

and we see that  $N \gg N_D \sim \bar{n}\bar{\lambda}_D^3$  implies  $R \gg \bar{\lambda}_D$ .

In order to compute  $n(r)$  in the inner region of the sphere of radius  $R$ , we write

$$\frac{n}{\bar{n}} = 1 + \frac{\delta n}{\bar{n}} \quad (45)$$

and assume that the relative density deviation  $|\delta n/\bar{n}|$  be much smaller than unity for  $r \ll R$ . Then we have

$$\ln \left( \frac{n}{\bar{n}} \right) \approx \frac{\delta n}{\bar{n}} \quad (46)$$

and Eq. (35) becomes

$$\frac{1}{x^2} \frac{d}{dx} \left[ x^2 \frac{d}{dx} \left( \frac{\delta n}{\bar{n}} \right) \right] = \frac{\delta n}{\bar{n}}. \quad (47)$$

This equation is linear in  $\delta n/\bar{n}$  and, for  $x > 0$ , it has a fundamental system of solutions given by  $e^{\pm x}/x$ . On the other hand, we must choose a solution that is non-singular at the center of the sphere, i.e. for  $x = 0$ , and thus the only possible linear combination of the two fundamental solutions is

$$\frac{\delta n}{\bar{n}} = -\epsilon \frac{\sinh x}{x}, \quad (48)$$

where the factor  $\epsilon$  remains to be determined. To this end, we remark that  $\epsilon$  represents the relative density deviation  $|\delta n/\bar{n}|$  at the origin and thus must be much smaller than unity. On the other hand, owing to the growing exponential in the hyperbolic sine, our assumption  $|\delta n/\bar{n}| \ll 1$  is no longer valid for large values of  $x$ , namely above some maximum value  $x_{max}$  such that

$$\epsilon \frac{e^{x_{max}} - e^{-x_{max}}}{2x_{max}} \sim 1. \quad (49)$$

The corresponding maximum radius  $r_{max}$  can be identified with the radius  $R$  of the sphere containing most of the particles, given by Eq. (44). Therefore we obtain  $x_{max} \sim R/\bar{\lambda}_D \gg 1$  and hence

$$\epsilon \sim 2 \frac{R}{\bar{\lambda}_D} e^{-(R/\bar{\lambda}_D)}. \quad (50)$$

Since  $\bar{\lambda}_D$  is the only fundamental length appearing in the problem, at the border of the sphere of radius  $R$  there will be a layer of a few Debye lengths, where the particle density  $n(r)$  drops substantially below  $\bar{n}$ .

Finally, for  $r \gg R$ , the elastic potential dominates over the electrostatic one and, according to the results obtained in the low charge limit, the particle density has a fast Gaussian decay.

We recall that the elastic restoring force towards the origin is generated by charges and currents located *outside* the system. In the case of unlimited charge in the plasma, this restoring force is perfectly balanced by Coulomb repulsion while, for a space-charge dominated system with a finite number of particles, the balance in the bulk of the plasma gets better and better as we move away from the Debye layer that separates the inside from the outside of the plasma. Therefore we can interpret this phenomenon by saying that, in the bulk of the plasma, the effect of external charges and currents is screened. At this point, the following remarks are in order:

1. There can be Debye screening even in a charged plasma.
2. This screening is a collective effect, since it follows from the Vlasov equation (similar results can be obtained also for  $f[H]$  different from the Boltzmann distribution).
3. In order to have Debye screening, the plasma must extend over many Debye lengths. This implies  $N \gg N_D$  and thus a non-neutral plasma will be dominated by space-charge effects.

#### 5.4 Debye screening for a test particle.

Here we consider the behaviour of our gas of charged harmonic oscillators in presence of a test particle with charge  $Q$ , located at the origin  $r = 0$ . The resulting density perturbation shows that Debye screening is possible even for charges located inside the plasma and is closely related to the pair correlation function to be computed in the next section. Instead of Eq. (23), the particle density  $n(r)$  corresponding to the Boltzmann distribution (21) will now satisfy the equation

$$n(r) = n_0 \exp \left\{ -\frac{e\Phi(r) + \frac{1}{2}kr^2 + eQ/r}{T} \right\}, \quad (51)$$

where  $Q/r$  is the Coulomb potential of the test charge and the self-consistent potential  $\Phi(r)$ , associated with the charge density  $en(r)$  of the plasma, is again solution of the Poisson equation (24). Following the same steps leading to Eq. (35), the differential equation for the particle density now reads

$$\frac{1}{x^2} \frac{d}{dx} \left\{ x^2 \frac{d[\ln(n/\bar{n})]}{dx} \right\} = \frac{n}{\bar{n}} - 1 + 2 \frac{eQ}{T\bar{\lambda}_D} \frac{\delta(x)}{x^2}, \quad (52)$$

where we have used the identities  $\nabla^2(1/r) = -4\pi\delta(r)$  and  $\delta(r) = \delta(r)/(2\pi r^2)$ , the latter being valid in the case of spherical symmetry, when the three-dimensional volume element can be written  $4\pi r^2 dr$ .

In the bulk of a space-charge dominated plasma, but far from the origin  $x = 0$  where the test charge is located, we can again assume that the relative density deviation  $|\delta n/\bar{n}|$  be much smaller than unity. Then, expanding the logarithm, we obtain

$$\frac{1}{x^2} \frac{d}{dx} \left[ x^2 \frac{d}{dx} \left( \frac{\delta n}{\bar{n}} \right) \right] = \frac{\delta n}{\bar{n}} + 2 \frac{eQ}{T\bar{\lambda}_D} \frac{\delta(x)}{x^2}, \quad (53)$$

whose solution is

$$\frac{\delta n}{\bar{n}} = -\frac{eQ}{T\bar{\lambda}_D} \frac{e^{-x}}{x}. \quad (54)$$

Recalling that  $x = r/\bar{\lambda}_D$ , we see that the density perturbation associated with the test charge decays exponentially with a characteristic length  $\bar{\lambda}_D$ , exhibiting again the phenomenon of Debye screening.

Let us remark that the coefficient  $eQ/T\bar{\lambda}_D$  in the solution (54) is a direct consequence of the last term in the linearized equation (53), containing a Dirac delta-function  $\delta(x)$ . On the other hand, for  $x \rightarrow 0$ , expression (54) diverges, contrary to the assumption of a small density deviation. Therefore the linearized equation (53) is no longer valid in the vicinity of the origin (where the convergence of any expansion in powers of  $\delta n/\bar{n}$  is questionable) and it is not clear whether it can be used to compute the effect of a singularity at the origin on the density perturbation at large distances from the origin. This point is usually overlooked in most textbooks on plasma physics [3,4], although the perturbative character of expression (54) is subsequently stressed in connection with a related divergence occurring in the collision integral of the kinetic equation (see in particular [2]). The only way to compute the amplitude of the density perturbation associated with the test charge  $Q$  is to solve (numerically) the exact equation (52). However, provided  $Q$  and  $e$  have the same sign, we can give a qualitative argument for the coefficient  $eQ/T\bar{\lambda}_D$  appearing in the approximate solution (54). In this case, as a consequence of Coulomb repulsion, the plasma charge contained in a sphere of smaller and smaller radius must tend to zero and, in the same limit, the self-consistent potential  $\Phi(r)$  is negligible with respect to the potential  $Q/r$  of the test charge. Therefore we can directly use Eq. (51) to show that, at very small distances from the origin, the plasma density becomes  $n(r) \sim n_0 \exp(-eQ/rT)$ . On the other hand, in the case of unlimited plasma, we know that  $n(r)$  approaches the asymptotic value  $\bar{n}$  for large values of  $r$  and that the density perturbation  $\delta n/\bar{n}$  must be proportional to  $(1/r) \exp(-r/\bar{\lambda}_D)$ , since this is the only bounded solution of the linearized equation (47). A possible interpolation formula, giving the particle density in both radial regions is then the following:

$$n(r) = \bar{n} \exp\left(-\frac{eQ}{T} \frac{e^{-(r/\bar{\lambda}_D)}}{r}\right). \quad (55)$$

Although this is not an exact solution of Eq. (52), it has the advantage of being well behaved and represents a very good approximation to the exact density profile, as computed by numerical methods: for values of  $r$  much larger than  $eQ/T$ , the density perturbation corresponding to this expression coincides to first order with Eq. (54).

The length  $eQ/T$  is the distance of closest approach to the test charge  $Q$  for particles with thermal kinetic energy  $T$ ; indeed, at this distance, the height of the electrostatic potential barrier due to Coulomb repulsion becomes equal to their initial kinetic energy. Our calculation shows that the plasma density tends to zero below this distance, while it relaxes to the unperturbed density  $\bar{n}$  at distances much larger than  $eQ/T$ , but not necessarily as large as the Debye length  $\bar{\lambda}_D$  (which typically exceeds the distance of closest approach for thermal particles by many orders of magnitude). In the case of unlimited plasma, this result is valid even when the test charge  $Q$  is not located at the origin, since for  $Q = 0$  the unperturbed density  $\bar{n}$  is uniform. In the general case, i.e. when the plasma density in the absence of test charge is not uniform, the same result remains qualitatively correct provided the unperturbed density does not vary too much over distances comparable to the distance of closest approach for thermal particles. This condition is equivalent to assuming that the external force plus the self-consistent force be negligible with respect to the local Coulomb repulsion in the vicinity of the test charge: then the previous argument based on Eq. (51) can be used again. In order to have Debye screening, however, the unperturbed density should remain practically constant over distances comparable to  $\bar{\lambda}_D$  and, as we have seen, this can only be possible in the bulk of a space-charge dominated plasma (see also [8]).

## 6 CLUSTER EXPANSION

In this section we discuss the relevance of Debye screening to the truncation of the BBGKY hierarchy, associated with the chain of equations (11). In particular, the previous analysis of the test particle problem is used as a guideline to compute the pair correlation function at thermodynamic equilibrium. The result indicates that, under reasonable conditions, the independent particle approximation can be used even in the absence of Debye screening.

We recall that the first two equations of the BBGKY hierarchy read

$$\left\{ \frac{\partial}{\partial t} + \left[ \frac{\mathbf{p}_1}{m} \cdot \frac{\partial}{\partial \mathbf{q}_1} - \frac{\partial U(\mathbf{q}_1)}{\partial \mathbf{q}_1} \cdot \frac{\partial}{\partial \mathbf{p}_1} \right] \right\} f_1(\mathbf{x}_1, t) = \int d\mathbf{x}_2 \frac{\partial \phi_{12}}{\partial \mathbf{q}_1} \cdot \frac{\partial f_2(\mathbf{x}_1, \mathbf{x}_2, t)}{\partial \mathbf{p}_1}, \quad (56)$$

$$\begin{aligned} \left\{ \frac{\partial}{\partial t} + \left[ \frac{\mathbf{p}_1}{m} \cdot \frac{\partial}{\partial \mathbf{q}_1} - \frac{\partial (U(\mathbf{q}_1) + \phi_{12})}{\partial \mathbf{q}_1} \cdot \frac{\partial}{\partial \mathbf{p}_1} \right] + \left[ \frac{\mathbf{p}_2}{m} \cdot \frac{\partial}{\partial \mathbf{q}_2} - \frac{\partial (U(\mathbf{q}_2) + \phi_{12})}{\partial \mathbf{q}_2} \cdot \frac{\partial}{\partial \mathbf{p}_2} \right] \right\} f_2(\mathbf{x}_1, \mathbf{x}_2, t) \\ = \int d\mathbf{x}_3 \left\{ \frac{\partial \phi_{13}}{\partial \mathbf{q}_1} \cdot \frac{\partial}{\partial \mathbf{p}_1} + \frac{\partial \phi_{23}}{\partial \mathbf{q}_2} \cdot \frac{\partial}{\partial \mathbf{p}_2} \right\} f_3(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, t). \end{aligned} \quad (57)$$

A systematic way to describe deviations from the independent particle approximation is the following Mayer cluster expansion:

$$\begin{aligned} f_1(\mathbf{x}_1, t) &\equiv f(1), \\ f_2(\mathbf{x}_1, \mathbf{x}_2, t) &\equiv f(1)f(2) + P(1, 2), \end{aligned} \quad (58)$$

$$f_3(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, t) \equiv f(1)f(2)f(3) + f(1)P(2, 3) + f(2)P(3, 1) + f(3)P(1, 2) + T(1, 2, 3),$$

where  $P(1, 2)$  denotes the pair correlation function,  $T(1, 2, 3)$  the three-particle correlation function and so on. Using the cluster expansion, Eq. (56) becomes

$$\left\{ \frac{\partial}{\partial t} + \left[ \frac{\mathbf{p}_1}{m} \cdot \frac{\partial}{\partial \mathbf{q}_1} - \frac{\partial V(\mathbf{q}_1, t)}{\partial \mathbf{q}_1} \cdot \frac{\partial}{\partial \mathbf{p}_1} \right] \right\} f(1) = \int d\mathbf{x}_2 \frac{\partial \phi_{12}}{\partial \mathbf{q}_1} \cdot \frac{\partial P(1, 2)}{\partial \mathbf{p}_1}, \quad (59)$$

where  $V$  is the sum of the external potential energy  $U$  and of the self-consistent electrostatic energy  $e\Phi$

$$V(\mathbf{q}_1, t) = U(\mathbf{q}_1) + e\Phi(\mathbf{q}_1, t) = U(\mathbf{q}_1) + \int d\mathbf{x}_2 \phi_{12} f(2). \quad (60)$$

This is a generalization of the Vlasov equation (14) and, if  $P(1, 2)$  can be expressed in terms of  $f(1)$  and  $f(2)$ , the resulting kinetic equation describes the irreversible effect of collisions. In a similar way, Eq. (57) for the two-particle distribution  $f_2(\mathbf{x}_1, \mathbf{x}_2, t) \equiv f_2(1, 2)$  can be written

$$\begin{aligned} \left\{ \frac{\partial}{\partial t} + \left[ \frac{\mathbf{p}_1}{m} \cdot \frac{\partial}{\partial \mathbf{q}_1} - \frac{\partial (V(\mathbf{q}_1, t) + \phi_{12})}{\partial \mathbf{q}_1} \cdot \frac{\partial}{\partial \mathbf{p}_1} \right] + \left[ \frac{\mathbf{p}_2}{m} \cdot \frac{\partial}{\partial \mathbf{q}_2} - \frac{\partial (V(\mathbf{q}_2, t) + \phi_{12})}{\partial \mathbf{q}_2} \cdot \frac{\partial}{\partial \mathbf{p}_2} \right] \right\} f_2(1, 2) \\ = \int d\mathbf{x}_3 \left\{ \frac{\partial \phi_{13}}{\partial \mathbf{q}_1} \cdot \frac{\partial}{\partial \mathbf{p}_1} + \frac{\partial \phi_{23}}{\partial \mathbf{q}_2} \cdot \frac{\partial}{\partial \mathbf{p}_2} \right\} [f(1)P(2, 3) + f(2)P(1, 3) + T(1, 2, 3)]. \end{aligned} \quad (61)$$

Let us notice the symmetry of this equation with respect to the indices 1 and 2, that is a consequence of our original definition (7) of the  $N$ -particle distribution function.

We now remark that, owing to Coulomb repulsion, in the limit  $\mathbf{q}_1 \rightarrow \mathbf{q}_2$  both  $f_2(\mathbf{x}_1, \mathbf{x}_2, t)$  and  $f_3(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, t)$  should tend to zero. Therefore, in the same limit, we must require that  $P(1, 2) \rightarrow -f(1)f(2)$  and  $T(1, 2, 3) \rightarrow -[f(1)P(2, 3) + f(2)P(1, 3)]$ ; in particular, this implies that the r.h.s of Eq. (61) tends to zero. At thermodynamic equilibrium, the single-particle distribution  $f(1)$  for a locally uniform plasma has the Boltzmann form

$$f(1) = f_B(1) = \frac{n}{(2\pi mT)^{3/2}} \exp\left(-\frac{p_1^2}{2mT}\right), \quad (62)$$

where  $n$  is the *local* density of the plasma. Moreover, for  $\mathbf{q}_1 \rightarrow \mathbf{q}_2$  we can neglect the external plus the self-consistent potential energy  $V$ , with respect to the electrostatic energy  $\phi_{12}$  associated with Coulomb repulsion. Then the equilibrium solution of Eq. (61) must have the following limiting form:

$$f_2(1, 2) \rightarrow f_B(1)f_B(2) \exp\left(-\frac{\phi_{12}}{T}\right). \quad (63)$$

This result parallels our previous discussion of the test particle problem, based on the self-consistency equation (51). In this case, however, the equation for the two-particle distribution  $f_2(1, 2)$  depends on the three-particle correlation  $T(1, 2, 3)$  and, in general, there is no self-consistency equation for  $f_2$  alone.

The strategy to truncate the chain of equations corresponding to the BBGKY hierarchy relies on the idea that, under reasonable conditions, one can consistently assume

$$T \ll Pf \ll fff. \quad (64)$$

Although this assumption is not justified in the limit of vanishing spatial separation  $r_{12} = |\mathbf{q}_1 - \mathbf{q}_2| \rightarrow 0$ , we now check its consistency in a statistical sense, namely for values of  $r_{12}$  of the order of the average inter-particle distance. Then we show that, using this assumption, the average value of the ratio  $P/f_f$  at thermodynamic equilibrium turns out to be

$$\langle \frac{P(1, 2)}{f(1)f(2)} \rangle \sim \exp\left(-\frac{\ell_T}{\bar{\ell}}\right) - 1, \quad (65)$$

where  $\ell_T$  is the distance of closest approach between thermal particles and  $\bar{\ell}$  the average inter-particle distance

$$\ell_T = \frac{e^2}{T}, \quad \bar{\ell} = n^{-1/3}. \quad (66)$$

To prove this, we also require that the local density  $n$  be approximately constant at least over distances comparable to  $\bar{\ell}$ . Then, provided  $\ell_T \ll \bar{\ell}$ , one can consistently use the independent particle approximation and formally derive the Vlasov equation, even in the absence of Debye screening. Let us remark that the ratio  $\ell_T/\bar{\ell}$  is equal to the ratio  $\langle \phi_{12} \rangle / T$  between the average Coulomb energy (associated with binary collisions) and the average kinetic energy per particle.

Expressing the two-particle distribution in terms of the pair correlation function and using Eq. (59), from Eq. (61) we obtain the following equation for  $P(1, 2)$ :

$$\begin{aligned} & \left\{ \frac{\partial}{\partial t} + \left[ \frac{\mathbf{p}_1}{m} \cdot \frac{\partial}{\partial \mathbf{q}_1} - \frac{\partial(V(\mathbf{q}_1, t) + \phi_{12})}{\partial \mathbf{q}_1} \cdot \frac{\partial}{\partial \mathbf{p}_1} \right] + \left[ \frac{\mathbf{p}_2}{m} \cdot \frac{\partial}{\partial \mathbf{q}_2} - \frac{\partial(V(\mathbf{q}_2, t) + \phi_{12})}{\partial \mathbf{q}_2} \cdot \frac{\partial}{\partial \mathbf{p}_2} \right] \right\} P(1, 2) \\ & - \frac{\partial \phi_{12}}{\partial \mathbf{q}_1} \cdot \left[ f(2) \frac{\partial f(1)}{\partial \mathbf{p}_1} - f(1) \frac{\partial f(2)}{\partial \mathbf{p}_2} \right] = \int d\mathbf{x}_3 \left\{ \frac{\partial \phi_{13}}{\partial \mathbf{q}_1} \cdot \frac{\partial}{\partial \mathbf{p}_1} [f(1)P(2, 3) + T(1, 2, 3)] \right. \\ & \left. + \frac{\partial \phi_{23}}{\partial \mathbf{q}_2} \cdot \frac{\partial}{\partial \mathbf{p}_2} [f(2)P(1, 3) + T(1, 2, 3)] \right\}. \end{aligned} \quad (67)$$

Neglecting  $V$  with respect to  $\phi_{12}$  (over the region where the density  $n$  is approximately constant) and assuming the ordering (64), this equation can be considerably simplified and becomes

$$\begin{aligned} & \frac{\partial P(1, 2)}{\partial t} + \left( \frac{\mathbf{p}_1}{m} \cdot \frac{\partial}{\partial \mathbf{q}_1} + \frac{\mathbf{p}_2}{m} \cdot \frac{\partial}{\partial \mathbf{q}_2} \right) P(1, 2) - \frac{\partial \phi_{12}}{\partial \mathbf{q}_1} \cdot \left[ f(2) \frac{\partial f(1)}{\partial \mathbf{p}_1} - f(1) \frac{\partial f(2)}{\partial \mathbf{p}_2} \right] \\ & = \int d\mathbf{x}_3 \left\{ \frac{\partial \phi_{13}}{\partial \mathbf{q}_1} \cdot \frac{\partial f(1)}{\partial \mathbf{p}_1} P(2, 3) + \frac{\partial \phi_{23}}{\partial \mathbf{q}_2} \cdot \frac{\partial f(2)}{\partial \mathbf{p}_2} P(1, 3) \right\}. \end{aligned} \quad (68)$$

As we shall see, the effect of the non-homogeneous term of order  $\phi_{12}f(1)f(2)$  is similar to that of a test particle with charge  $Q = e$ , while the integral on the r.h.s. represents the collective response of the system

and gives rise to Debye screening of the pair correlation function. Starting from the Boltzmann distribution in the  $N$ -particle phase space, it can be shown that the general form of  $P(1, 2)$  for a locally uniform plasma at thermodynamic equilibrium is

$$P(1, 2) = f_B(1)f_B(2)h(r_{12}), \quad (69)$$

where  $h(r_{12})$  is an arbitrary function of the spatial separation  $r_{12} = |\mathbf{q}_1 - \mathbf{q}_2|$ . Then the previous equation for the pair correlation reads

$$\frac{(\mathbf{p}_1 - \mathbf{p}_2)}{m} \cdot \left\{ \frac{\partial h(r_{12})}{\partial \mathbf{q}_1} + \frac{1}{T} \frac{\partial \phi_{12}}{\partial \mathbf{q}_1} \right\} = -\frac{n}{T} \int d\mathbf{q}_3 \left\{ \frac{\mathbf{p}_1}{m} \cdot \frac{\partial \phi_{13}}{\partial \mathbf{q}_1} h(r_{23}) + \frac{\mathbf{p}_2}{m} \cdot \frac{\partial \phi_{23}}{\partial \mathbf{q}_2} h(r_{13}) \right\}. \quad (70)$$

This must be valid for all  $\mathbf{p}_1$  and  $\mathbf{p}_2$  and, in particular, when  $\mathbf{p}_2 = 0$  we obtain

$$\frac{\partial h(r_{12})}{\partial \mathbf{q}_1} + \frac{1}{T} \frac{\partial \phi_{12}}{\partial \mathbf{q}_1} = -\frac{n}{T} \int d\mathbf{q}_3 h(r_{23}) \frac{\partial \phi_{13}}{\partial \mathbf{q}_1}. \quad (71)$$

Taking the divergence of both sides with respect to  $\mathbf{q}_1$  yields

$$\nabla_1^2 h(r_{12}) - \frac{4\pi e^2}{T} \delta(r_{12}) = \frac{4\pi n e^2}{T} \int d\mathbf{q}_3 h(r_{23}) \delta(r_{13}) = \frac{4\pi n e^2}{T} h(r_{12}), \quad (72)$$

which has the same form of Eq. (53) for the density perturbation at large distances from a test charge. The solution is therefore

$$h(r_{12}) = -\frac{e^2}{T} \frac{e^{-(r_{12}/\lambda_D)}}{r_{12}} \quad (73)$$

and we see that there can be Debye screening for the spatial part  $h(r_{12})$  of the pair correlation function, provided the density  $n$  remains approximately constant over the local Debye length  $\lambda_D = \sqrt{T/4\pi n e^2}$ . As we know, this is possible only in the bulk of a space-charge dominated plasma.

Comparing this result to the asymptotic form (63), valid for small spatial separations  $r_{12}$ , we conclude that

$$h(r_{12}) = \frac{P(1, 2)}{f(1)f(2)} \sim \exp\left(-\frac{\ell_T}{r_{12}} e^{-(r_{12}/\lambda_D)}\right) - 1. \quad (74)$$

Therefore, replacing  $r_{12}$  by the average inter-particle distance  $\bar{\ell}$  and neglecting Debye screening yields Eq. (65). On the other hand, when  $\ell_T \ll \bar{\ell}$ , the Debye damping factor  $\exp(-\bar{\ell}/\lambda_D)$  would not play any significant role, since

$$\frac{\bar{\ell}}{\lambda_D} = \frac{n^{-1/3}}{\sqrt{T/4\pi n e^2}} = \sqrt{4\pi \frac{\ell_T}{\bar{\ell}}} \ll 1. \quad (75)$$

In plasma physics, one usually defines the *discreteness parameter*  $g$  as the inverse of the number of particles contained in a sphere of radius equal to the Debye length  $\lambda_D$ . When this parameter is small, the cluster expansion becomes an expansion in powers of  $g$  and the Vlasov equation is the zero-order result of this procedure. This gives the (wrong) impression that the ordering (64) be a consequence of Debye screening. On the contrary, we can express  $g$  in terms of the more fundamental ratio  $\ell_T/\bar{\ell}$

$$g = \frac{1}{n\lambda_D^3} = \left(\frac{\bar{\ell}}{\lambda_D}\right)^3 = \left(4\pi \frac{\ell_T}{\bar{\ell}}\right)^{3/2}, \quad (76)$$

showing again that the Vlasov equation can be formally derived even in the absence of Debye screening. For example, assuming a number of particles per bunch  $N \sim 10^{12}$  in the case of LEP at 50 GeV, the particle density in the rest frame of the beam is  $n \sim 10^8 \text{ cm}^{-3}$  and, according to the discussion of Sec. 5.2, the local Debye length is of the order of  $10^3$  times the vertical beam size in the regular arc, i.e.  $\lambda_D \sim 50 \text{ cm}$ . Therefore we obtain  $g \sim 10^{-13}$ , corresponding to a ratio  $\ell_T/\bar{\ell} \sim 10^{-10}$ .

We now review the physical picture emerging from the discussion on the pair correlation function and the origin of the time scale limitations associated with the truncation of the BBGKY hierarchy. The Vlasov equation describes the motion of a (representative) charged particle in the self-consistent field created by all the other particles, considered as a smoothed out fluid. The most obvious way this representative particle can detect the granularity of the 'fluid' is by a direct collision with another particle, leading to a significant deflection of its trajectory and thus establishing some degree of correlation between them. However this is possible only for a close binary encounter, at a distance  $\ell_T$  such that the Coulomb potential energy  $e^2/\ell_T$  be comparable to the average kinetic energy  $T$  of the representative particle. In other words,  $\ell_T$  should be of the order of the distance of closest approach between thermal particles. Now, if the average inter-particle distance  $\bar{\ell}$  is much larger than  $\ell_T$ , these close encounters are very unlikely and it takes a long (relaxation) time before the representative particle can react to the granularity of the fluid. Note that this argument does not require any screening effect, since the average Coulomb interaction between particles with spatial separations much larger than  $\ell_T$  is already included in the self-consistent field.

However, there is a more efficient mechanism for the granularity of the 'fluid' to show up, similar to the cumulative effect produced on a Brownian particle by the weak, random collisions with many neighbouring atoms or molecules. Owing to the long range nature of the Coulomb interaction, the cumulative effect of weak collisions decreases only as the inverse of the distance and, compared to the relaxation time for strong binary collisions, that for weak collisions can be shown to be shorter by a factor  $\sim 8 \ln(\ell_{\max}/\ell_T)$  [4]. Here we must rely on the existence of a maximum spatial separation between particles,  $\ell_{\max}$ , to cut-off the logarithmic divergence associated with the Rutherford cross section for Coulomb scattering. In an ordinary plasma, the cut-off is provided by Debye screening and thus the maximum separation is  $\ell_{\max} \simeq \lambda_D$ ; the same is true for a non-neutral plasma dominated by space-charge effects. In the case of a low-current storage ring, however, the local Debye length is much larger than the transverse beam size and it is the latter to play the role of a cut-off length; the corresponding relaxation time is therefore longer and the Vlasov equation applies over a wider time scale.

\* \* \*

## REFERENCES

- [1] S.G. Brush, Statistical Physics and the Atomic Theory of Matter (from Boyle and Newton to Landau and Onsager), (Princeton Press, Princeton, 1983).
- [2] L.D. Landau and E.M. Lifshitz, Kinetic Theory (Pergamon Press, London, 1981).
- [3] D.C. Montgomery and D.A. Tidman, Plasma Kinetic Theory (McGraw-Hill, New York, 1964).
- [4] T.J.M. Boyd and J.J. Sanderson, Plasma Dynamics (Nelson, London, 1969).
- [5] C.N. Lashmore-Davis, Proc. 1985 CAS CERN Accel. School on Advanced Accel. Physics, CERN 87-03 Vol. I (1987) p. 219.
- [6] A. Hofmann, Proc. 1987 CAS CERN Accel. School, Second Advanced Accel. Physics Course, CERN 89-01 (1989) p. 25.
- [7] D. Habs, Proc. 1986 School on Frontiers of Particle Beams, South Padre Island, Lecture Notes in Physics, Vol. 296, Springer-Verlag (1988) p. 310.
- [8] R.C. Davidson, J. Plasma Phys. 6, 229 (1971).
- [9] F. Ruggiero, E. Picasso and L.A. Radicati, Ann. Phys. (N.Y.) 197, 396 (1990).

## LANDAU DAMPING

*R.D. Kohaupt*

Deutsches Elektronen-Synchrotron (DESY), Hamburg, Fed. Rep. Germany

### ABSTRACT

Landau damping plays a fundamental role in storage ring physics. In this paper the theoretical background is described in a framework of transfer functions, the basic tool for instability considerations.

#### 1. FUNDAMENTAL CONSIDERATIONS

The origin of the confusion concerning Landau damping is its history. Landau damping was discussed for the first time by Landau in 1946 in plasma physics. In the meantime many authors worked on it in different fields of physics. Therefore everything which has "to do" with the original paper can be called Landau damping.

On the other hand Landau damping has nothing to do with "normal" damping (by friction, etc.). The physical effect of Landau damping is concerned with the decay of coherence in an assembly of oscillators with different frequencies. Let us call "Landau damping" only those effects of collective systems appearing in accelerator physics.

#### 2. THE FRAMEWORK OF STABLE AND UNSTABLE LOOPS

##### 2.1 Transfer Functions and Stability

We consider Fig. 1 where a particle bunch moves from the right to the left.

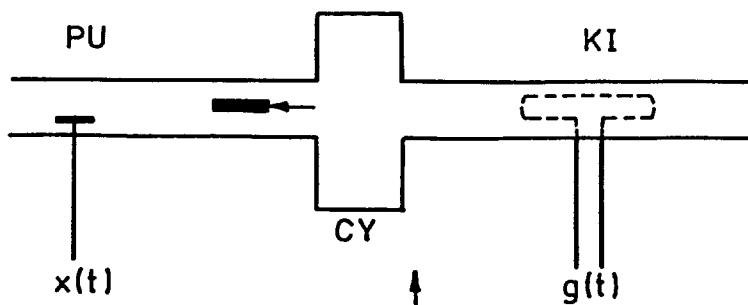


Figure 1: Bunch passage

A kicker kicks the bunch, and after passing the cavity CY we measure the transverse displacement  $x(t)$  of the bunch.

We assume

$$x(t) \equiv 0 , g(t) \equiv 0 \text{ for } t \leq 0 \quad (1)$$

Then  $x(t)$  is a linear response of  $g$ , which can be written as

$$x(t) = \int_0^\infty dt' T(t, t') g(t') \quad (2)$$

where  $T(t, t')$  is the transfer function of the system.

Let us impose "boundary conditions" on  $T(t, t')$ . After a change of variables  $t, t' \rightarrow t; t - t'$  we require that:

- 1) T is a real function
  - 2) T is stationary:  $T = T(t - t')$
  - 3) T is causal:  $T(t - t') \equiv 0$  for  $t - t' \leq 0$ .

In addition, in accelerator physics it is reasonable that

$$g(t) \rightarrow 0 \quad \text{and} \quad |x(t)| < e^{\alpha t}, \quad \alpha \geq 0 \quad (3)$$

for  $t \rightarrow \infty$  .

Because of the properties (1), (2), (3) we can introduce the general Fourier transform:

$$F[g(t)] \rightarrow \bar{g}(z) ; \quad F[T] \rightarrow \bar{T}(z) ; \quad F[x(t)] \rightarrow \bar{x}(z) \quad (4a)$$

$$\text{so that} \quad \tilde{x}(z) = \tilde{T}(z)\tilde{g}(z) \quad (4b)$$

holds.

Explicitely we define

$$\tilde{x}(z) = \frac{1}{\sqrt{2\pi}} \int_0^\infty dt x(t) e^{-izt}, \quad Im z \leq -\alpha \quad (5a)$$

$$\text{and } x(t) = \frac{1}{\sqrt{2\pi}} \int_K dz \tilde{x}(z) e^{izt}, \quad t \geq 0 \quad (5b)$$

where the path of integration goes along a line in the complex  $z$ -plane parallel to the real axis with  $\operatorname{Im} z < -\alpha$ .

As a consequence, mainly of the first principles, the functions  $\tilde{x}(z)$ ,  $\tilde{T}(z)$ ,  $\tilde{z}(z)$  are analytic functions of  $z$  at least for  $\operatorname{Im} z \leq -\alpha$ .

From reality it follows that

$$\bar{x}(-z^*) = \bar{x}^*(z) \quad . \quad (6)$$

Due to the analytic properties (as a consequence of the first principles) the time behaviour of  $x(t)$  is completely determined by the singularities of  $\tilde{x}(z)$  in the complex  $z$ -plane.

Let us consider the singularity structure of functions which occur in accelerator physics (Fig. 2).

These singularities are classified by:

- 1) poles (of first order) at  $z_k$
- 2) a cut along the real axis.

The poles in the upper half-plane are connected with stable (damped) physical objects.

For  $\operatorname{Im} z < -\alpha$ ,  $\tilde{x}(z)$  is regular as already pointed out. In order to obtain  $x(t)$  we have to perform the integration (5b). Instead of the path  $K$  we consider the "closed loop" integration along  $C$ , because function theory tells us the result:

$$\int_C = \int_K + \int_{\text{cut}^-} + \int_\infty = 2\pi i \sum_k R_k e^{iz_k t} \quad (7)$$

where  $R_k$  are the residua at the poles. The integral along the cut can be rewritten as

$$L(t) = \int_{\text{cut}^-} = \int_a^b d\omega (\tilde{x}^+(\omega) - \tilde{x}^-(\omega)) e^{i\omega t} \quad (8)$$

where  $\tilde{x}^+$ ,  $\tilde{x}^-$  are the "upper" and "lower" limits of  $\tilde{x}$  at the cut, being continuous functions of  $\omega$  along the real axis. Therefore (8) is a final Fourier integral with the property

$$L(t) \rightarrow 0 \quad \text{for} \quad t \rightarrow \infty . \quad (9)$$

The integral  $\int_\infty$  along the "infinite bow" vanishes if we assume  $|\tilde{x}(z)| \rightarrow 0$  for  $|z| \rightarrow \infty$ .

Then from (7) and (8) it follows that

$$x(t) = 2\pi i \sum_k R_k e^{iz_k t} + L(t) . \quad (10)$$

Therefore, whether the system is stable or not depends on whether the poles  $z_k$  are located in the upper or lower complex half-plane.

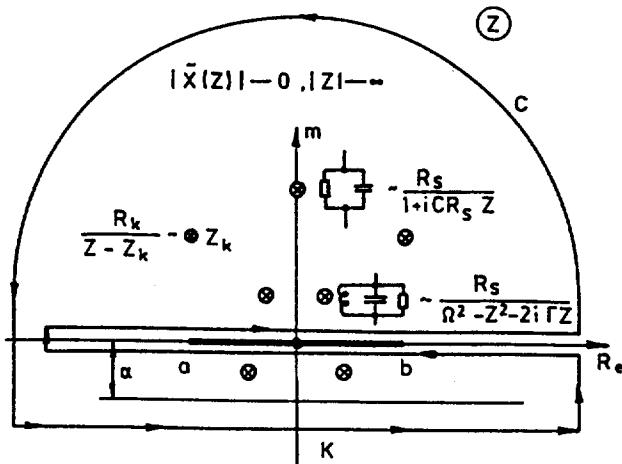


Figure 2: Singularities in the complex  $z$ -plane

## 2.2 Equations of Motion and Stability

Let us consider an equation of motion of the following general form:

$$(\underline{D} + k^2) \cdot x = T \cdot x + g \quad (11)$$

where  $\underline{D}$  denotes a second order differential operator in time. The transfer  $T \cdot x$  describes a term generated by  $x$  and acting back on the system. This forms a typical closed loop leading to instabilities. The function  $g$  denotes an external force.

Before we can take the generalized Fourier transform of (11) we need the corresponding rules to proceed. From mathematics we obtain

$$\begin{aligned} F(x) &= \ddot{x}(z) \\ F(\dot{x}) &= iz\ddot{x}(z) - x(0) \\ F(\ddot{x}) &= -z^2 \ddot{x}(z) - izx(0) - \dot{x}(0) . \end{aligned} \quad (12)$$

According to (1) and the properties of  $T(t - t')$  we can drop all terms containing  $x(0)$ ,  $\dot{x}(0)$  etc.

The transformation of (11) into the generalized Fourier representation yields, after we have resolved for  $\ddot{x}(z)$ ,

$$\ddot{x}(z) = \frac{g(z)}{\tilde{D}(z) + k^2 - \tilde{T}(z)} . \quad (13)$$

Taking  $\tilde{g}(z) = \text{const.}$  which corresponds to a  $\delta$ -kick in time, the poles of (13) are determined by

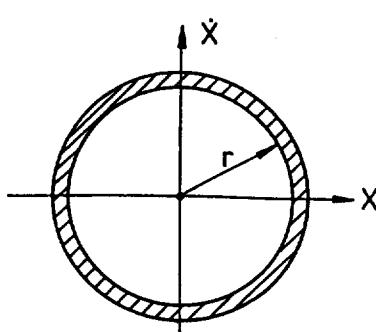


Figure 3: Particle motion in phase space

the zeros of the denominator

$$\tilde{D}(z) + k^2 - \tilde{T}(z) = 0 . \quad (14)$$

As a simple example we consider a linear oscillator being influenced by a driving term  $\tilde{G}(z)$  and a feedback term  $\tilde{F}(z)$  according to

$$\tilde{T}(z) = \tilde{G}(z) + \tilde{F}(z) . \quad (15)$$

Then following (13) we obtain simply

$$-z^2 + k^2 = \tilde{G}(z) + \tilde{F}(z) . \quad (16)$$

Equation 16 shows that generally  $\tilde{G}(z)$  and  $\tilde{F}(z)$  together determine the solution. Thus, depending on the special properties of  $\tilde{G}$ ,  $\tilde{F}$ , there is an influence of one term on the other, which means that the instability and feedback system cannot be considered separately.

### 2.3 Instability and Landau Damping

We consider transverse dipole oscillations in the transverse phase space  $x, \dot{x}$  (fig. 3).

The center-of-mass motion of particles contained in a ring with radius  $r$  is denoted with  $\Delta(r, t)$ , the tune of the ring is  $Q(r)$ . The density  $\rho(r)$  is normalized according to

$$\int dr r \rho(r) = 1 . \quad (17)$$

The dipole moment can then be defined as

$$d(t) = \int dr r \rho(r) \Delta(r, t) . \quad (18)$$

The interaction of the beam with the vacuum chamber is assumed to be

$$W(t) = \int dt' G(t - t') d(t') . \quad (19)$$

Then the equation of motion reads

$$\ddot{\Delta}(r, t) + Q^2(r) \Delta(r, t) = W(t) + g(t) . \quad (20)$$

Passing to the generalized Fourier transform yields

$$(Q^2(r) - z^2) \tilde{\Delta}(r, t) - \tilde{G}(z) \tilde{d}(z) = g(z) . \quad (21)$$

From (18) we obtain of course:

$$\tilde{d}(z) = \int dr r \rho(r) \tilde{\Delta}(r, z) . \quad (22)$$

Using (21) and (22) we can resolve for  $\tilde{d}(z)$

$$\tilde{d}(z) = \frac{\int dr r \frac{\rho(r)}{Q^2(r)-z^2} \cdot \tilde{g}(z)}{1 - \tilde{G}(z) \int dr r \frac{\rho(r)}{Q^2(r)-z^2}} . \quad (23)$$

The time dependent  $d(t)$  is then defined by

$$d(t) = \frac{1}{\sqrt{2\pi}} \int_K dz \tilde{d}(z) e^{izt} . \quad (24)$$

Since  $d(t)$  is completely determined by the singularity structure of  $\tilde{d}(z)$ , we first consider the properties of the integral

$$S(z) = \int dr r \frac{\rho(r)}{Q^2(r) - z^2} . \quad (25)$$

We find:

- 1)  $S(z) \rightarrow 0$  for  $|z| \rightarrow \infty$
- 2)  $S(z) \rightarrow S^\pm(\omega)$  for  $z \rightarrow \omega \pm i\delta$
- 3)  $S^\pm(\omega)$  are continuous functions of  $\omega$ .

Therefore  $\tilde{d}(z)$  has a cut along the real axis determined by  $S(z)$ .

Possible poles occur for

$$\tilde{G}(z) S(z) = 1 . \quad (26)$$

For a finite frequency spread, (26) implies that  $S(z)$  has only finite (maybe large) values.

Therefore, for sufficiently small values of  $\tilde{G}(z)$ , (27) has no solution, which means that  $\tilde{d}(z)$  has no poles, and therefore no instabilities. This is the effect of Landau damping that we use in accelerator physics: the system is not able to organize a common or coherent action.

The threshold for instabilities is given for those  $\tilde{G}$   $\tilde{T}(z)$  which just solve Eq. (27), see Ref. 1.

#### 2.4 Landau Damping as a "Cure"

- a) Consider the case  $\tilde{G} \equiv 0$  (no interaction).  
Then  $d(t)$  decays in time according to

$$d(t) \sim \int d\omega (S^+(\omega) - S^-(\omega)) e^{i\omega t} \quad (27)$$

This decay is not necessarily exponential. Nevertheless, we can define a decay time  $\tau_L$  by

$$|d(\tau_L)| = \frac{|d(0)|}{2} \quad (28)$$

- b) Consider the case of no frequency spread. Then we may have poles

$$z = \pm \omega_o + i \delta_G \quad (29)$$

where  $\delta_G$  denotes a growth rate.

Usually one compares  $1/\tau_L$  and  $\delta_G$  in order to find out whether the system is stable or not. This generally makes no sense. Including also a feedback term  $\tilde{F}(z)$  we necessarily have to consider

$$(\tilde{G}(z) + \tilde{F}(z)) S(z) = 1 \quad (30)$$

since all three effects interfere in general.

The correct discussion of Eq. (31) ensures that the applied cures do not fail for the real machine.

\* \* \*

REFERENCE

1. *R.D. Kohaupt*, DESY M-86-02, 1986



## ION TRAPPING AND CLEARING

*A. Poncet*

CERN, Geneva, Switzerland

### ABSTRACT

After a brief review of the adverse effects of neutralization in particle storage rings, the basic topics of ion production by ionisation of the residual gas are recalled: ion production rate, natural clearing rates, ion kinematics and conditions of trapping for bunched and unbunched particle beams with positive or negative space charge. Different methods of clearing are described and their performance discussed, namely, DC clearing electrodes, empty buckets (in electron storage rings) and beam shaking. Examples of neutralization effects and diagnostics are taken from CERN machines, such as ion clearing current measurements on electrodes, tune shifts and emittance measurements.

### 1. INTRODUCTION

In accelerators and storage rings, ions created by the circulating particles from neutral molecules of the residual gas may be trapped in the beam-space-charge potential and may generate all sorts of ill effects: reduced beam lifetime (increased pressure), emittance growth and losses through excitation of resonances, and coherent beam instabilities. While they can occur in proton beams (e.g. CERN ISR trapping electrons), these neutralization phenomena mainly affect machines with negative beam space charge such as electron storage rings and antiproton accumulators.

Low energy machines are more subject to ion trapping because of their small size, which leaves little space between bunches for ions to escape the beam potential, and suffer most because of their inherent high sensitivity to space charge effects. To illustrate this point, the incoherent space charge tune shift can be written as [1]:

$$\Delta Q = \frac{N}{k} \left( n - \frac{1}{\gamma^2} \right)$$

(where  $k$  relates to beam transverse dimensions and bunching factor,  $\gamma$  is the relativistic factor,  $n$  the neutralization coefficient defined as the ratio of trapped charges to beam charges, and  $N$  the number of particles in the beam).

$\Delta Q$  can be unacceptably large if  $\gamma$  is small (low energy) and/or  $n$  is high. For instance  $\Delta Q$  is  $\sim 1/n$  in the CERN 600 MeV EPA (Electron Positron Accumulator [2]) for its nominal electron beam, where in the absence of clearing  $n$  can reach values close to one. This has to be compared with the value  $\Delta Q \leq 0.1/\gamma^2$ , i.e.  $\sim 10^{-7}$  for perfect clearing. A large tune shift

is accompanied by a large tune spread, owing to the non-linearity of the ion focusing forces on the beam particles. This results in the excitation of a large number of resonances, as can be usually seen in the tune diagram. Figure 1 illustrates this effect as seen from the variation of equilibrium emittances with the beam intensity in EPA.

A good description and bibliography on neutralization can be found in reference 1. In a one hour lecture, we can only attempt to describe the main aspects of ion trapping and clearing, with some emphasis on recent developments such as resonant beam shaking and ion-clearing-current diagnostics. Therefore, this lecture has to be understood as being complementary to the one given in reference 1.

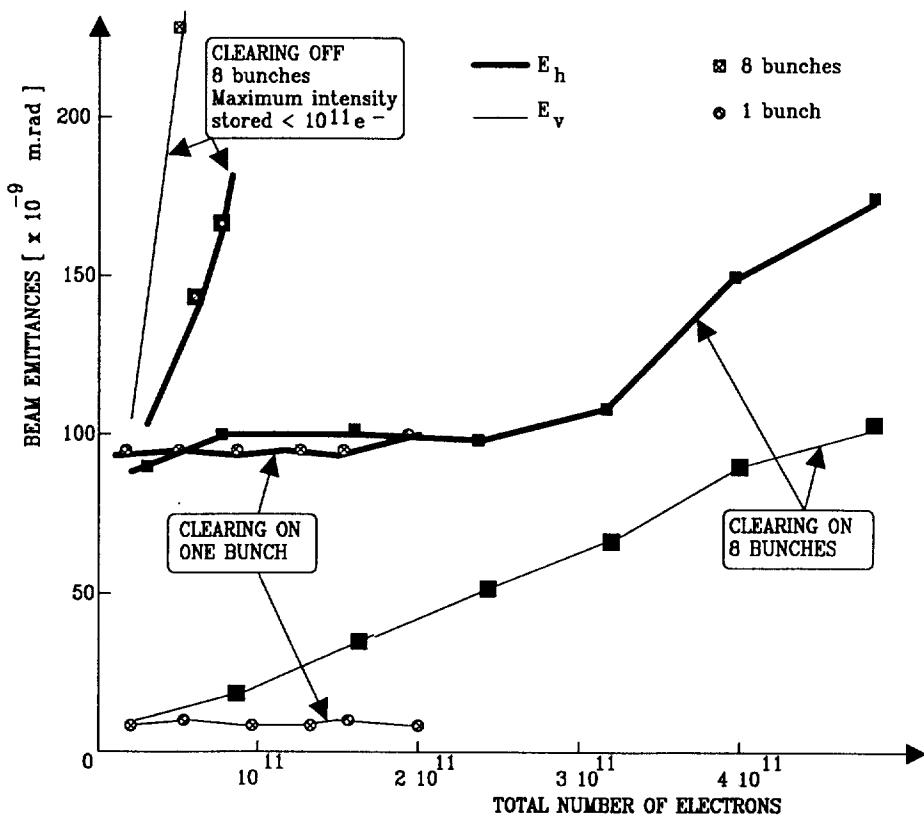


Fig. 1 EPA emittances with and without clearing

## 2. ION PRODUCTION

Beam particles with relativistic velocities interact with the nuclei and the electrons of the gas atoms via the Coulomb forces. The energy transfer  $\Delta E(b)$  in a collision, which depends on the impact parameter  $b$  (distance between target and projectile), may be sufficiently large to break the liaison between the nucleus and the electrons, leading to

ionisation:

$$\Delta E(b) > Z' \cdot 13.5 \text{ eV} \text{ (ionisation potential } E_i) .$$

In S.I. units [3]:

$$\Delta E(b) = \frac{e^4}{8\pi^2 \epsilon_0^2 \beta^2} \cdot \frac{Z^2 Z'^2}{m_0} \cdot \frac{1}{b^2} \text{ (Joule)} \quad (1)$$

where  $Z$  and  $Z'$  are respectively the charge numbers of the projectile and the target,  $m_0$  the mass of the target,  $e$  the elementary charge,  $c$  the speed of light, and  $\epsilon_0$  the vacuum permittivity. If the target is an electron:  $m_0 = m_e$ ,  $Z'=1$ ; for a nucleus  $m_0 \sim 2m_p Z'$ .

One sees from (1) that at identical impact parameter:

$$\Delta E(b)_{\text{electron}} = \frac{2m_p}{m_e} \cdot \frac{1}{Z}, \quad \Delta E(b)_{\text{nucleus}} \gg \Delta E(b)_{\text{nucleus}} \left( \frac{m_p}{m_e} \sim 1836 \right) . \quad (2)$$

This shows that collisions with orbital electrons are the main cause of energy loss, and therefore ionisation rate calculations need only consider electrons. Usually expression (1) is integrated over a range of possible impact parameters:

$$\frac{dE}{dt} = \int_{b_{\min}}^{b_{\max}} \Delta E(b) [2\pi b \beta c N_a Z' db]$$

where the quantity between brackets represents the number of electrons at distance  $b$  from the projectile during time  $dt$  and  $N_a$  is the density of atoms of charge  $Z'$ .

By consideration of the collision time with the orbital period of electrons,  $b_{\max}$  is obtained from [3]:

$$b_{\max} = \frac{\gamma \beta c h}{E_i}$$

where  $h$  is the Planck constant =  $4.14 \times 10^{-15}$  eV·s, and  $E_i = Z' \cdot 13.5$  eV.  $b_{\min}$ , giving the maximum energy transfer for trappable particles, namely the beam space charge potential  $U$ , is obtained from expression (1):

$$b_{\min} = \frac{r_0 c Z}{\beta} \left( \frac{2m_0 Z'^{1/2}}{U} \right)$$

since species created with  $\Delta E(b) > U$  can escape to the vacuum chamber wall, and are therefore not trappable.

Table 1

Values of parameters  $r_0$ ,  $m_0$ ,  $Z'$  for different beam trapping events.

$r_p$  and  $r_e$  are the classical proton and electron radii ( $r_{p,e} = \frac{e^2}{4\pi\epsilon_0 m_{p,e} c^2}$ )

Trappable particle	Electron (positive ion beam)	Proton ( $e^-$ or antiproton beam)	Ion ( $e^-$ or antiproton beam)
$r_0$	$r_e$	$r_p$	$r_p$
$m_0$	$m_e$	$m_p$	$2m_p$
$Z'$	1	1	$Z'$

The production rate is therefore:

$$R_p = \frac{1}{E_0} \frac{dE}{dt} = \frac{2\pi m_e c^3 r_e^2}{\beta} \cdot \frac{N_a Z'}{E_0} \cdot \ln \left( \frac{\gamma \beta^2 h}{E_1 r_0 Z} \left( \frac{U}{2m_0 Z'} \right)^{1/2} \right) \quad (3)$$

$E_0$  being the average energy for the formation of an ion-electron pair ( $\sim 35$  eV).

A consequence of this expression is that — everything being equal (energy, beam potential) — a positive space charge beam (e.g. protons) will trap less electrons than a negative one will trap positive ions, since the fraction of electrons born in the ionisation process with sufficient energy to escape the beam potential is larger.

As an example, the logarithmic term in expression (3) is:

For EPA:

$\gamma = 1200$ ,  $U = 10$  Volts

positrons:  $\ln(\quad) = 18$ ; electrons  $\ln(\quad) = 22$

(i.e. the positron beam traps  $\sim 19\%$  electrons less than an electron beam traps protons).

For the CERN Antiproton Accumulator (AA) [4]:

$\gamma \sim 3.7$ ,  $U = 10$  Volts

proton:  $\ln(\quad) \sim 12$ ; antiprotons:  $\ln(\quad) \sim 18$

i.e. a proton beam used for AA tests will trap  $\sim 35\%$  less than an antiproton beam. This proportion is roughly confirmed by clearing current measurements (sum of the currents drawn from the clearing electrodes, i.e. the number of charges created by ionisation, per second).

This consideration, together with expression (2), illustrates the fact that positive ions are created with much less energy (in fact with quasi-thermal energies  $< 0.01$  eV) than

electrons, and are therefore generally all trappable. Indeed a quick numerical application of expression (1) would show that the impact parameter has to be  $4.7 \times 10^{-14}$  m - i.e. a quasi head-on collision with the nucleus, and therefore highly improbable — for the  $H_2^+$  ion to be born with 10 eV energy (the electron would obtain 20 keV).

The neutralization coefficient of a beam is the ratio of the ion production rate  $R_p$  to the clearing rate  $R_c$  ( $s^{-1}$ ). Since the ions are virtually born at rest, the production rate of ions can therefore be obtained from experimentally determined ionisation cross sections  $\sigma_i$ :

$$R_p = \sum_{i=1}^n \sigma_i N_{mi} \beta c$$

$N_{mi}$  being the molecular density of gas species i.

As a typical example, the values for EPA (600 MeV,  $\gamma = 1200$  for electrons) at  $P = 10^{-9}$  mbar [5] are shown in Table 2.

Table 2  
EPA ionisation cross sections

Gas	$\sigma_i$ ( $m^2$ )	$R_p$ ( $s^{-1}$ )	Ionisation time (s)
$H_2$	$0.4 \cdot 10^{-22}$	0.4	2.5
CO	$1.54 \cdot 10^{-22}$	1.5	0.7

Due to a logarithmic dependance on the energy of the primary particle, these cross sections would only grow by 20% between 0.6 and 6 GeV/c.

## 2.1 Beam heating

Distant collisions with large impact parameter — much more probable than close ones leading to ionisation — are important, since they feed energy differentially to ions. In some circumstances (neutralization pockets) this may be a clearing mechanism, i.e. when the trapped species get enough energy to escape the beam potential:

$$R_c = \frac{1}{eU} \frac{dE}{dt} = \frac{1}{eU} \int_{b_{min}}^{b_{max}} \Delta E(b) (2\pi b \beta c N_p db) .$$

This represents the "natural" clearing rate for a singly charged species. The expression between brackets is the number of projectiles passing at distance b to the ion target during time dt.  $N_p$  is the projectile density of charge Z = 1.

To good approximation,  $b_{\max}$  and  $b_{\min}$  can be chosen to have the same values as the ion and nucleus radii respectively, leading to [3]:

$$R_c = \frac{2\pi m_0 c^3 r_0^2}{\beta} \cdot \frac{N_p Z'}{eU} \cdot \ln(3 \cdot 10^4 \cdot Z'^{-2/3}) \quad (4)$$

with  $m_0$  and  $r_0$  being  $m_e$ ,  $r_e$  if the trapped species is an electron;  $m_p$ ,  $r_p$  for a proton;  $2m_p$ ,  $r_p$  for an ion of charge  $Z'$ .

As an example, typical clearing times for the EPA machine with  $6 \cdot 10^{11}$  electrons (300 mA), 1 mm beam radius, giving a beam potential of  $\sim 50$  V, are shown in Table 3.

**Table 3**  
EPA natural ion clearing rates

Clearing rate $R_c$ ( $s^{-1}$ )	Clearing Time $\tau_c$ (s)
$H^+$	$3 \cdot 10^{-3}$
$H_2^+$	$6 \cdot 10^{-3}$
$CO^+$	0.04
$CO_2^+$	0.07

The process is thus slow compared to typical ionisation rates. But it may be important to explain why in some circumstances, e.g. pockets with very low gas pressure ( $5 \cdot 10^{-11}$  Torr, see AA case in reference [6]) fully ionised light ions can accumulate to a dangerous level, as they replace heavy ones which are chased away by beam heating.

## 2.2 Gas Cooling

Seldom taken into account, gas cooling could perhaps be an important process for high pressures and long ion resting times. Charge-exchange phenomena by which a positive ion captures an electron from a gas molecule may occur at ion energies of only a few eV. The new ion is created with the primary molecule's energy, while the newly created neutral species carries away the initial ion energy. Resonant capture cross sections between an ion and its own neutral molecule can be very high at low energy:

$$\sigma = 1.2 \times 10^{-15} \text{ cm}^2 \text{ for } H_e^+ \text{ in } H_2$$

for an ion energy of 3 eV. With this cross section it would take only 27 seconds at  $10^{-8}$  mbar for an ion to capture an electron and thus leave a cool ion behind. Capture cross sections are even higher for heavy gas species [7].

### 2.3 Limits on ion accumulation

In the vast majority of cases (electron storage rings with typical pressures of  $10^{-9}$  mbar and ionisation times of a second or less), ionisation is the dominant effect in the absence of any clearing mechanism. The production rates are:

for singly ionised species (density  $N_i^+$ ):  $\frac{d(N_i^+)}{dt} = N_m N_p \sigma_i c - N_i^+ N_p \sigma_i c = N_p \sigma_i c (N_m - N_i^+)$

doubly ionised:  $\frac{d(N_i^{++})}{dt} = N_i^+ N_p \sigma_i c - N_i^{++} N_p \sigma_i c = N_p \sigma_i c (N_i^+ - N_i^{++})$

etc. until ( $Z'$  being the total number of electrons of the gas atom):

$$\frac{d(N_i^{Z' +})}{dt} = N_i^{(Z' - 1) +} N_p \sigma_i c .$$

In the steady state (assuming constant ionisation cross sections):

$$N_i^+ = N_i^{++} = \dots = N_i^{(Z' - 1) +} < N_m .$$

Therefore partially ionised ions can, at most, reach the molecular density  $N_m$ . Only the fully ionised state  $N_i^{Z' +}$  could get close to the particle density (usually much larger than the gas density) divided by the final charge state:  $N_i^{Z' +} < N_p/Z'$  corresponding to full neutralization of the particle beam. ("Over" neutralization is excluded, since the resulting potential would chase the species to the chamber wall.)

The degree of neutralization of a particular beam can be estimated from the incoherent tune shift. Almost full neutralization has been measured on the CERN AA when all the clearing electrodes are turned off (Fig. 2).

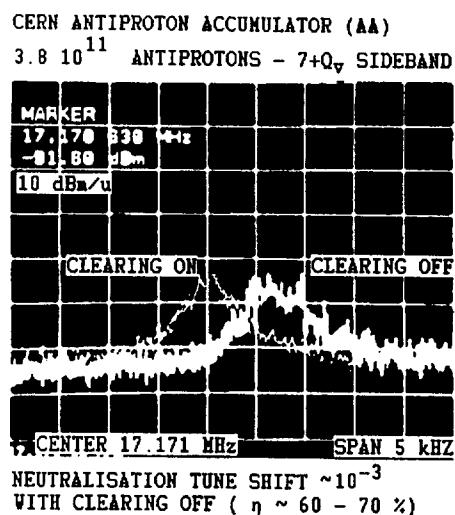


Fig. 2 AA transverse vertical Schottky scan showing the band  $(7 + Q_v)f_{rev}$  with clearing electrodes turned OFF (full neutralization) and ON (little neutralization)

### 3. ION DYNAMICS (DC BEAM)

Born with near thermal velocities, ions are generally not free to simply drift in the beam. Their motion is mostly governed by the beam space charge potential when the beam is not fully neutralized, and by external forces such as magnetic fields in dipoles and quadrupoles. The Lorentz force due to the beam magnetic field is weak and can be neglected in practical cases.

#### 3.1 Ion oscillatory motion and azimuthal drift in magnetic field-free regions

The transverse distribution of beam charge results in an electric field  $E$  to which the charged trapped species is sensible. This centripetal force (directed towards the centre of the beam) provokes a radial oscillatory motion with an amplitude equal to the radius at birth. Its frequency is proportional to the square root of the local field derivative times the ion charge, divided by its mass (plasma frequency).

The longitudinal modulation due to changing beam sizes and the varying vacuum chamber dimensions give rise to longitudinal fields which drive the trapped ions around the machine. Usually negligible, the effect of the beam magnetic field may also contribute to the longitudinal motion for very high intensities. To illustrate this with numbers, we consider the simple case of a round beam in a circular vacuum chamber, with a uniform transverse distribution of charges (Fig. 3).

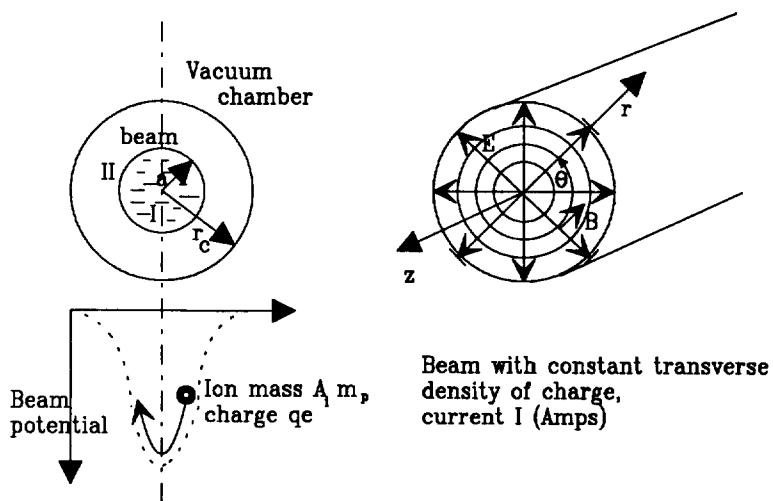


Fig. 3 Beam potential, magnetic and electric fields of a uniform beam

The Lorentz force acting on the ion is:

$$\vec{F} = qe (\vec{E} + \vec{v} \times \vec{B})$$

with components

$$\begin{cases} F_r = qe(E_r - B \cdot \dot{z}) \\ F_\theta = 0 \\ F_z = qe(B \cdot \dot{r} + E_z) \end{cases}$$

From the Gauss theorem, the electric field can be written as

in region I:  $E_r(r) = \frac{I}{2\pi\epsilon_0\beta c} \frac{r}{a^2}$

in region II:  $E_r(r) = \frac{I}{2\pi\epsilon_0\beta c} \frac{1}{r}$ .

The magnetic field is simply  $B = \frac{\beta E_r}{c}$ .

The beam potential at the centre is [1]:

$$U = \frac{I}{2\pi\epsilon_0} \left( \ln \frac{r_c}{a} - \frac{1}{2} \right)$$

The forces on the ion reduce to:

$$\begin{aligned} F_r &= A_i m_p \ddot{r} = \frac{qe I}{2\pi\epsilon_0 \beta c a^2} (1 - \beta \frac{\dot{z}}{c}) r \\ F_z &= A_i m_p \ddot{z} = \frac{qe I}{2\pi\epsilon_0 \beta c^2} \dot{r} r + qe E_z. \end{aligned} \quad (5)$$

As the ions can only have a maximum potential energy equal to the beam potential times their charge, i.e. typically up to a few hundred eV, they are non relativistic. Equations (5) can be uncoupled by neglecting  $\beta \dot{z}/c \ll 1$ . The ion motion is thus transversally oscillatory with a frequency ("plasma" frequency) of:

$$\omega_i = 2\pi f_i = \left( \frac{I}{2\pi\epsilon_0 \beta c} \cdot \frac{1}{a^2} \cdot \frac{qe}{A_i m_p} \right)^{1/2} \quad (6)$$

$$r(t) = r_0 \cos(\omega_i t + \alpha).$$

The second expression (5) can be integrated to yield the longitudinal drift velocity (neglecting some oscillatory terms) [8]:

$$\dot{z}(t) = \dot{z}(0) + \frac{1}{4c} \omega_i^2 r(0)^2 + \frac{qe}{A_i m_p} E_z t.$$

The second term can only be important for very high plasma frequencies (50-100 MHz), i.e. for intense positive beams ( $I > 1 \text{ A}$ ) trapping electrons, as in the CERN ISR. As a numerical example, a 1 A negative beam 5 mm in radius in a 0.16 m diameter vacuum chamber would have a space charge field of 12 kV/m at beam edge, resulting in an  $\text{H}_2^+$  ion plasma frequency of 1.7 MHz. The beam potential at centre would be  $\sim 200 \text{ V}$  with typical azimuthal gradients (longitudinal field), resulting from beam and chamber size variations, of one to several V/m. With 1 V/m, 0.1 ms would be sufficient to get a drift velocity of 5000 m/s, well in excess of the thermal velocity  $v_0$ . This illustrates the fact that the ion motion in field-free regions is governed by the beam space charge. In particular so called neutralization pockets may exist in places where the beam potential is deepest (bellows, chamber enlargements): ions created there do not have enough energy to overcome the potential barriers to eventually reach the nearest clearing electrodes.

Neutralization pockets — or potential barriers — may also be created by highly insulating ceramic vacuum chambers becoming electrostatically charged [6]. Metallization of their inner surface is therefore important, as is the necessity to keep the vacuum chamber cross section as uniform as possible (shielded bellows, screens ...), joining in this respect the conditions imposed by impedance considerations. A precise knowledge of the potential variation around a machine is therefore important, as this also determines the strategic locations of clearing electrodes. Better expressions exists for more precise calculations of the potential of a beam at any location in a rectangular vacuum chamber. For elliptic beams with Gaussian transverse distributions, closed solutions have been given recently for the electric field [9] and the beam potential [10].

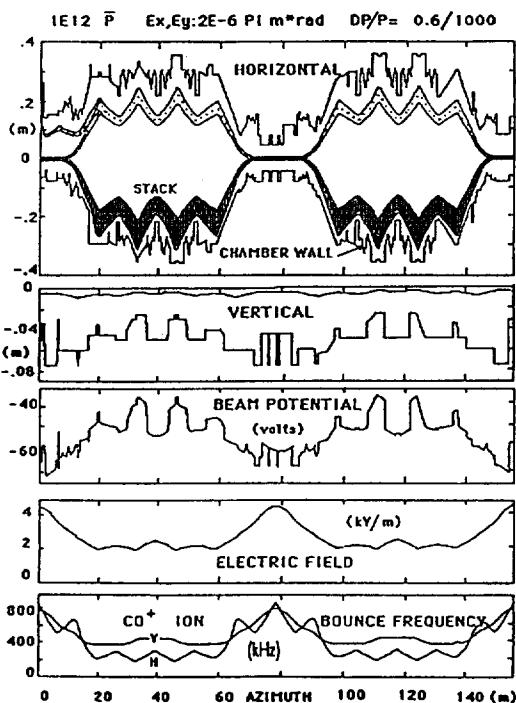


Fig. 4 AA nominal beam and vacuum chamber sizes, beam potential and ion bounce frequencies [11]

Using the beam envelope (Twiss) parameters and the vacuum chamber dimensions as input to a computer program, these formulae can be used to produce a plot of the beam potential, maximum electric field, ion plasma frequencies, etc... around a machine. The example above is representative of the CERN Antiproton Accumulator (Fig. 4).

### 3.2 Drift motion in dipoles and quadrupoles, magnetic mirror (containment effect)

The motion of a non-relativistic trapped species of charge  $qe$  and mass  $A_i m_p$  is well covered in reference 1 and classical electrodynamics textbooks, but will now be summarised for the more usual magnetic fields.

#### 3.2.1 Quadrupole fields (gradient dB/dr)

The main effect is a longitudinal drift (perpendicular to the field lines) with velocity:

$$V_D = W_{kin} \frac{1}{qeB^2} \left( \frac{dB}{dr} \right) \quad (7)$$

neglecting the small additional term due to the curvature of the field lines.  $W_{kin}$  is the ion's kinetic energy perpendicular to the beam axis. For ions born at the centre of the beam, the drift can be very slow indeed.

Example: minimum ion energy (thermal):  $W_{kin} = 4 \cdot 10^{-2}$  eV

(ion born at beam centre)  $B = 1$  T

$$\frac{dB}{dr} = 10 \text{ Tm}^{-1}$$

yields a drift velocity of  $V_D = 0.4$  m/s (for  $W_{kin} = 10$  eV,  $V_d = 400$  m/s).

Because of this possibility of relatively high neutralization, clearing electrodes are sometimes installed in quadrupoles, possibly combined with beam position electrostatic pickups. This is the case of the CERN AA, in which another effect has been experimentally verified in measuring the ion current drawn by these electrodes: namely an enhanced ionisation rate due to a containment effect of electrons. As hinted in reference 1, electrons from the primary ions, of the same charge and thus repelled by the beam, may spiral around the quadrupole field lines towards the poles and be reflected back, in a so-called mirror effect. As already seen, these electrons have sufficient energy above a beam threshold intensity, and long resting times to further ionise gas atoms.

Figure 5 below illustrates this effect as seen from the discharge current on a quadrupole pick up/clearing electrode in the CERN AA.

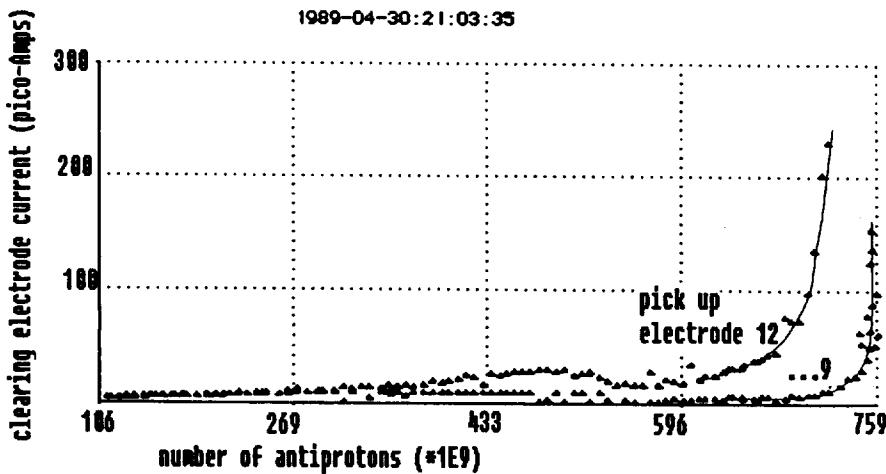


Fig. 5 Clearing electrode currents on two quadrupole clearing electrodes (pick-ups), plotted versus beam intensity in the CERN AA, showing enhanced discharge at high beam intensity

Without clearing, this additional effect could enhance the neutralization coefficient, through an increased ion production rate.

### 3.2.2 Uniform fields ( $B$ )

The cyclotron motion around the field lines has a frequency [1]:

$$\omega_{ci} = \frac{qeB}{A_i m_p} \quad (8)$$

$$\text{with a radius } r_i = \frac{A_i m_p v_\perp}{q e B}$$

$v_\perp$  being the species' velocity perpendicular to the field lines.

The motion of the radius of gyration along the lines of force ( $v_{\parallel i}$ ) is not affected; vertically, in dipoles, the spinning particle behaves as in field-free regions.

The longitudinal cross-field drift velocity, due to the combined action of the beam's space charge field  $E(r)$ , and  $B$  is independent of the charge and mass of the ion to a first approximation:

$$v_D = \frac{E(r)}{B} . \quad (9)$$

If one takes into account the non-linearity of the electric field  $E(r)$ , there is a slight dependence on the ion mass: heavy ions drift out more slowly than light ones [12].

Trapped species drift in opposite directions on each side of the beam, and the velocity at the centre is zero. As an example, for a 0.1 A beam with a radius of  $5 \cdot 10^{-3}$  m in a field of 1 Tesla, the field at the beam edge is  $1.2 \times 10^3$  V/m and the drift velocity 1200 m/s. But the fact that the drift velocity falls to zero at beam centre may result in a relatively high neutralization in dipoles. The recent experiments of beam shaking (see § 6) in the CERN AA supports this fact, as hinted by the behaviour of the clearing currents drawn at the extremities of long bending magnets [13], the only ones around the ring to react to the shaking perturbation. In addition, the observation that neutralization effects can be reduced mainly with a vertical shaking of the beam near ion plasma frequencies (CERN AA and EPA) fits with the hypothesis that it affects ions in dipoles, where the vertical motion is the only degree of freedom that a coherent beam force can excite.

### 3.2.3 Dipole fringe fields

It has been shown recently [14] that slow ions with low transverse kinetic energy (i.e. born near the beam centre) and low longitudinal velocity drifting from no-field regions, can be reflected by the longitudinal gradient  $\partial B / \partial z$  of the dipole field (fringe field). An ion oscillating horizontally in the beam space charge potential and drifting towards a dipole starts a cyclotronic motion around the vertical field lines, with a decreasing radius as it proceeds towards this magnetic field. Motion reversal may occur if after one orbital period the ion has sufficiently escaped the field. Fringe fields can therefore represent potential barriers for trapped species, and straight sections between dipoles must incorporate clearing electrodes. This effect, added to the field reversal, could be very detrimental in terms of ion trapping for undulators in electron storage rings.

## 4. ION DYNAMICS (BUNCHED BEAMS)

### 4.1 Field-free regions

Up to now we have only considered unbunched beams. Bunching obviously does not change the ion production rate but with our neutralization coefficient definition:

$$\text{inside the bunch: } n_{\max} (N_i = N_p) = \frac{1}{B}$$

$$\text{with the bunching factor defined as } B = \frac{2\pi R}{nl_b} \gg 1$$

( $R$  being the machine radius,  $n$  the number of bunches, and  $l_b$  the bunch length).

A trapped ion will therefore be submitted to the repetitive bunch space charge centripetal (focusing) force, independent of neutralization. If the ion motion is slow, and/or the ion heavy, or if the bunch spacing is uniform and small with respect to the

machine circumference, the ion will mainly respond to the DC component of the Fourier expansion of the passing charge. In this case the ion dynamics will be that of an unbunched beam on the average, plus some stable oscillations. If these conditions are not met, the ions may perform resonant oscillations, thus becoming unstable and being lost to the chamber wall. The ion motion in fact obeys Hill's equation, similar to the beam particles in a synchrotron. By analogy the beam bunches represent thin focusing lenses for the ions. With uniform bunch repartitions and in the linear approximation of a uniform beam, the analysis (well described in reference [1]), leads to the concept of a critical mass (or rather mass-to-charge ratio) above which ions perform stable oscillations and thus can be trapped:

$$A_c > \frac{N_p r_p}{8n^2} \cdot \frac{\pi R}{b^2 (1 + \frac{a}{b})} . \quad (10)$$

Equation (10) insures vertical stability of ion motion in a uniform elliptic beam of horizontal and vertical sizes  $b$  and  $a$  respectively. For a similar horizontal stability criterium,  $b$  and  $a$  have to be interchanged in (10). Since usually  $b < a$ , the vertical stability criterium is the most stringent. This expression shows that the critical mass increases (i.e. neutralization is likely to be less severe) with the number of beam particles, with reduced beam emittances, and with a smaller number of bunches. Figure 6, relevant to the CERN EPA machine, confirms this behaviour of the critical mass with varying number of bunches. For a given beam intensity, emittances (and neutralization) are lower for a small number of equidistant bunches, which corresponds to an increased critical mass.

On the other hand, for a given number of bunches, the calculated critical mass (circled numbers) remains more or less constant as the number of beam particles increases. This is because the increase in  $N_p$  in Eq. (10) is compensated by an increase in beam sizes  $a$  and  $b$ : neutralization remains therefore more or less constant as intensity increases, meaning more ions in absolute numbers, i.e. an increasing strength of non-linear forces on individual particles. This leads to increasing emittances as particle accumulation proceeds.

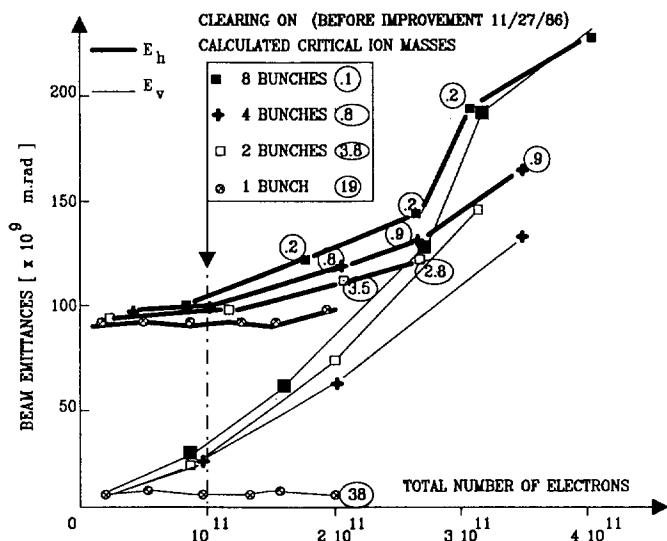


Fig. 6 Beam emittances and ion critical masses in EPA for equidistant electron bunches

As already partly explained, the ion current measured on a clearing electrode provides a good method of diagnosing neutralization (Ref. [12,13]). Ions whose masses are below the critical value are lost to the wall before they can be channelled to the next clearing electrode. Figure 7 shows that the current (and thus the neutralization) decreases with decreasing number of equidistant bunches and equal number of particles in EPA (measurement of 3.10.1987). The figure also shows that neutralization is lower with four consecutive bunches than with four equidistant ones (measurement of 8.11.1987). This illustrates the "missing-bunches scheme" as a means of clearing (see § 5.2).

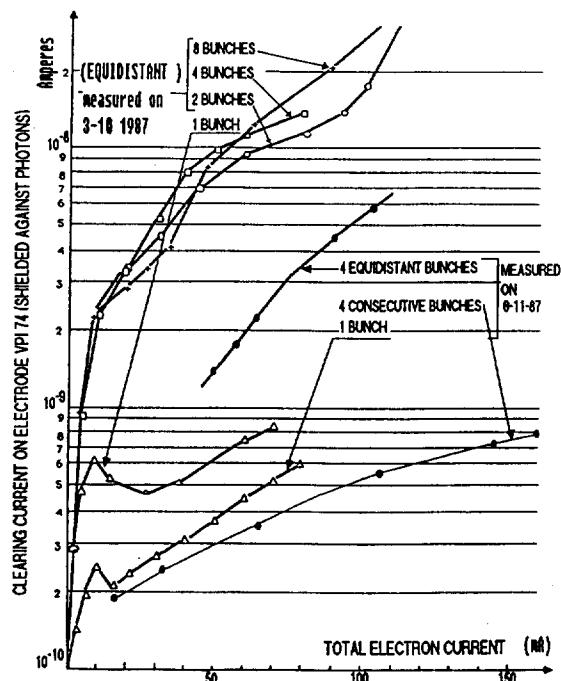


Fig. 7 Ion current measured on a clearing electrode (shielded against photons) for different bunch distributions in EPA

For completeness, it must be stressed that so far the criterium of ion stability has assumed linear and regular periodic forces. Computer simulations can easily include the non-linearity of the force, important at large amplitude, and show that a large fraction of ions which ought to be stable according to the linear theory are in fact unstable if created with large initial position and velocity ( $r_0, \dot{r}_0$ ). In addition, bunches are rarely equally populated, and this represents random gradient errors for the ion motion, similarly to synchrotrons, leading to "enlarged" ion stopbands [15]. Finally, ions represent additional thick defocusing lenses for themselves, and a detailed analysis shows that this influences the limit of accumulation (although usually only weakly).

#### 4.2 In magnets (bunched beams)

We have seen that in uniform fields (dipoles), the vertical motion of the centre of gyration is unaffected by the presence of the magnetic field  $B$ . Therefore the vertical stability criterium (10) holds.

In addition, in some rare cases and rather fortuitously, there may be a resonant condition between the horizontal cyclotron motion of the ions and the frequency of the passing bunches leading to horizontal instability [15]. In quadrupoles, the same argument holds to first order along the field lines with, in addition, an increased longitudinal drift velocity as resonant ions gain energy (see formula (7)).

## 5. CLEARING MEANS

### 5.1 Clearing electrodes

Clearing electrodes consisting of negatively polarised plates fitted into the vacuum chamber provide a transverse electric field which diverts beam-channelled ions onto them, where they are neutralized and return into the gas phase. Figure 8 qualitatively represents the potential variation across a vacuum chamber of radius  $r_c$ , with an electrode on one side with a potential of  $U_{ce}$ .

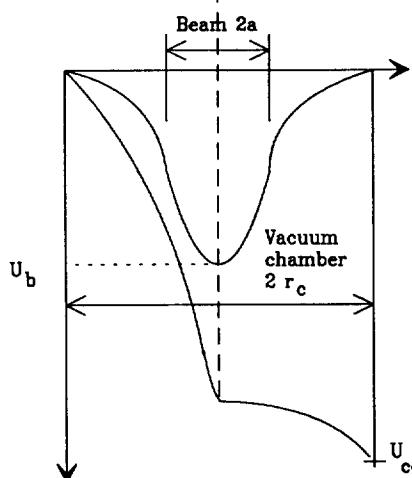


Fig. 8 Transverse potential distribution across a vacuum chamber,  
due to the beam space charge and a clearing electrode

A minimum condition for capture of the passing ion is that the transverse field provided by the electrode equals the maximum beam space charge field:

$$\frac{U_{ce}}{2r_c} > E(a) \sim \frac{I}{2\pi\beta c e_0 a} \quad (11)$$

In fact, because ions may have — as we have seen — transverse and longitudinal velocities corresponding to energies of up to a few eV (possibly up to the field at the centre if singly ionised), and because a clearing electrode is necessarily limited in size, the field provided by the electrode must usually be larger than the calculated beam field. For instance in the EPA ring, where electrodes are of the button type with a diameter of 20 mm and installed flush to the beam, transverse clearing in field-free regions is complete with the following parameters [16]:

$$U_{ce} = -6 \text{ kV}$$

electrode field: 30 kV/m on beam axis  
beam max. field: 12 kV/m for  $I = 0.3 \text{ A}$   
and  $\sim 10^{-8} \text{ m}\cdot\text{rad}$  horizontal emittance with 10% coupling.

Measurement of the electrode clearing current as a function of the applied voltage provides a verification of the required maximum to be applied for full clearing, since the current will saturate once a sufficient field is reached (Fig. 9).

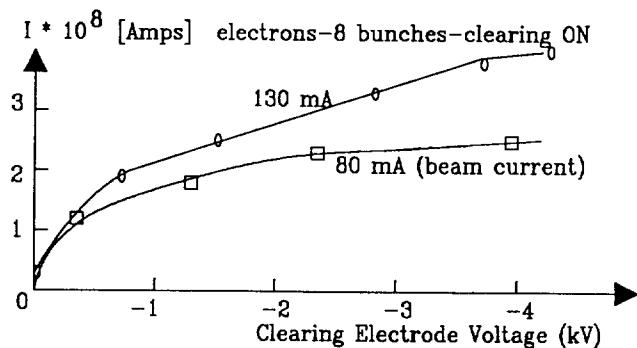


Fig. 9 Extracted ion current vs. electrode voltage in EPA

The number and the optimum location of electrodes are in principle dictated by the tolerable degree of residual neutralization. In practice, even with a large number of electrodes, uncleared pockets always remain and contribute to typical residual neutralizations of a fraction to a few percent.

No small electron storage ring exists which has reached a fully satisfactory ion-free situation, even with clearing electrodes. Perhaps one reason for this is that clearing systems have not been up to now very complete, partly owing to the fact that clearing electrodes complicate the mechanical design of the vacuum chamber and may contribute to the machine impedance. Recently at CERN (EPA), button-type clearing electrodes presenting negligible coupling characteristics to the beam have been designed. They are made of a ceramic body, coated with a highly resistive glass layer (thick-film hybrid technology), and are terminated with a highly-lossy wide-band filter [17] as shown in Fig. 10.

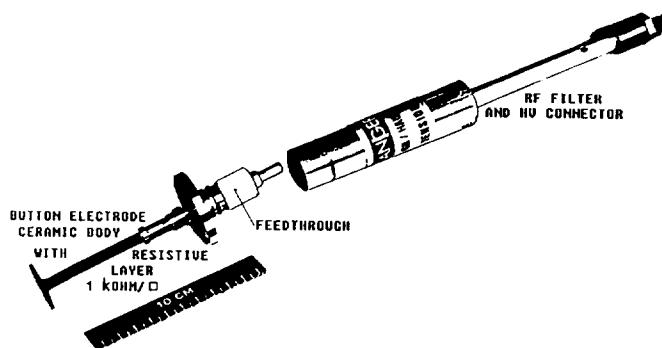


Fig. 10 EPA button type clearing electrode

Nevertheless, it still remains to be demonstrated that full beam clearing can be achieved with a clearing system. The CERN AA has reached a very low neutralization level for a DC beam machine (< 1%) with the help of an ever increased number of electrodes, but neutralization pockets still remain (Fig. 11).

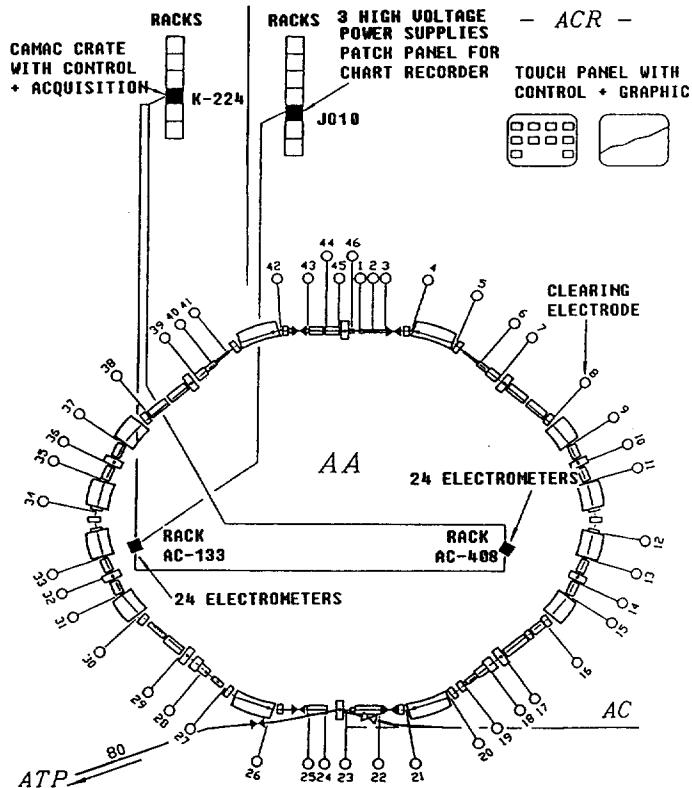


Fig. 11 The CERN AA clearing system

## 5.2 Missing bunches (electron storage rings)

Many small electron storage rings prone to ion trapping have partly solved their problem by introducing one or several gaps in the bunch train by not filling certain buckets at injection. In complement to reference 1 and using the same notations, over one revolution period of a train of  $p$  consecutive bunches, the motion (here vertical) of a trapped ion is the solution of:

$$\begin{bmatrix} y \\ \dot{y} \\ \ddot{y} \end{bmatrix}_1 = M_{Ty} \begin{bmatrix} y \\ \dot{y} \\ \ddot{y} \end{bmatrix}_0$$

with the transfer matrix

$$M_{Ty} = \begin{pmatrix} 1 & 0 \\ -a_y & 1 \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{f} \\ 0 & 1 \end{pmatrix}^p \begin{pmatrix} 1 & (h-p)\frac{1}{f} \\ 0 & 1 \end{pmatrix} \quad (12)$$

where the terms in parenthesis represent the linear kick received at the p bunch passages where [1]:

$$a_y = \frac{N}{n} \cdot \frac{r_p c}{b^2(1 + \frac{a}{b})} \cdot \frac{1}{A_i} \quad (13)$$

interleaved with drifts of duration  $1/f$ ,  $f$  being the radiofrequency. The period of  $p$  successive kicks plus drifts is terminated by the drift in the time interval  $(h-p) 1/f$ , where  $h$  is the cavity harmonic, i.e. the maximum number of bunches that the machine can handle. The Floquet's condition of stability for the ion of mass-to-charge  $A_i$ :

$$-2 < \text{Tr } (M_{Ty}) < 2$$

does not lead to a simple criterion defining which  $A_i$  are stable. Rather, the trace of the transfer matrix (12) is of order  $p$  in  $N$ , the total number of circulating particles. This means that there are  $p$  stable bands of ion mass-to-charge ratios for a given  $N$ , or that a given ion will be stable or unstable, depending on the number of beam particles.

Figure 12 illustrates the conditions of linear stability for various ions in the EPA ring, as a function of the number of beam particles and consecutive bunches [12].

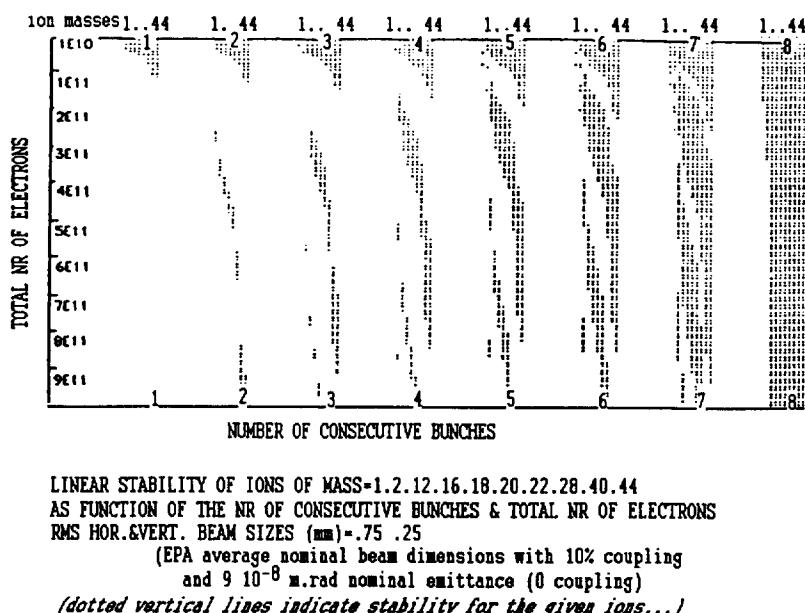


Fig. 12 Linear stability of ions in EPA as a function of total beam intensity and number of consecutive bunches for nominal beam emittances

Although not always absent, ion trapping is indeed less severe in EPA with missing bunches. This stems from the fact that since a given trapped ion can be further ionised, it then has a good chance of falling in an unstable band and thus of being cleared.

6. RESONANT TRANSVERSE SHAKING OF THE BEAM

Rather recently neutralization effects have been considerably reduced when exciting vertical coherent oscillations with a transverse kicker at a given frequency [18], both in electron storage rings (bunched) and antiproton accumulators (unbunched). This technique of "RF knock out" has been determinant in solving ion problems in the CERN AA, where it has been studied both theoretically [19] and experimentally [20] and where it is permanently implemented, with the following parameters:

shaking: vertical  
shaking frequency: 490 kHz  
sideband frequency (fractional tune  $q_v$ ): 480 kHz  
length of kicker electrodes: 0.6 m  
kicker field:  $\sim 20$  V/cm.

Although still at an early stage of both understanding and development, some beam shaking experimental observations can be summarised as follows:

- 1) Beam shaking works best when applied vertically: one possible reason is that neutralization is high in dipole fields (low ion drift velocity) where the motion along the lines of force is the only practical degree of freedom.
- 2) To work, beam shaking relies on the longitudinal motion of the ions. Due to changing beam dimensions, the ion "plasma frequency" spectrum is wide compared to the "knock out" frequency: ions have to "sweep" through this resonance. For this they must be free to move longitudinally. This is probably the reason why shaking works best in conjunction with clearing electrodes since it ensures a low level of neutralization permitting longitudinal field gradients which drive the ions around.
- 3) Beam shaking depends on the non-linearity of the beam space-charge field: this allows the "lock on" of the sweeping ions onto the resonance, where they keep large oscillation amplitudes, thus reducing their density in the beam centre.
- 4) Beam shaking is efficient even with RF kicker fields of only a few 10 V/cm, provided it is applied close to a beam betatron side band whose frequency lies close to the ion plasma frequency. In this case, the beam resonant response ensures sufficiently large non-linear forces on the ion. Experimentally it is found that for a weakly exciting RF field, shaking works best above a band ( $n+Q$ ) or below a band ( $n-Q$ ). This observation of asymmetry of weak resonant shaking is important in that it validates the non-linear character of the ion motion and the "lock on" conditions.

To illustrate this in a simple way, we use a quasi-linear description of the two-body resonant condition (due to D. Möhl, see also Ref. [19]) for an unbunched beam. We consider only one ion species  $i$ , of mass-to-charge ratio  $A_i$ , with the following definitions:

$\Omega$ : circular revolution frequency of circulating beam ( $\Omega = 2\pi f$ )

$Q_i = 2\pi f_i / \Omega$ : the ion bounce number in the beam potential well

$$\Omega^2 Q_i^2 = \frac{2N_p r_p c^2}{\pi b(a+b) R A_i} .$$

If  $Q_v$  is the beam particle unperturbed incoherent tune and  $Q_p$  the beam particle bounce number in the ion-potential well where

$$\Omega^2 Q_p^2 = \frac{2N_i r_p c^2}{\pi b(a+b) \gamma R}$$

and

$Q = (Q_v^2 + Q_p^2)^{1/2}$  is the perturbed beam tune ,

then a beam particle and an ion obey the coupled set of linear differential equations

$$\text{particle: } \left( \frac{\partial^2}{\partial t^2} + \Omega \frac{\partial^2}{\partial \theta^2} \right) y_p + Q^2 \Omega^2 y_p - Q_p^2 \Omega^2 \bar{y}_i = F e^{i\omega t} \quad (14)$$

$$\text{ion: } \left( \frac{d^2}{dt^2} \right) y_i + Q_i^2 \Omega^2 (y_i - \bar{y}_p) = 0$$

where the bar on  $y$  denotes the average vertical position of each beam and the  $F e^{i\omega t}$  term is the harmonic of the external driving force close to beam and ion resonance:

$$\omega \sim (n \pm Q) \Omega \sim Q_i \Omega .$$

Assuming solutions of the form:

$$y_p = E_p e^{i(n\theta + \omega t)}$$

$$y_i = E_i e^{i\omega t}$$

the ion amplitude becomes:

$$y_i = \frac{Q_i^2 \frac{F}{\Omega^2} e^{i\omega t}}{(x^2 - Q_i^2) ((n+x)^2 - Q^2) - Q_p^2 Q_i^2} \quad (15)$$

with

$$x = \frac{\omega}{\Omega} \rightarrow (n \pm Q) \sim Q_i .$$

Therefore, shaking works when  $y_i$  becomes large, i.e. when the denominator  $\rightarrow 0$ . But, as shown in Fig. 13, as the ions gain large amplitude, we have non-linear detuning such that  $x^2 > \Omega_i^2$  (lock on). Therefore for  $y_i$  to become large, requires that:

$$(n + x)^2 - \Omega^2 > 0$$

i.e.:  $x < n - \Omega$  for excitation near a "slow-wave" beam frequency:  $\omega \leq (n-\Omega)\Omega$ ;  $n > \Omega$   
 $x > n + \Omega$  for a "fast-wave" frequency :  $\omega \leq (n+\Omega)\Omega$ ;  $n > \Omega$ .

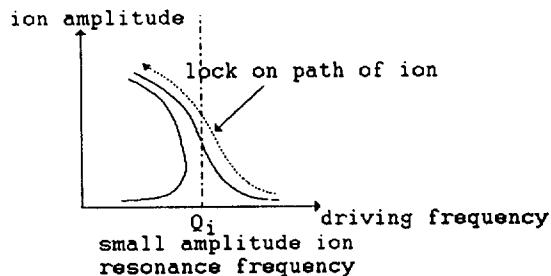


Fig. 13 Qualitative amplitude response curve of an ion versus the driving frequency, near resonance

This asymmetry has been verified both in the CERN and Fermilab antiproton accumulators and in the EPA ring. Figure 14 for the Fermilab AA, shows that shaking below a "slow wave" excited by ions is more efficient than shaking above, since it suppresses the instability signal.

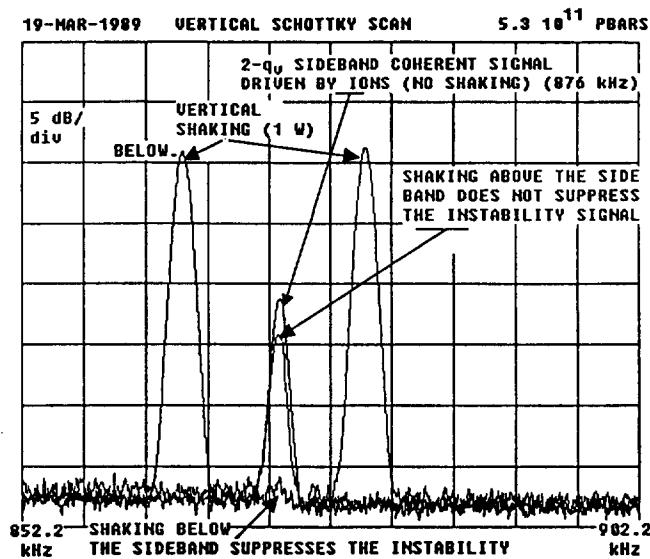


Fig. 14 Vertical Schottky scan showing the suppression of an ion-driven dipolar instability by shaking in the FNAL Antiproton Accumulator

The analysis developed above simply hints at the reality. In fact, when one considers ions "locked" on to resonance, the amplitude of their coherent oscillation  $y_i$  depends only little on the amplitude of the excitation, and the oscillating external force  $F$  in (14), when it is applied below or above a sideband [19]:

$$F \rightarrow F_{\text{eff}} = F \pm Q_i^2 \bar{y}_i, \quad (x + n)^2 \gtrless Q^2.$$

To conclude on beam shaking as a means to suppress ion effects, it must be stressed that this technique is still at an early stage of development although already applied permanently on Antiproton Accumulators and electron storage rings [20]. On the CERN EPA, shaking has made it possible to overcome a neutralization threshold (Fig. 15) which no other method (clearing electrodes, transverse kicks...) could achieve. From tune-shift measurements, clearing electrodes reduce the neutralization coefficient down to typically 2-3%. Shaking the beam reduces it further to well below 1%.

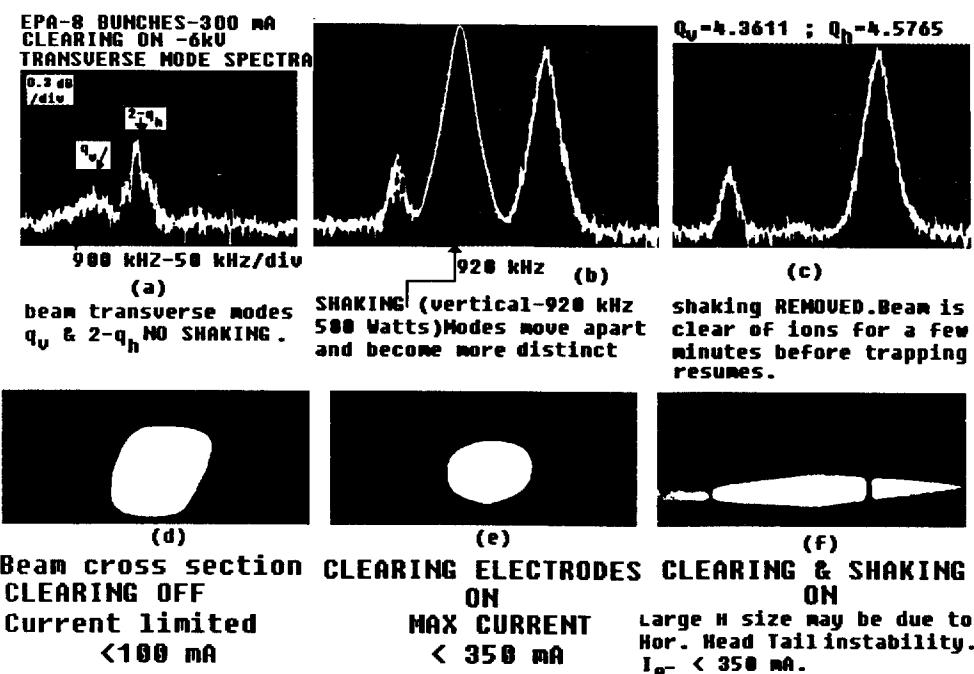


Fig. 15 Transverse Schottky scans and beam cross-sections in EPA showing the effect of shaking the beam (920 kHz vertically)

## 7. CONCLUSION

Often very detrimental in their effects, ions may represent a real challenge to machine designers. We have seen that they can be partly eliminated with clearing electrodes, but this requires a careful design if one wants to avoid increasing the machine impedance. Empty buckets on bunched machines and, more recently, beam shaking are other means which are used to further reduce their numbers. The latter is still at the experimental stage but represents a real hope for some future machines where the very small beam sizes and high intensities may render very difficult the design of clearing electrodes.

ACKNOWLEDGEMENTS

S. Turner has been very helpful in shaping up the manuscript. K. Hübner and D. Möhl made many valuable comments and corrections. My many thanks to them, and to all the people listed in the references.

\* \* \*

REFERENCES

- 1) Y. Baconnier, Neutralization of Accelerator Beams by Ionization of the Residual Gas, CERN/PS/PSR/84-24, CERN Accelerator School, Gif-suf-Yvette, France, 1984, CERN 85-19, pp. 267-288.
- 2) S. Battisti, M. Bell, J.-P. Delahaye, A. Krusche, H. Kugler, J.H.B. Madsen, A. Poncet, 12th. Int. Conf. on High En. Acc., Fermilab, p. 2050, 1983.
- 3) J.D. Jackson, Classical Electrodynamics, John Wiley & Sons, Inc., New York, 1962.
- 4) The PS Staff, The CERN PS complex as an Antiproton Source, IEEE Trans. Nucl. Sci., Vol. NS-32, p. 2218, 1985.
- 5) F.F. Rieke, W. Prepejchal, Ionisation Cross Sections of Gaseous Atoms and Molecules for High-Energy Electrons and Positrons, Phys. Rev. A, Vol. 6, No.2 (1972).
- 6) F. Pedersen, A. Poncet, Proton-Antiproton Instability in the AA, CERN PS/AA/ME Note 81, 1981.
- 7) S.C. Brown, Basic Data of Plasma Physics, Technology Press, J. Wiley & Sons, Inc., New York, 1959.
- 8) B. Rinnert, Unpublished Note, ISR/VA/tn, 1977.
- 9) M. Gygi-Hanney, B. Zotter, Field Strength in a Bi-Gaussian Beam, CERN/LEP/Theory Note 44, 1987.
- 10) R. Alves-Pires, Conformal Mapping for Two-dimensional Electrostatic Beam Potential Calculations, CERN PS/87-66 (AA), 1987.
- 11) R. Alves-Pires, Beam Dimensions and Beam Potential in the CERN Antiproton Accumulator Complex, CERN PS/87-70 (AA), 1987.
- 12) A. Poncet, Trapping of Ions in the EPA Electron Beam: Stability Conditions and Diagnosis, CERN/PS 88-14 (ML), 1988.
- 13) F. Pedersen, A. Poncet, L. Søby, The CERN Antiproton Accumulator Clearing System, CERN PS/89-17 (ML), 1989. 1989 Part. Acc. Conf., Chicago, 1989.
- 14) Y. Miyamara, K. Takayama, G. Horikoshi, Dynamical Analysis on the Longitudinal Motion of Trapped ions in Electron Storage Rings, Nucl. Inst. and Meth. in Ph. Res., A 270, p. 217, 1988.
- 15) G. Brianti, Y. Baconnier, O. Gröbner, E. Jones, D. Potaux, H. Schönauer, Proceedings of the Workshop on pp in the SPS, SPS pp.1, p. 121, 1980.
- 16) J.-C Godot, K. Hübner, A. Poncet, Comparison of the Electric Field in the Beam to the Available Clearing Fields in EPA, Note PS/LPI/87-25, 1987.
- 17) F. Caspers, J.P. Delahaye, J.C. Godot, K. Hübner, A. Poncet, EPA Beam-Vacuum Interaction and Ion Clearing System, CERN PS/88-37 (ML), 1988, EPAC, Rome, 1988.

- 18) Y. Orlov, The Suppression of Transverse Instabilities in the CERN AA by Shaking the  $\bar{p}$  Beam, CERN PS/89-01 (AR), 1989, 1989 Part. Acc. Conf., Chicago, 1989.
- 19) R. Alves-Pires et al. On the Theory of Coherent Instabilities due to Coupling between a Dense Cooled Beam and Charged Particles from the Residual Gas, CERN/PS/89-14 (AR), 1989, 1989 Part. Acc. Conf. Chicago, 1989.
- 20) J. Marriner, D. Möhl, Y. Orlov, S. van der Meer, A. Poncet, Experiments and Practice in Beam Shaking, CERN/PS/89-48 (AR), 1989, 14th Int. Conf. on High Energy Accelerators, Tsukuba, Japan, 1989.
- 21) T. Kasuga, Ion Clearing System of UVSOR Storage Ring, Japanese Journal of Applied Physics, Vol. 25, No. 11, p. 1711, 1986.

## SCHOTTKY NOISE FROM VERY COLD BEAMS

*S. Baird*

CERN, Geneva, Switzerland

### ABSTRACT

The aim of this paper is to introduce the idea of the modification of the statistical noise signals, observed on coasting beams, due to interactions between the particles. These interactions will be most noticeable as the particle density is increased by the action of a strong beam cooling system. Some examples from existing machines, using electron cooling devices will also be given. The discussions will be limited to longitudinal noise signals.

### 1. INTRODUCTION

This subject has been treated in detail in several previous papers [1-3], and this article is only intended as a general introduction to the topic. For detailed derivations the reader should see the above references.

The noise signals observed on low density, uncooled beams, due to their statistical density fluctuations, have been well described elsewhere [4]. Such signals are called Schottky noise by analogy with the noise signals generated in an electron tube, from the random emission of individual electrons at the cathode [5]. Likewise a circulating beam is made up of a large, but finite number of particles, and for a low-phase-space density beam the motion of any one particle can be considered as independant of all the others in the beam. However as the beam phase space density is increased by some external cooling system, interactions between the particles must be taken into account in the interpretation of the observed noise signals, as they will reduce the degree of randomness of the particle motion. Such correlations between particles will, at high beam densities, strongly modify the statistical fluctuations within the beam itself and, therefore, any observed signals.

### 2. HOW DO WE DEFINE "HOT" AND "COLD" BEAMS?

Before it is possible to discuss statistical noise signals from very cold beams, it is necessary to define what constitutes a cold beam, and at the same time a hot beam. In discussing the beam state in terms of temperature it is useful to use some definitions from the statistical physics treatment of a classical gas [6].

An Ideal Gas: A gas of molecules whose mutual interaction is almost negligible. (The molecules have to be allowed to exchange energy slowly to bring the gas into thermal equilibrium).

Equilibrium: A macroscopic state, which does not change in time except for random fluctuations.

Longitudinal temperature:  $KT = p^2/m$  where  $K$  = Boltzmann's constant,  $T$  = temperature,  $p$  = molecule momentum,  $m$  = molecule mass.

For longitudinal motion in the beam frame of reference then this formula becomes:

$$KT = \langle dp^2 \rangle / m$$

The beam frame of reference is defined as the frame moving along in the beam direction at the average beam velocity.

Mean Free Path: The average distance a molecule travels in the gas before it collides (interacts) with another molecule.

In the beam frame of reference, a hot beam can be defined as a beam which behaves like an ideal gas, where the interactions between the particles are effectively negligible, and the mean free path between particle interactions is long. In a cold beam, the mean free path between collisions is short, and therefore the particle interactions cannot be ignored. Here the words "collisions" and "interactions" are used to describe the interaction of a particle with its neighbours, via its environment, or via direct Coulomb fields. This is normally characterised by a longitudinal impedance.

In accelerator physics terms the importance of the beam's response to its own internally induced collective interactions can be characterised by a longitudinal stability criterion, the Keil-Schnell criterion [7]. This states that:

$$\frac{|Z_n|}{n} < \frac{m_e c^2 \beta^2 \gamma |\eta|}{e^2 N w} \times \left[ \frac{dp}{P} \right]^2 \quad (1)$$

For a given longitudinal impedance of the machine ( $Z_n/n$ ), this inequality defines the limiting values of particle number ( $N$ ) and momentum spread ( $dp/p$ ) which result in stable longitudinal motion. It should be noted that this relation is only an approximation to the stable area in the stability diagram [7]. The consequences of this approximation will be re-examined in section 5.

In this way a hot beam can be described as one in which collective effects can be ignored, which means that the beam is unable to organise itself into a coherent response

to the collective interactions. In a cold beam the collective effects inside the beam can no longer be ignored, since the beam is now able to develop a coherent response to its own collective interactions. The Keil-Schnell criterion defines the range of values of  $N$  and  $dp/p$  for which one could expect the beam behaviour to pass from one regime to the other.

As an example one can estimate the minimum stable momentum spreads as a function of intensity for a proton or an antiproton beam in LEAR [8] at 105 MeV/c (5 MeV):

For  $N = 1 \times 10^9$  particles  $(dp/p)_{\min} = 1.8 \times 10^{-4}$  (FWHM)

For  $N = 1 \times 10^8$  particles  $(dp/p)_{\min} = 5.7 \times 10^{-5}$  (FWHM)

It should also be noted that at low momenta in small machines the longitudinal impedance is dominated by the space charge term given by :

$$\frac{Z_i}{n} = \frac{Z_0 g}{2\beta\gamma^2}$$

Where  $Z_0 = 377$  ohms and  $g \approx 2.4$  [9].

By rearranging (1) one can also estimate, for a given impedance and momentum spread, the maximum value of  $N$  for which the beam will remain longitudinally stable. This value  $N_{cr}$  is

$$N_{cr} = \frac{m_e c^2 \beta^2 \gamma d\omega^2}{e^2 N w^3 \eta} \times \left[ \frac{Z_n}{n} \right]^{-1} \quad (2)$$

It is now immediately apparent that typical beams in machines using stochastic cooling can be considered as hot. In LEAR the minimum momentum spreads obtained at 105 MeV/c using stochastic cooling are around  $dp/p = 1$  to  $2 \times 10^{-3}$  FWHM. However, using electron cooling it is possible to approach the boundary between the hot and the cold beam regimes, as the momentum spreads theoretically obtainable with electron cooling are in the  $10^{-4}$  to  $10^{-5}$  range.

### 3. LONGITUDINAL DENSITY FLUCTUATIONS (A SIMPLE PICTURE)

Before examining the effects of the particle interactions on the observed density fluctuations in a cold beam, as seen by a longitudinal pick-up, it is useful to review the case of a hot beam. In this section, a simplified picture will be presented, and in sections 4 and 5 it will be explained in more detail.

### 3.1 Density Fluctuations in a "hot" beam

Consider, in the beam frame of reference, the longitudinal particle motion as that of an ideal gas contained in a large box with volume  $V_b$ . Now the pick-up is a small volume  $V_p$  inside this box. Let  $N$  = the total number of particles,  $p = V_p/V_b$  and  $q = 1 - p$ .

Then the average number of particles in the pick-up is  $NP$ . However, due to the random motion of the particles, this number varies in time. Over many observations of the number of particles in the pick-up, then the probability,  $P(n)$ , of finding a certain number  $n$  inside the pick-up is given by:

$$P(n) = \frac{1}{\sqrt{2\pi Npq}} \times \exp \frac{-(n-Np)^2}{2Npq}$$

If  $A_n$  and  $\sigma_n$  are defined as,  $A_n = (n - Np)$  the amplitude of the instantaneous observed density fluctuation, and  $\sigma_n = \sqrt{(Npq)}$ , then the probability of observing a density fluctuation of amplitude  $A_n$  can be written as:

$$P(n) = \frac{1}{\sqrt{2\pi \sigma_n^2}} \times \exp \frac{-A_n^2}{2 \sigma_n^2} \quad (3)$$

These density fluctuations, due to the randomness of the particle motion give rise to the observed Schottky noise. However this has now to be applied to an accelerator or storage ring where the same particles circulate many times through the pick-up. In a storage ring containing a continuous beam, and considering only longitudinal motion, then a current will be detected by the pick-up, due to the changing charge density inside it. This current will be given by the rate of change of charge ie.

$$I = e(dn/dt)$$

If  $\phi$  is the angular position around the ring then this can be rewritten as:

$$I = e(d\phi/dt) \cdot (dn/d\phi) = e\omega(dn/d\phi)$$

where  $\omega$  is the particle revolution frequency.

Now provided the particle distribution is completely random, i.e. there is no relation between  $n$  and  $\omega$ , and  $\omega$  is constant for each particle, then the contribution from each particle is a periodic  $\delta$ -function with a period of  $2\pi/\omega$ . Thus  $(dn/d\phi)$  can be expressed as a sum over all integer multiples of  $\omega$ , by expansion of the  $\delta$ -function, summed over all  $N$  particles. In this case the above equation can be rewritten as:

$$I = e \sum_1^N \sum_{k=-\infty}^{k=\infty} \exp(i(k\theta+\phi))$$

where  $\phi$  is the initial position, or phase, of each particle around the ring. Therefore,

$$I = Ne\omega_0 + 2e \sum_{k=1}^N \sum_{k=1}^{k=\infty} \omega \cos(k\omega_0 t + \phi)$$

In reality each particle will have a slightly different revolution frequency,  $\omega$ , and will thus contribute its own series of harmonics. Therefore the observed current will contain a DC part,  $k = 0$ , and a series of "bands",  $k \neq 0$ , around successive harmonics of the average revolution frequency  $\omega_0$ . These bands are known as Longitudinal Schottky bands. However, the particles are distributed randomly all around the accelerator ring. Therefore, after the summation over all  $N$  particles, only the DC term,  $Ne\omega_0$ , remains, due to the random initial phase factor,  $\phi$ , for each particle.  $Ne\omega_0$  is simply the DC beam current. However the rms current per band does not vanish in the summation process due to the  $\cos^2$  term. So, for one band, one value of  $k$ , the rms current is given by:

$$I_{rms} = \sqrt{I^2} \text{ where } I^2 = 4e^2 \omega_0^2 \sum_{k=1}^N \cos^2(k\omega_0 t + \phi)$$

Therefore,

$$I_{rms} = 2e\omega_0 \sqrt{N/2}$$

This is the Schottky noise current, which can be observed by tuning a spectrum analyser to the frequency of the band ( $k\omega_0$ ) under consideration, and, for a hot beam, in which the motion of each particle is independent of the other particles, the observed Schottky power (power  $\propto I_{rms}^2$ ) is proportional to the circulating beam intensity,  $N$  [4].

The expression for the total current per band is independant of the harmonic number,  $k$ , of the band, but the current distribution inside one such band provides valuable information on the momentum profile of the beam since,

$$dp/p = d\omega/\omega_0 (1/k) (1/\eta)$$

The measured power is proportional to the total beam current,  $N$ , and the average revolution frequency,  $\omega_0$ , squared. For a given momentum spread, as the observation frequency is increased,  $k$ , and the width of the bands increases, until at high frequencies they overlap, when  $k\Delta\omega/\omega_0 > \omega_0$ . Since the rms current per band is constant, the current density (i.e. the height) of the bands decreases with  $1/k$ , until overlap occurs.

Combining the expression for  $I_{rms}$  with  $\sigma_n$ , the width of the probability distribution (3)  $P(n)$  where  $\sigma_n = \sqrt{Npq}$  gives:

$$I_{rms} \propto 2e\omega_0 \sigma_n \quad (3a)$$

### 3.2 Density Fluctuations in a "cold" beam

In the previous section the mean free path of each particle in between collisions was considered as long compared with the dimensions of the pick-up "box". However for a cold beam, in which  $d\rho/p$  ( $d\rho/\omega$ ) has been reduced so that  $N \gg N_{cr}$ , the mean free path between interactions will be short when compared with the pick-up "box" volume. Therefore the density fluctuations, in the pick-up, will be dominated by these collisions. In a gas this is equivalent to saying that the particle motion now resembles Brownian motion, and that the density fluctuations look like thermal noise. In other words the observed noise current is proportional to the square root of the longitudinal beam temperature, or the noise power is proportional to  $(d\rho)^2$ .

Such a result can be derived in detail [1,2] by including particle interactions in (3). If the beam is in or very close to longitudinal thermal equilibrium, then the interactions between the particles can be included by multiplying the unperturbed probability distribution of (3) by an equilibrium statistical Boltzmann distribution of the form:

$$\exp \frac{-E}{KT} \quad (4)$$

where  $E = kV(i,j)$  and  $T$  = longitudinal beam temperature.

$V(i,j)$  is the longitudinal interaction potential between the  $i$ th and the  $j$ th particles, via the longitudinal impedance  $Z_n$ . This is then summed over all pairs of particles to obtain  $E$ . Including this term (4) in (3), one obtains a modified probability distribution of the form:

$$P(n) = C \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp \frac{-A_n^2}{2\sigma_n^2} \left(1 - \frac{N}{N_{cr}}\right) \quad (5)$$

where  $N_{cr}$  is identical to the  $N_{cr}$  from (2) as derived from the Keil-Schnell criterion. Apart from a normalisation constant,  $C$ , Eq.(5) is very similar to (3) except that:

$$\sigma_n^2 \text{ is replaced by } \sigma'_n{}^2 = [\sigma_n^2 \times N_{cr}/(N_{cr} - N)] \quad (6)$$

In this case the equation for the Schottky noise current per band (3a), which was shown at the end of the previous section on "hot" beam density fluctuations, becomes :

$$I_{rms} \propto 2e\omega_0 [\sigma'_n{}^2 \times N_{cr}/(N_{cr} - N)] \quad (6a)$$

It is useful to examine (6) and (6a) in a little more detail. The expression (6) will always be finite and positive if  $N_{cr}$  is negative. However, if  $N_{cr}$  is positive then  $\sigma'_n{}^2$  will only remain finite and positive for values of  $N < N_{cr}$ . From Eq. (2), the sign of

$N_{cr}$  is determined by the sign of  $\eta$ , so these two situations correspond to being either below or above transition energy, provided that the reactive (imaginary) contribution to the impedance ( $Z_n$ ) is large and positive and the resistive (real) part is small [7]. This is the case for a small machine where the longitudinal impedance is dominated by the space charge term i.e. a capacitive impedance. Under these conditions, above transition energy, as  $N \rightarrow N_{cr}$ , then  $\sigma'_n{}^2 \rightarrow \infty$ , therefore the beam exhibits unstable fluctuations, and the justification for the beam being in thermal equilibrium breaks down. In practice this would be observed as a longitudinal beam instability, the negative mass instability. Below transition, then  $\sigma'_n{}^2$  is always finite and positive, therefore the density fluctuations will be stable and not lead to any global beam instability. The effects of longitudinal impedance on beam stability above and below transition have been well reviewed in several papers, see for example [7].

This leads naturally to the conclusion [8] that small rings, operating below transition energy, should be the most suitable places to create and store very high phase space density beams.

If  $N \ll N_{cr}$ , then, under all circumstances,  $\sigma_n{}^2 \propto N$ , and the observed noise signal will be normal random Schottky noise, where  $I_{rms}{}^2 \propto N$ . As  $N \rightarrow N_{cr}$ ,  $\sigma'_n{}^2 \rightarrow \infty$ , for a machine operating above transition, however, in a machine operating below transition it appears to be possible to allow the case that  $N > |N_{cr}|$ , in which case  $\sigma_n{}^2 \propto N_{cr}$  (remember  $\sigma_n{}^2 \propto N$ ). From (2) it is apparent that  $N_{cr} \propto (dw)^2$ , and therefore  $\sigma_n{}^2 \propto (dw)^2$ . The observed noise signal now behaves like thermal noise and signal power is proportional to the longitudinal beam temperature i.e.  $I_{rms}{}^2 \propto (dw)^2$ .

Such phenomena were first observed in electron cooling experiments at Novosibirsk [10] and more recently in several other electron cooling experiments [11-14]. Some examples of observed spectra will be given in section 5.

One must however consider the justification for allowing  $N \gg N_{cr}$ , as at first sight this appears to blatantly contradict the longitudinal stability criterion (1). This very important question is addressed in the following section.

#### 4. LONGITUDINAL STABILITY AND COHERENT MOTION IN A COLD BEAM

Initially this discussions will be limited to the simple case of a very cold beam with no external cooling force. This beam will be considered as having no frequency spread and initially, a uniform longitudinal line density. If a single particle is displaced longitudinally with respect to its neighbours then there will be a resulting force on these, which will spread this displacement around the beam in the form of a density fluctuation. To determine if the beam is longitudinally stable it is necessary to see whether this fluctuation damps itself or grows in amplitude. The treatment will follow closely that given in [7].

Consider a uniform beam with no frequency or energy spread, to which a small longitudinal density modulation is introduced of the form  $\exp[i(n\theta - \Omega t)]$  where  $n$  is the number of density oscillations per turn,  $\theta$  is the angular position around the ring and  $\Omega$  is the angular velocity of the density modulation. This introduces a fluctuation in the beam current,

$$I = I_0 + I_1 \exp(i(n\theta - \Omega t)) \quad (7)$$

which in turn, due to the presence of a non-negligible impedance  $Z_n$ , which has both real and imaginary components ( $Z_r + iZ_i$ ), induces a voltage in the surrounding vacuum chamber given by:

$$W = -(Z_r + iZ_i)I_1 \exp(i(n\theta - \Omega t)) \quad (8)$$

This voltage acts back upon the beam and causes a fluctuation in the particles' angular velocity  $\omega$ , which is given by:

$$\omega = \omega_0 + \omega_1 \exp(i(n\theta - \Omega t)) \quad (9)$$

It now remains to see whether the effect of this voltage is to damp (stable motion) or amplify (unstable motion) the initial density modulation, and under what conditions the motion is stable or unstable. This voltage will change the particles' energy so that the rate of change of energy per turn is given by:

$$dE/dt = eW\omega/2\pi \quad (10)$$

However the change in energy will modify the angular velocity  $\omega$ :

$$d\omega/dt = (d\omega/d\theta)(d\theta/dt) + (d\omega/dE)(dE/dt) \quad (11)$$

But  $d\omega/dE = (\eta\omega_0)/(\beta^2 E)$  so that (11) becomes

$$d\omega/dt = (d\omega/d\theta)\omega_0 + [(\eta\omega_0)/(\beta^2 E)]dE/dt \quad (12)$$

Therefore combining (8), (9) and (12) gives:

$$\omega_1 (\Omega - n\omega_0) = \frac{-(i\eta\omega_0^2 [Z_r + iZ_i] I_1)}{(2\pi\beta^2 E)} \quad (13)$$

However, for a one-dimensional beam the change in local density per unit time for a short section of the beam ( $d\theta$ ) is just the rate of change of current ( $dI/dt$ ). Now the continuity equation states that if there is a change of density then there must be an equal flux in (or out) of the section, defined by  $d\theta$ . Hence,

$$dI/dt = -\text{Div}(I\omega) = -d(I\omega)/d\theta \quad (14)$$

Provided the density perturbation is small so that

$$I \rightarrow I_0, \text{ and } \omega \rightarrow \omega_0,$$

then (14) becomes:

$$\frac{dI}{dt} = -\omega_0 \frac{dI}{d\theta} - I_0 \frac{d\omega}{d\theta} \quad (15)$$

Substituting into (15) from (7) and (9):

$$I_1 \Omega = n I_1 \omega_0 + n \omega_1 I_0 \quad (16)$$

Combining (13) and (16) gives:

$$(\Omega - n\omega_0)^2 = \frac{-(ie\eta\omega_0)^2 n [z_r + iz_i] I_0}{(2\pi\beta^2 E)} \quad (17)$$

Therefore coherent particle motion, which gives rise to global density fluctuations is only possible for specific values of  $(\Omega - n\omega_0) = d\Omega$ . It is important to remember that  $d\Omega$ , is a complex variable ( $d\Omega_r + id\Omega_i$ ) and therefore the final equation for the density modulation is:

$$I = I_0 + I_1 \exp(d\Omega i t) \exp(i(n\theta - [n\omega_0 + d\Omega_r t])) \quad (18)$$

Where,

$$(d\Omega_r + id\Omega_i)^2 = \frac{-(ie\eta\omega_0)^2 n [z_r + iz_i] I_0}{(2\pi\beta^2 E)} \quad (19)$$

If  $z_i$  is zero, i.e. the machine impedance is purely resistive then from (19)  $d\Omega_i$  is always non-zero in which case there will always be an exponentially growing term  $\exp(d\Omega_i t)$  in (18) and there will always be an instability. However when the longitudinal impedance is dominated by the imaginary part i.e. a reactive impedance, and in a small low energy machine this will be mainly the space charge contribution, this leads to a positive value for  $z_i$ . In this case, below transition energy,  $d\Omega$  is purely real and  $d\Omega_i$  is zero, so there is no term  $\exp(d\Omega_i t)$  in (18), there is only the oscillatory term, where  $d\Omega_r$  is the angular velocity of the perturbation around the beam. Above transition energy  $d\Omega$  becomes imaginary and  $d\Omega_i$  is non-zero, now the term  $\exp(d\Omega_i t)$  leads to the well known "negative mass instability", which was mentioned in section 3. Figure 1 shows such an instability observed during electron cooling studies at ICE, as self-bunching of the coasting proton when operating the machine above transition [15].

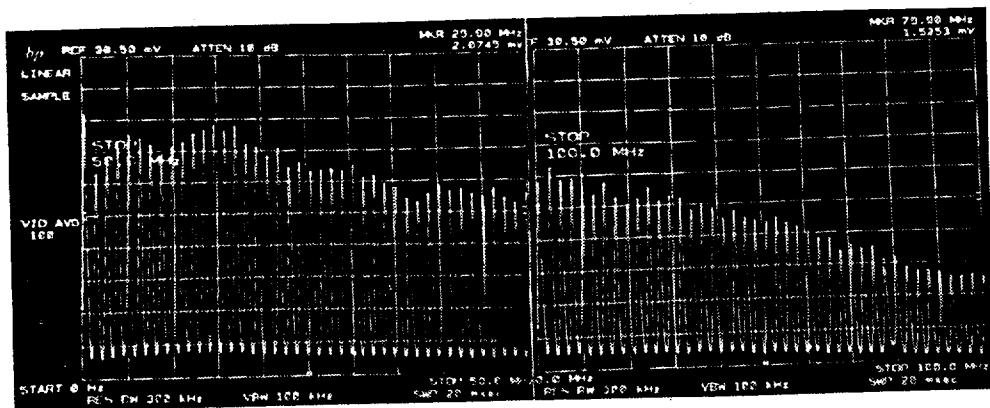


Fig. 1 Longitudinal frequency spectrum of a self-bunched beam observed in ICE, when operating above transition.

In conclusion it can be seen that, for a machine operating below transition with a space charge dominated longitudinal impedance, the collective interactions between the particles will give rise to density waves around the beam whose angular velocity is given by:

$$d\Omega_r^2 = \frac{(en\omega_0^2 n[z_i] I_a)}{(2\pi\beta^2 E)}$$

Until now  $\Omega_r$  and  $\omega_0$  have been angular frequencies, however dividing both sides of this equation by  $(2\pi)^2$  gives the frequency of the density waves. Now  $I_a$  is the average current and can be written as  $N\omega_0$ , where  $N$  is the total number of particles. This gives the frequency of the density waves as:

$$d\Omega_r^2 = \frac{(e^2 n\omega_0^3 n[z_i] N)}{(2\pi\beta^2 E)} \quad (20)$$

This is very similar to the result found in [1-2]. It should be noted that there are two solutions to (20), one for  $d\Omega_r$  positive and the other for  $d\Omega_r$  negative. These correspond to two waves one moving with the beam direction and the other against.

From (18) and (19) it can be seen that all beams will exhibit some form of coherent response, either stable or unstable, the number of particles  $N$  determines the amplitude of the fluctuation. This is because the effect of a finite frequency or energy spread has

been ignored. The effect of this frequency spread is to damp out the coherence for small values of  $N$  and/or large values of  $d\omega$ . This is very fortunate as it means that beams with finite frequency spread will be stable under certain conditions. If the frequency spread  $d\omega$  that is introduced is greater than the  $d\omega_r$ , the frequency of the coherent density wave, then the density wave will not be able to remain coherent and will be damped (Landau damping) [16]. Conversely if  $d\omega$  is less than  $d\omega_r$ , then the coherent motion will remain. This modulation of  $\omega_0$  gives rise to two coherent peaks in the longitudinal noise spectrum at frequencies  $(\omega_0 \pm d\omega_r)$ . The splitting of these peaks will be given by  $2d\omega_r$ , where:

$$|d\omega_r| = \frac{\sqrt{(e^2 \eta \omega_0^3 n [z_i] N)}}{\sqrt{2\pi \beta^2 E}} \quad (21)$$

These conditions are described in a "longitudinal stability diagram" [7] which is a scaled plot of imaginary against real impedance where the scaling is given by:

$$V + iU = \frac{2Ne^2 \omega_0}{m_e c^2 \beta^2 \gamma \eta (dp/p)^2 n} (z_r + iz_i) \quad (22)$$

Such a plot is shown in Fig. 2. Any pair of  $z_r$ ,  $z_i$  which give a point inside the closed curve (curve 1 in Fig. 2) will lead to a stable beam, for the given values of  $N$ ,  $dp/p$  etc. The exact shape of the curve depends on the longitudinal beam distribution, and the size of the stable area is determined by the current  $N$  and the momentum spread  $(dp/p)^2$ . If  $N$  is increased or  $dp/p$  is decreased then the stable area shrinks. Normally  $z_n$  is fixed for a particular set of machine conditions so the beam stability is determined solely by the beam intensity and momentum spread.

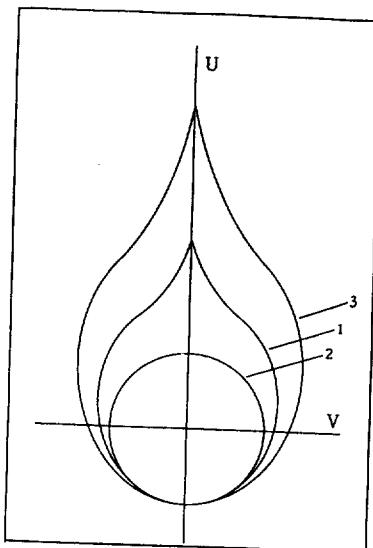


Fig. 2 Longitudinal stability diagram. Curve 1 defines the theoretical limit for longitudinal stability. Curve 2 shows the curve for  $N = N_{cr}$ , defined by the Keil-Schnell longitudinal stability criterion. Curve 3 shows the possible modification of the stable region due to electron cooling.

The Keil-Schnell stability criteria is estimated by approximating the curve 1 as a circle (curve 2 in Fig. 2). Therefore a point on the edge of this circle corresponds to the condition  $N = N_{cr}$ , as given in (2). The case of a machine in which the longitudinal impedance is principally due to the space charge term,  $Z_s$ , will be small and  $Z_i$  will be large and positive. On the stability diagram this corresponds to moving along the positive  $U$  axis. This suggests that there may be an area of stability for  $N > N_{cr}$ , outside the Keil-Schnell circle but still inside the stable area.

It has also been shown [3] that a strong electron cooling can damp the collective motion if the cooling time is less than half of the coherent fluctuation growth time. The effect of this would be to open up the stable area of the stability diagram along the  $U$  axis (curve 3 in Fig. 2). This would mean that  $N$  could be increased still further and the particle motion would remain stable, the collective density waves described in section 4 would be present, and the normal Schottky noise signals would become thermal noise signals. The action of the cooling system would maintain the beam in this quasi-stable state. Immediately the cooling was switched off however, then the beam would find itself in a now unstable state and  $dp/p$  would increase rapidly until the beam again found itself inside the reduced stable area of the stability diagram. Under these conditions, without strong longitudinal cooling, it is no longer possible for the beam to exceed the standard stability limit, defined by  $N > N_{cr}$ , therefore the beam will no longer be able to organise a coherent response to the collective interactions, the density waves will disappear and the observed noise signals will revert to the random noise Schottky signals.

##### 5. COMPARISON WITH MEASURED SPECTRA

For a machine with a space-charge dominated impedance, operating below its transition energy, one would expect to observe the following behaviour in the longitudinal noise spectrum. If the number of particles  $N$  is increased in the presence of electron cooling, the cooling must be strong enough to maintain a small momentum spread  $dp/p \ll 10^{-4}$ . Then, from (2) and (20) one should observe three distinct regimes, which correspond to three ranges of values of  $N$ :

For  $N < N_{cr}$ , the longitudinal noise spectrum will be the typical random noise Schottky spectrum, whose spectrum distribution will mirror the longitudinal density distribution. This corresponds to the case that  $d\Omega_r < dw_0$ , where  $dw_0$  is the natural unperturbed beam frequency spread [4].

For  $N = N_{cr}$ , the longitudinal spectrum will begin to distort away from the normal Gaussian type distribution, as the noise spectrum moves into the thermal noise domain and the two coherent density waves will begin to develop, causing the two peaks at  $\pm d\Omega_r$  to begin to appear. This corresponds to the case that  $d\Omega_r = dw_0$ .

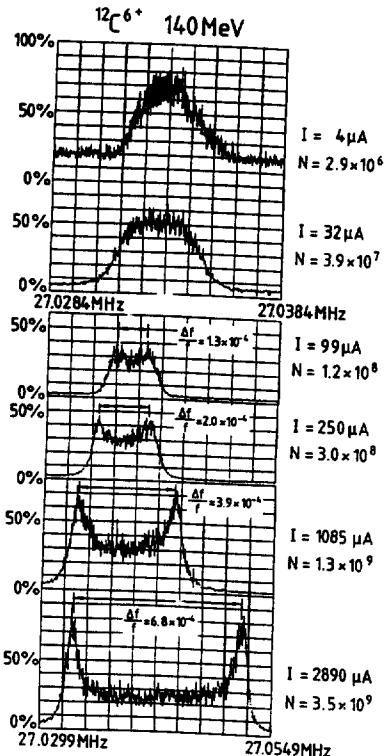


Fig. 3 Longitudinal noise spectra observed in TSR  $^{12}\text{C}^{6+}$  ions with electron cooling at various beam intensities.

For  $N > N_{cr}$  then the longitudinal spectrum will be dominated by the two peaks at  $\pm \Delta\Omega_r$ , around each revolution harmonic, the random Schottky spectrum will have almost completely disappeared. The splitting of the two peaks will be given by (21) and will increase as  $\sqrt{N}$ . This corresponds to the case that  $\Delta\Omega_r > \Delta\omega_0$ .

This is very well illustrated by a very nice series of measurements taken at the Heidelberg storage ring TSR, for various intensity  $^{12}\text{C}^{6+}$  ions stored at 140 MeV/c in the presence of electron cooling. These measurements are shown in Figure 3. The distortion of the noise spectrum as  $N$  is increased, follows the outline given above and in the zone  $N > N_{cr}$ , the peak splitting,  $2\Delta\Omega_r$ , indeed scales as  $\sqrt{N}$ . From these spectra we can estimate that  $N_{cr} \approx 1 \times 10^8$  for  $^{12}\text{C}^{6+}$  ions in this case.

Similar results have been seen in other electron cooling experiments. Figure 4 shows longitudinal noise spectra taken from the LEAR at CERN during electron cooling experiments at 105 MeV/c.

Rearranging (21) one can calculate the imaginary part of the impedance  $Z_i/n$ :

$$\frac{Z_i}{n} = \frac{\Delta\Omega_r^2 2\pi\beta^2 E}{e^2 \eta \omega_0^2 n^2 N} \quad (23)$$

The observation of these two peaks at 105, 200 and 309 MeV/c during the LEAR electron cooling experiments on proton beams [11,12] gives 6900 Ohms at 105 MeV/c, 4100 ohms at 200 MeV/c, and 1600 ohms at 309 MeV/c. This is in reasonable agreement with the values of  $Z_1/n$ , which can be estimated for LEAR as 4500, 2250, and 1450 ohms at 105, 200 and 309 MeV/c respectively. These values are estimated using the formula from [9].

$$|Z_1|/n = Z_0 g / 2\beta\gamma^2 \text{ where } Z_0 = 377 \text{ ohms and } g \approx 2.4.$$

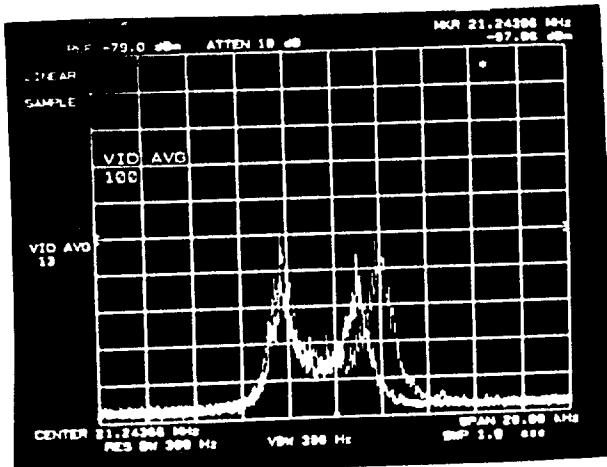


Fig. 4 Typical longitudinal noise spectra observed in LEAR at 105 MeV/c for protons with electron cooling (the two curves are for 1.2 and  $0.6 \cdot 10^9$  particles respectively).

\* \* \*

#### REFERENCES

- [1] S. Chatopadhyay, Some fundamental aspects of fluctuations and coherence in charged particle beams in storage rings. CERN 84-11 (1984)
- [2] V. Parkhomchuk & D. Pestrikov, Thermal noise in an intense beam in a storage ring, Sov. Phys. Tech. Phys. 25(7) (1980).
- [3] S. Cocher & I. Hofmann, On the stability and diagnostics of heavy ions in storage rings with high phase space density, GSI 89-61 (1989)
- [4] D. Boussard, Schottky Noise and beam transfer function diagnostics. CAS Oxford, England September 1985. CERN 87-03 Vol 2 (1987)
- [5] W. Schottky, Über spontane Schwanungerscheinungen in verschiedenen Elektrizitätshalbleitern. Ann. Phys. 57, 541 (1918)
- [6] F. Reif, Statistical Physics: Berkeley Physics Course. Macgraw Hill (1985)
- [7] A. Hofmann, Single beam collective phenomena. Theoretical aspects of the behaviour of beams in accelerators and storage rings, Erice, November 1977. CERN 77-13 (1977)

- [8] D. Mohl, Limitations and instabilities in cooled beams., Proceedings of the Workshop on Crystalline Ion Beams, Eds R. W. Hasse, I. Hofmann, D. Liesen. Wertheim, Fed. Rep. of Germany, October 1988
- [9] J-L. Laclare, Some limitations in LEAR. Physics at LEAR with low-energy cooled antiprotons, Erice, May 1982, Eds U. Gastaldi & R. Klapisch, Plenum Press (1984)
- [10] G. Budker et al., Proc. 10th International Conference on High Energy Charged Particle Accelerators, Serpukhov, USSR (1977), p. 498
- [11] H. Poth et al., First results of electron cooling experiments at LEAR, Z. Phys. A. 332 (1989)
- [12] J. Bosser et al., Electron cooling tests at LEAR, CERN/PS/AR Note 88-21 (1988)
- [13] J. Bosser et al. Summary of the MD on comparative measurements of electron cooling with protons and anti-protons. CERN/PS/AR Note 89-6 (1989)
- [14] The TSR group, First electron cooling of heavy ions at the new Heidelberg storage ring TSR, presented at the XIV Int. Conf. on High Energy Accelerators Tsukuba, Japan August 1989, MPI V33 (1989)
- [15] H. Poth, Review of electron cooling experiments. Presented at ECOOL 1984, Karlsruhe, Federal Republic of Germany. September 1984
- [16] R. D. Kohaupt, Landau damping, at this School.



## LASER COOLING

*Ejvind Bonderup*

Institute of Physics, Aarhus University, Aarhus, Denmark

### ABSTRACT

The velocity distribution of an ionic or atomic beam may be modified through interaction with a laser field with a frequency close to a transition frequency in the ion or atom. Efficient longitudinal cooling may be achieved in an intense field forming a standing wave in a reference frame moving with the average particle velocity. The chapter concentrates on the theoretical description of this effect, and experimental results are presented only briefly towards the end. The origin of the force on an atom from a light wave is first explained in terms of the Lorentz force. In a strong field, quantitative evaluations may conveniently be performed within the dressed-atom picture, based on the eigenstates for the atom-laser system. At low velocities of the particle relative to the standing wave, the force may be interpreted in terms of motion on potential curves with occasional interruptions owing to spontaneous emission. In the final section is a summary of how the force depends on laser detuning and intensity and on relative velocity, and typical numerical values are given for key quantities.

### 1. INTRODUCTION

This chapter deals with the possibility of decreasing the longitudinal velocity spread of a beam of atoms or ions through the interaction of the particles with a strong laser field with a frequency close to a transition frequency in the atom or ion. The experimental investigation of this method of modifying particle beams has just begun, and at present it is not clear to what extent laser cooling may become a supplement to electron cooling and stochastic cooling in the future. Towards the end of the chapter, the results of two experiments will be mentioned but the emphasis will be on the description of the mechanism behind the effect.

Laser cooling can be applied only to selected atoms and ions since the states involved in the laser induced transitions must form a so-called closed system, i.e., spontaneous transitions to other states should be neg-

ligible. To obtain a symmetric compression of the longitudinal velocity distribution of a particle beam, one must apply two laser waves, travelling in the directions parallel and antiparallel to the particles. In the laboratory, the frequencies of the two travelling waves should differ by such an amount that the Doppler-shifted frequencies become equal in a reference frame moving with the average particle velocity. In the moving frame, the superposition of the travelling waves with the same frequency leads to a standing wave, and the basic problem is therefore to understand the influence of a standing laser field on particles moving with velocities of the order of the velocity spread in the beam. The laser affects only particles moving sufficiently slowly with respect to the standing wave. The velocity region accessible to laser cooling increases with the detuning of the radiation frequency away from the atomic transition frequency, provided the increase in detuning is accompanied by a sufficient increase in light intensity. Typically, velocities of  $10^3$  m/sec may be damped out during an interaction time of  $10^{-3}$  seconds.

We first consider an atom at rest, and the origin and direction of the force from a light wave are explained within a simple classical model with an electron bound in a harmonic oscillator. For weak fields, the transition to quantum mechanics then corresponds to the replacement of the binding frequency by a transition frequency and the introduction of an oscillator strength, as in the theory of atomic polarizability. The difference between the forces from a travelling wave and a standing wave of the same strength is illustrated. A transparent description of the force on an atom moving through a strong standing laser field may be given in terms of the so-called dressed-atom states which are eigenstates of the interacting atom-laser system, and this picture will be introduced and discussed in some detail. Finally, we summarize how the direction and strength of the force depend on the particle velocity with respect to the standing wave, on the light intensity, and on the detuning of the laser frequency.

## 2. ORIGIN OF FORCES

### 2.1 Classical model

Consider an electron with charge  $-e$  and mass  $m$  bound in a harmonic oscillator with frequency  $\omega_0$ , and damping constant  $\Gamma$ . In addition to the elastic force, the electron experiences the electric and magnetic fields from a light wave. Assuming a linearly polarized plane wave, we may write the electric field in the form

$$\vec{E}(z,t) = \vec{\epsilon}_x \frac{1}{2} \{ E(z)e^{-i\omega t} + E^*(z)e^{i\omega t} \} = \vec{\epsilon}_x \text{Re} \{ E(z)e^{-i\omega t} \}, \quad (1)$$

where  $\vec{\epsilon}_x$  denotes a unit vector in the  $x$  direction. The two cases  $E(z)=E_0 \cos(kz)$  and  $E(z)=E_0 \exp(ikz)$ , with  $E_0$  real, correspond to a standing wave and a travelling wave, respectively. The wavelength of the light is assumed to be long compared to the displacement of the electron from its equilibrium position  $(0,0,z)$ , such that the fields may be evaluated at this fixed point in space. For nonrelativistic particle motion, the magnetic force is much weaker than the electric force, and we therefore at first neglect it. To this approximation, the electron displacement  $x$  satisfies the equation of motion

$$\ddot{x} + \Gamma \dot{x} + \omega_0^2 x = - \frac{e}{m} \operatorname{Re} \{ E(z) e^{-i\omega t} \}. \quad (2)$$

After the disappearance of transients, we obtain for the dipole moment  $\vec{d}(t)=d(t)\vec{\epsilon}_x$ ,

$$d(t) = -ex(t) = \operatorname{Re} \{ \alpha(\omega) E(z) e^{-i\omega t} \}, \quad (3)$$

where the complex polarizability  $\alpha(\omega)$  is given by the expression

$$\alpha(\omega) = \frac{e^2/m}{\omega_0^2 - \omega^2 - i\omega\Gamma} . \quad (4)$$

The time average of the electric force on an electron vanishes. This is in contrast to the magnetic part of the Lorentz force which results from the interaction of the induced oscillating dipole with the magnetic field from the light wave. According to the Maxwell equation  $\operatorname{curl} \vec{E} = -c^{-1} \partial \vec{B} / \partial t$ , the electric field in Eq. (1) is accompanied by the magnetic field

$$\vec{B}(z,t) = -\vec{\epsilon}_y \operatorname{Re} \left\{ \frac{ic}{\omega} \frac{\partial E(z)}{\partial z} e^{-i\omega t} \right\} . \quad (5)$$

Denoting a time average by the symbol  $\langle \rangle$ , we then obtain the force

$$\vec{F} = -\frac{e}{c} \langle \dot{x} \vec{\epsilon}_x \times \vec{B}(z,t) \rangle = \frac{1}{2} \operatorname{Re} \{ \alpha(\omega) E(z) \frac{\partial E^*(z)}{\partial z} \} \vec{\epsilon}_z . \quad (6)$$

## 2.2 Transition to quantum mechanics in weak fields

A quantal treatment of an atom, perturbed by a weak field of the form in Eq. (1), leads to a very similar polarizability [1],

$$\alpha(\omega) = \frac{e^2}{m} \sum_n \frac{f_{n0}}{\omega_{n0}^2 - \omega^2 - i\omega\Gamma_n} , \quad (7)$$

where  $\omega_{n0}$  is the transition frequency from the ground state  $|0\rangle$  to the excited state  $|n\rangle$  of linewidth  $\Gamma_n$ , and where the quantity  $f_{n0}$ , proportional to the square of a dipole matrix element, is the dipole oscillator strength,

$$f_{n0} = \frac{2m}{\hbar} \omega_{n0} |\langle n | \sum_{j=1}^z x_j | 0 \rangle|^2 . \quad (8)$$

Here  $x_j$ ,  $j=1,2,\dots,z$ , denotes a coordinate of the  $j$ 'th electron, and the oscillator strengths fulfil the sum rule

$$\sum_n f_{n0} = z . \quad (9)$$

### 2.3 Scattering force in travelling wave

In a travelling wave, with  $E(z)=E_0 \exp(ikz)$ , Eq. (6) leads to the scattering force or radiation pressure

$$\vec{F}_{sc} = \frac{1}{2} E_0^2 k \operatorname{Im} \alpha(\omega) \vec{\epsilon}_z . \quad (10)$$

As the names indicate, this force is related to the momentum transfer resulting from the scattering of the light wave by the atom. Energy and momentum are flowing in the direction of motion of the travelling wave, indicated by the unit vector  $\vec{\epsilon}_z$ , and an energy transfer  $\Delta E$  to the atom is accompanied by a momentum transfer  $(\Delta E/c)\vec{\epsilon}_z$ . In a weak field, the energy absorbed by the atom is reemitted in spontaneous decays but since this emission is of equal strength in opposite directions, there is, on the average, no momentum change associated with these processes. The magnitude of the force on the atom therefore equals  $P/c$ , where  $P$  denotes the power transfer from the light when the electric field drives the oscillating dipole, i.e.,

$$\vec{F}_{sc} = \frac{1}{c} P \vec{\epsilon}_z = \frac{1}{c} \langle -e E(z,t) \hat{x} \rangle \vec{\epsilon}_z . \quad (11)$$

From this formula we immediately retrieve Eq. (10).

For a one-electron atom, we now introduce the expressions (7) and (8) into Eq. (10). If the light frequency  $\omega$  is close to the transition frequency  $\omega_{n0}$  for a closed two-state system, we find the approximate formula

$$\vec{F}_{sc} = \frac{1}{4} \frac{e^2}{\hbar c} \frac{\Gamma_n \omega_{n0} |x_{n0}|^2}{\delta^2 + \Gamma_n^2 / 4} E_0^2 \vec{\epsilon}_z , \quad (12)$$

where the constant  $e^2/\hbar c$  equals the finestructure constant, and where we have introduced the detuning

$$\delta \equiv \omega - \omega_{n0} . \quad (13)$$

On the average, a momentum transfer from a travelling wave to an atom occurs only if the absorption is followed by spontaneous and not by induced emission since the induced radiation is emitted in the direction  $\vec{\epsilon}_z$ . As a result, the scattering force in Eq. (12) must saturate when the absorption

rate  $F_{sc}/(\hbar\omega_{n0}/c)$  becomes of the order of the spontaneous emission rate  $\Gamma_n$ , i.e., when the Rabi frequency  $\omega_R = eE_0|x_{n0}|/\hbar$  becomes of the order

$$\omega_R = eE_0|x_{n0}|/\hbar = 2^{1/2}(\delta^2 + \Gamma_n^2/4)^{1/2}. \quad (14)$$

At this electric field strength, we are at the limit of the perturbation treatment since the population of the excited state becomes significant. With the numerical factor  $2^{1/2}$  in Eq. (14), a combination of the equations (12) and (14) leads to the correct value of the saturated scattering force,

$$F_{sc}^{\text{sat}} = \frac{1}{2} \frac{\hbar\omega}{c} \Gamma_n, \quad (15)$$

where the factor 1/2 accounts for the population of the upper level at saturation.

At this point it may be appropriate to give the order of magnitude of important quantities. With a photon energy  $\hbar\omega=2\text{eV}$  and a lifetime  $\Gamma_n^{-1}=10^{-8}$  sec we obtain  $F_{sc}^{\text{sat}} \sim 0.5 \cdot 10^{-19}$  newton. For a projectile of mass number A, the corresponding acceleration is  $\sim 0.3 \cdot A^{-1} \cdot 10^8$  m/sec<sup>2</sup>, such that within a microsecond a neon atom experiences a velocity change of  $\sim 1.5$  m/sec. For a two-state system we have a relation between the matrix element  $x_{n0}$  and the linewidth  $\Gamma_n$  in Eq. (14),  $\Gamma_n = (4/3)\alpha\omega^3|x_{n0}|^2/c^2$ , where  $\alpha$  is the finestructure constant. In the case considered, we then find that at resonance ( $\delta=0$ ), saturation is obtained at an electric field strength of  $\sim 260$  V/m. This again corresponds to an energy flux of  $0.9 \cdot 10^{-4}$  Watt/mm<sup>2</sup>.

#### 2.4 Dipole force in standing wave

We now turn to a standing wave, and with  $E(z)=E_0 \cos(kz)$  we obtain from the general formula (6) the so-called dipole force

$$\vec{F}_d = \frac{1}{4} \{\text{Re } \alpha(\omega)\} \frac{\partial}{\partial z} (E_0 \cos(kz))^2 \vec{\epsilon}_z. \quad (16)$$

Since this force is proportional to the real part of the polarizability, the interpretation in terms of absorption and emission is quite different from the one given in connection with the scattering force. If the standing wave is considered as the superposition of two waves travelling in opposite directions, i.e.,  $E(z)=E_0(\exp(ikz)+\exp(-ikz))/2$ , the force in Eq. (6) results from the cross terms whereas the two direct contributions cancel. This indicates that the dipole force is associated with the momentum change in processes where absorption from one of the waves is followed by induced emission into the other one.

When the frequency  $\omega$  is close to the atomic transition frequency  $\omega_{n0}$  in a closed two-state system, the dipole force may be expressed as

$$\vec{F}_d = -\frac{1}{8} \frac{e^2}{m\omega_{n0}} f_{n0} \frac{\delta}{\delta^2 + \Gamma_n^2/4} \frac{\partial}{\partial z} (E(z))^2 \vec{\epsilon}_z . \quad (17)$$

The sign of the dipole force thus depends on the sign of the detuning: For blue detuning, the force is in the direction out of the strong electric field intensity. A comparison of the magnitudes of the dipole force in a standing wave and the scattering force in a travelling wave of equal amplitude and at equal detuning may be obtained from the Eqs. (12) and (17), and replacing the derivative  $\partial/\partial z$  by the wave number  $k$  of the light wave, we find

$$F_d/F_{sc} \sim |\delta|/\Gamma_n . \quad (18)$$

The scattering force from a travelling wave saturates at the value in Eq. (15), and if the detuning is large compared to the linewidth, this occurs when the Rabi frequency becomes of order  $|\delta|$ . As we shall see later, from Eq. (51), the dipole force has also reached saturation in the situation  $\omega_R > |\delta| > \Gamma_n$ , and the large factor  $|\delta|/\Gamma_n$  in Eq. (18) then represents the ratio of the saturated dipole force  $\sim |\delta| \hbar \omega / (2c)$  and the saturated scattering force (15).

Although the dipole force can become strong, one might not expect it to cool a beam of particles travelling through the standing wave since the expression (17) is periodic as a function of position, with vanishing average value. The reason why cooling does occur is that the forces become velocity dependent. The motion of the particle may only be disregarded as long as the spatial variation of the fields is negligible during the time necessary for the disappearance of transients, i.e., for the attainment of local statistical equilibrium. In strong fields, this time is of order  $\Gamma_n^{-1}$ , and in Chapter 4 we shall find an appreciable dipole force on an atom moving through a strong standing laser field already at velocities quite low compared to  $\Gamma_n/k$ , corresponding to motion through a distance  $k^{-1}$  during the time  $\Gamma_n^{-1}$ .

### 3. ATOM IN RADIATION FIELD

#### 3.1 Quantization of radiation field

As an introduction to the dressed-atom picture, we briefly recall the quantization of the radiation field in the absence of charges [2]. In Coulomb gauge,  $\text{div}\vec{A}=0$ , the electric and magnetic fields may be expressed in terms of the vector potential,

$$\vec{E}(\vec{r}, t) = -\frac{1}{c} \frac{\partial \vec{A}(\vec{r}, t)}{\partial t} , \quad \text{and} \quad (19)$$

$$\vec{B}(\vec{r}, t) = \text{curl } \vec{A}(\vec{r}, t) . \quad (20)$$

Introducing a spatial Fourier decomposition of the vector potential within the quantization volume  $V$ ,

$$\vec{A}(\vec{r}, t) = \frac{1}{V} \sum_{\vec{k}} \vec{A}_{\vec{k}}(t) e^{i\vec{k} \cdot \vec{r}}, \quad \vec{k} \cdot \vec{A}_{\vec{k}}(t) = 0 , \quad (21)$$

we obtain for the energy in the radiation field,

$$H_r = \frac{1}{8\pi} \int d^3 r (\vec{E}^2 + \vec{B}^2) = \frac{1}{8\pi} \sum_{\vec{k}} (k^2 |\vec{A}_{\vec{k}}|^2 + \frac{1}{c^2} |\dot{\vec{A}}_{\vec{k}}|^2) . \quad (22)$$

This expression simplifies if we introduce the function

$$\vec{f}_{\vec{k}} = \vec{A}_{\vec{k}} + \frac{i}{kc} \dot{\vec{A}}_{\vec{k}} . \quad (23)$$

Since the vector potential  $\vec{A}(\vec{r}, t)$  is real, the Fourier coefficients  $\vec{A}_{\vec{k}}$  fulfil the relation  $\vec{A}_{-\vec{k}}^* = \vec{A}_{\vec{k}}$ , and for the energy  $H_r$  we therefore have

$$H_r = \frac{1}{8\pi} \sum_{\vec{k}} k^2 |\vec{f}_{\vec{k}}|^2 . \quad (24)$$

The vector  $\vec{f}_{\vec{k}}$  is perpendicular to  $\vec{k}$ , and thus it may be written in the form

$$\vec{f}_{\vec{k}}(t) = \left( \frac{4\pi c}{k} \right)^{1/2} \sum_{\mu=1}^2 (q_{\vec{k}\mu}(t) + i p_{\vec{k}\mu}(t)) \vec{\epsilon}_{\vec{k}\mu} , \quad (25)$$

where  $\{\vec{\epsilon}_{\vec{k}1}, \vec{\epsilon}_{\vec{k}2}, \vec{k}/k\}$  is an orthonormal set, and where  $q_{\vec{k}\mu}$  and  $p_{\vec{k}\mu}$  are real. They fulfil very simple equations of motion which are obtained from the wave equation satisfied by the vector potential,

$$\Delta \vec{A} - \frac{1}{c^2} \frac{\partial^2 \vec{A}}{\partial t^2} = \vec{0} . \quad (26)$$

The equivalent equation  $\ddot{\vec{A}}_{\vec{k}} = -k^2 c^2 \vec{A}_{\vec{k}}$  gives for the function  $\vec{f}$  in Eq. (23),

$$\dot{\vec{f}}_{\vec{k}} = -ikc \vec{f}_{\vec{k}} , \quad (27)$$

and this relation may immediately be rewritten in terms of the quantities  $q_{\vec{k}\mu}$  and  $p_{\vec{k}\mu}$ ,

$$\dot{q}_{\vec{k}\mu} = kc p_{\vec{k}\mu} , \quad (28)$$

$$\dot{p}_{\vec{k}\mu} = -kc q_{\vec{k}\mu} .$$

As is easily verified, these equations of motion are reproduced by the canonical equations with the Hamiltonian equal to the field energy  $H_r$ , written as a function of the canonical variables  $q_{\vec{k}\mu}$  and conjugate momenta  $p_{\vec{k}\mu}$ ,

$$H_r = \sum_{\vec{k}\mu} \frac{1}{2} \hbar c (p_{\vec{k}\mu}^2 + q_{\vec{k}\mu}^2) . \quad (29)$$

To make the transition to quantum mechanics, we introduce the canonical commutation relations

$$[q_{\vec{k}\mu}, p_{\vec{k}'\mu'}] = i\hbar \delta_{\vec{k}\vec{k}'} \delta_{\mu\mu'} . \quad (30)$$

Finally we define photon annihilation operators

$$a_{\vec{k}\mu} = (2\hbar)^{-1/2} (q_{\vec{k}\mu} + ip_{\vec{k}\mu}) , \quad (31)$$

and in terms of these operators and their adjoints  $a_{\vec{k}\mu}^\dagger$  the Hamiltonian takes the form

$$H_r = \sum_{\vec{k}\mu} \hbar c (a_{\vec{k}\mu}^\dagger a_{\vec{k}\mu} + \frac{1}{2}) . \quad (32)$$

To describe the interaction of an atom with the radiation field, we also need the expression for the vector potential  $\vec{A}$  in terms of creation and annihilation operators. From the relation  $\vec{A}_{-\vec{k}}^* = \vec{A}_{\vec{k}}$  we find that  $\vec{A}_{\vec{k}}$  equals  $(\vec{f}_{\vec{k}} + \vec{f}_{-\vec{k}}^*)/2$ , and the equations (21), (25), and (31) then yield

$$\begin{aligned} \vec{A} &= \frac{1}{\sqrt{V}} \sum_{\vec{k}} \frac{1}{2} (\vec{f}_{\vec{k}} e^{i\vec{k} \cdot \vec{r}} + \vec{f}_{\vec{k}}^* e^{-i\vec{k} \cdot \vec{r}}) \\ &= \sum_{\vec{k}\mu} \left( \frac{2\pi\hbar c}{V\hbar} \right)^{1/2} [a_{\vec{k}\mu} e^{i\vec{k} \cdot \vec{r}} + a_{\vec{k}\mu}^\dagger e^{-i\vec{k} \cdot \vec{r}}] \vec{\epsilon}_{\vec{k}\mu} . \end{aligned} \quad (33)$$

### 3.2 Dressed atoms

A canonical description of an electron moving in a potential  $V(\vec{r})$  and interacting with the radiation field is based on the Hamiltonian

$$H = \frac{1}{2m} (\vec{p} + \frac{e}{c} \vec{A}(\vec{r}, t))^2 + V(\vec{r}) + H_r , \quad (34)$$

where  $\vec{p}$  denotes the canonical momentum of the particle, and where the Hamiltonian  $H_r$  for the radiation field is given in Eq. (29), and in quantized form in Eq. (32). Neglecting a very small term proportional to  $\vec{A}^2$ , we may, in Coulomb gauge, rewrite the operator  $H$  for a one-electron system in the form

$$H = \frac{\vec{p}^2}{2m} + V + H_r + \frac{e}{mc} \sum_{\vec{k}\mu} \left( \frac{2\pi\hbar c}{V\hbar} \right)^{1/2} [a_{\vec{k}\mu} e^{i\vec{k} \cdot \vec{r}} + a_{\vec{k}\mu}^\dagger e^{-i\vec{k} \cdot \vec{r}}] \vec{\epsilon}_{\vec{k}\mu} \cdot \vec{p} . \quad (35)$$

We now specialize to an atom interacting with a laser field in a single mode. At first, spontaneous emission is neglected such that only a single

term need be retained in the summation in Eq. (35), and for brevity we drop the indices  $\vec{k}\mu$ . The eigenstates of the Hamiltonian  $H$  are then linear combinations of product states of the type  $|a\rangle|n\rangle$ , where the first factor is an eigenstate of the atomic Hamiltonian  $H_a = \vec{p}^2/2m + V$ , and where  $|n\rangle$  is the state describing the presence of  $n$  laser photons. Let the laser frequency be close to the transition frequency  $(E_e - E_g)/\hbar$  in a closed atomic two-state system with upper and lower states  $|e\rangle$  and  $|g\rangle$ . The energies of the coupled atom-laser system may then be obtained from diagonalization of the Hamiltonian  $H$  within the two-dimensional subspaces spanned by the nearly degenerate product states  $|g\rangle|n+1\rangle$  and  $|e\rangle|n\rangle$ , i.e., from diagonalization of the  $2 \times 2$  matrices

$$\underline{H}_n = \begin{Bmatrix} E_g + \hbar\omega(n+\frac{3}{2}) & M_n^* \\ M_n & E_e + \hbar\omega(n+\frac{1}{2}) \end{Bmatrix} . \quad (36)$$

Here the interaction matrix element  $M_n$  is given by the expression

$$M_n = \frac{e}{m} \left( \frac{2\pi\hbar}{V\omega} \right)^{1/2} e^{i\vec{k} \cdot \vec{R}} \vec{\epsilon} \cdot \langle e | \vec{p} | g \rangle \sqrt{n+1} , \quad (37)$$

where the vector  $\vec{R}$  indicates the position of the atomic nucleus. In Eq. (35) the variation of the exponential over atomic distances has been neglected (dipole approximation). The commutator relation  $[x, H_a] = i\hbar p_x/m$  leads to a simple connection between matrix elements of the momentum operator and of the position operator,

$$\langle e | \vec{p} | g \rangle = im\omega_0 \langle e | \vec{r} | g \rangle , \quad (38)$$

where  $\omega_0$  is the transition frequency,  $\hbar\omega_0 = E_e - E_g$ . Applying this equation in the expression for the off-diagonal element  $M_n$ , we may write the eigenvalues of the matrix  $\underline{H}_n$  in the form

$$E_n = \frac{1}{2} [E_g + \hbar\omega(n+\frac{3}{2}) + E_e + \hbar\omega(n+\frac{1}{2})] \pm \frac{\hbar}{2} \sqrt{\delta^2 + \omega_R^2} . \quad (39)$$

The detuning  $\delta$  is defined as in Eq. (13), with  $\omega_{n0}$  replaced by  $\omega_0$ , and the Rabi frequency  $\omega_R$  is given by the formula

$$\hbar\omega_R = 2|\vec{\epsilon} \cdot \vec{d}| \left( \frac{2\pi\hbar\omega}{V} \right)^{1/2} \sqrt{n+1} , \quad (40)$$

where the dipole matrix element  $\vec{d}$  equals  $-e\langle e | \vec{r} | g \rangle$ . In Eq. (40) no distinction is made between the frequencies  $\omega$  and  $\omega_0$ .

To make the connection to the Rabi frequency defined in Eq. (14), we first show that in the classical limit the single mode laser field under consideration corresponds to a travelling wave of the type leading to Eq. (14). In fact, from the Eqs. (19), (33), (27), and (25) we find, omitting

the fixed values of the indices  $\vec{k}\mu$ ,

$$\begin{aligned} \vec{E}(\vec{r}, t) &= -V^{-1/2} k \operatorname{Im} \{ \vec{f}(0) e^{i(\vec{k} \cdot \vec{r} - kct)} \} \\ &= -\vec{\epsilon} \left( \frac{4\pi\omega}{V} \right)^{1/2} [p(0)\cos(\vec{k} \cdot \vec{r} - \omega t) + q(0)\sin(\vec{k} \cdot \vec{r} - \omega t)] . \end{aligned} \quad (41)$$

For the travelling wave discussed previously,  $\vec{E}(\vec{r}, t) = \vec{\epsilon}_x E_0 \cos(kz - \omega t)$ , the energy density, averaged in space and time, equals  $E_0^2/8\pi$ . If this value is identified with the quantal result  $n\hbar\omega/V$ , corresponding to the presence of  $n$  photons, the Rabi frequencies introduced in the Eqs. (14) and (40) indeed become equal.

As expected, the dressed-atom picture may be developed for any kind of laser mode, and not only for travelling plane waves. The classical fields are then expanded in another orthonormal system than the complex exponentials in Eq. (21) but the transformation of the equations of motion to canonical form with a Hamiltonian describing a set of uncoupled harmonic oscillators may be carried through with any set of orthonormal basis functions [3]. The dressed-atom energies are obtained through an appropriate replacement of the function  $V^{-1/2} \exp(i\vec{k} \cdot \vec{R})$  in Eq. (37). In the case of a standing wave, the function is of the form  $(2/V)^{1/2} \cos(\vec{k} \cdot \vec{R})$ , and the right-hand side of Eq. (40) is multiplied by the factor  $\sqrt{2}\cos(\vec{k} \cdot \vec{R})$ . A classical field

$$\vec{E}(\vec{r}, t) = \vec{\epsilon}_x E_0 \cos(\vec{k} \cdot \vec{r}) \cos(\omega t) \quad (42)$$

corresponds to an average energy density  $E_0^2/16\pi$ , and it therefore leads to the Rabi frequency

$$\omega_R(\vec{R}) = \hbar^{-1} |\vec{d} \cdot \vec{\epsilon}_x| E_0 \cos(\vec{k} \cdot \vec{R}) . \quad (43)$$

Strictly speaking, the Rabi frequency depends on the number of photons present but in laser fields the average number  $\bar{n}$  is very large, e.g.,  $\bar{n} \gtrsim 10^{10}$ , and the corresponding relative fluctuation  $(\bar{n})^{-1/2}$  is therefore so small that this dependence may be neglected. For an atom in the standing wave (42), with wave vector  $\vec{k}$  in the  $z$  direction, we then obtain the dressed-atom energies, setting the energy  $E_g + \frac{1}{2}\hbar\omega$  equal to zero, [4]

$$\begin{aligned} E_{1n} &= (n+1)\hbar kc - \frac{1}{2} \hbar\delta + \frac{1}{2} \hbar\Omega(z) , \\ E_{2n} &= (n+1)\hbar kc - \frac{1}{2} \hbar\delta - \frac{1}{2} \hbar\Omega(z) . \end{aligned} \quad (44)$$

Here,  $z$  denotes the  $z$  coordinate of the atomic position, and  $\Omega(z)$  the generalized Rabi frequency

$$\Omega(z) = (\delta^2 + |\vec{d} \cdot \vec{\epsilon}_x|^2 E_0^2 \cos^2(kz)/\hbar^2)^{1/2}. \quad (45)$$

With an appropriate choice of the relative phase of the atomic states  $|e\rangle$  and  $|g\rangle$ , the off-diagonal elements in the matrix  $H_n$  in Eq. (36) become real with the value  $M_n = \hbar\omega_R(z)/2$ , and the eigenstates of  $H_n$  are then given by the expressions

$$|1,n;z\rangle = C(z)|e\rangle|n\rangle + S(z)|g\rangle|n+1\rangle, \quad (46)$$

$$|2,n;z\rangle = -S(z)|e\rangle|n\rangle + C(z)|g\rangle|n+1\rangle,$$

with

$$C(z) = \{[1 - \delta/\Omega(z)]/2\}^{1/2}, \quad (47)$$

$$S(z) = \{[1 + \delta/\Omega(z)]/2\}^{1/2}.$$

#### 4. COOLING IN STANDING WAVE

##### 4.1 Application of the dressed-atom picture

The dipole force acting on an atom in a standing laser field has a very simple interpretation within the dressed-atom picture [4]. The energies  $E_{1n}$  and  $E_{2n}$  in Eq. (44) depend on atomic position, and therefore the atom experiences a force in the  $z$  direction

$$F_{in} = (-1)^i \frac{1}{2} \hbar \frac{\partial \Omega}{\partial z}, \quad i = 1, 2. \quad (48)$$

The sign, but not the magnitude of the force, depends on whether the system is in the upper state 1 or in the lower state 2.

So far, we have neglected spontaneous emission. The transition amplitude for such an event is proportional to a dipole matrix element  $\langle j, n'; z | \vec{r} | i, n; z \rangle$ , and since the process always involves an atomic transition  $|e\rangle \rightarrow |g\rangle$  and does not affect the number of laser photons present, a dressed state  $|i, n; z\rangle$  can only decay to a state  $|j, n-1; z\rangle$ . The level indices  $i$  and  $j$  can differ or be equal, and from the expressions (46) we find for the corresponding transition rates  $\Gamma_{ji}$

$$\Gamma_{21} = \Gamma C^4(z), \quad (49)$$

$$\Gamma_{12} = \Gamma S^4(z),$$

where  $\hbar\Gamma$  denotes the linewidth of the upper atomic state. In statistical equilibrium, the populations  $\Pi_1$  and  $\Pi_2$  of the upper and lower levels are determined by the relation  $\Pi_1 \Gamma_{21} = \Pi_2 \Gamma_{12}$ , and one obtains

$$\Pi_1 = 1 - \Pi_2 = S^4(z)/(C^4(z) + S^4(z)) . \quad (50)$$

The implicit assumption that populations do not couple to off-diagonal elements of the density matrix is valid for generalized Rabi frequencies  $\Omega(z)$  large compared to  $\Gamma$  [5].

For positive detuning,  $\delta > 0$ , the ratio  $S/C$  is larger than unity, according to the expressions (47), and levels of type 1 are more heavily populated than levels of type 2. As a result, the average force obtained from the equation (48) is in the direction away from regions of high electric field intensity, in agreement with Eq. (17). The average force is given by the formulas

$$F = -\frac{1}{2} \hbar \frac{\partial \Omega}{\partial z} (\Pi_1 - \Pi_2) = -\frac{1}{2} \hbar \delta \frac{\partial \Omega^2 / \partial z}{\Omega^2 + \delta^2} . \quad (51)$$

In the weak field limit,  $\omega_R(0) = eE_0 |x_{n0}| / \hbar < |\delta|$ , Eq. (51) reduces to Eq. (17), with the oscillator strength  $f_{n0}$  given by Eq. (8), when we insert the condition  $|\delta| > \Gamma_n$ . In the opposite limit of strong fields,  $|\omega_R(z)| > |\delta|$ , the force saturates at the value  $\frac{1}{2} \hbar k \delta \sin(2kz) / \cos^2(kz)$ .

If the atom moves infinitely slowly in the  $z$  direction, local statistical equilibrium results at every position. According to the Eqs. (45) and (51), the force  $F$  then becomes antisymmetric around the maxima and minima in the electric field energy, and the average value over half a wavelength of the field vanishes. At finite particle velocities  $v$ , however, the system does not have time to relax to local equilibrium, and the populations  $\Pi_i(z, v)$  must be determined from the differential equation

$$\frac{\partial}{\partial z(v)} \begin{Bmatrix} \Pi_1 \\ \Pi_2 \end{Bmatrix} = \begin{Bmatrix} -\Gamma_{21} & \Gamma_{12} \\ \Gamma_{21} & -\Gamma_{12} \end{Bmatrix} \begin{Bmatrix} \Pi_1 \\ \Pi_2 \end{Bmatrix} . \quad (52)$$

The symmetry with respect to the extrema of the electric field energy is now broken, and an average force results. If the wavelength of the field is long compared to the distance travelled by the projectile during the lifetime  $\Gamma^{-1}$  of the upper atomic state, i.e.,  $v < \Gamma/k$ , the populations only differ from the equilibrium values in Eq. (50) by small terms proportional to  $v$ . The average force then also becomes proportional to velocity, and it is straightforward to show that it is retarding (accelerating) for blue (red) detuning.

The spontaneous events determine the populations of the dressed states and thereby the average force. The transitions do not in themselves lead to extraction of kinetic energy from the atom since the energy of an emitted photon is balanced by the decrease in dressed-state energy from  $E_{in}$  to  $E_{j(n-1)}$ . At sufficiently high velocities, nonradiative transitions between dressed states, corresponding to the same value of  $n$ , also become important.

Such a process results in changes in both populations and in kinetic energy. These nonadiabatic transitions will be briefly discussed in Section 4.2

As regards the direction of the average dipole force on a slowly moving atom, we shall give another argument which is based on the simple pictorial representation of the dressed-atom description [6]. In Fig. 1, the energy levels in Eq. (44) are sketched as functions of position for consecutive values of  $n$ . When spontaneous emission is neglected, the system evolves along a single curve. The energy required to climb a hill is taken from the kinetic energy of the atom, but this kinetic energy is fully regained on the way down the other side of the hill. Apart from modulations, the particle therefore maintains its velocity as it travels through the standing wave.

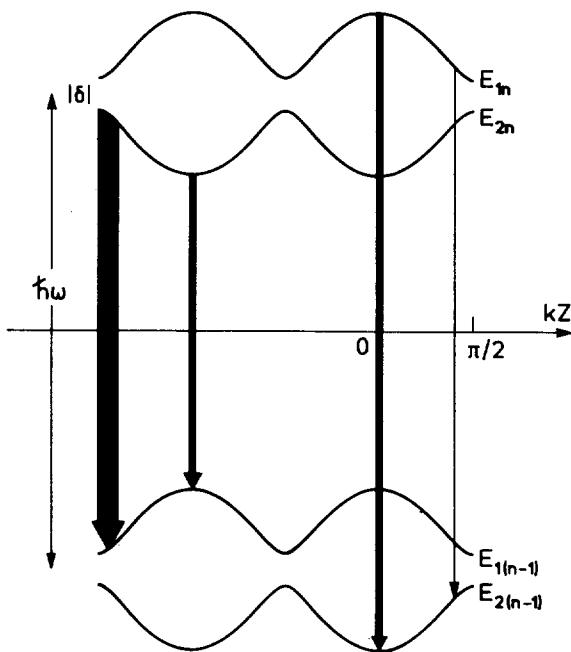


Fig. 1: Two sets of dressed-state energies as functions of position. The relative values indicated for the transition rates  $\Gamma_{21}$  and  $\Gamma_{12}$  at various positions correspond to a positive detuning  $\delta$ , in the limit  $\omega_R(0) \gg \delta$ .

This situation is changed when spontaneous processes are introduced. A decay from a level  $E_{1n}$  to a level  $E_{1(n-1)}$  does not affect the force but a change of level type, from 1 to 2 or vice versa, leads to a reversal of its direction. Let us consider a positive detuning,  $\delta > 0$ . As is observed from the expressions (46) and (47), a system travelling along the upper level  $E_{1n}$  is in the pure product state  $|g\rangle|n+1\rangle$  at the field nodes where  $\Omega$  equals  $\delta$ , and therefore no decay can occur at these positions. The probability of decaying to the energy level  $E_{2(n-1)}$  is proportional to the quantity  $C^4(z)$  which attains its maximum at the hill tops. Here, a decay to a level of type 2 brings the system from an energy maximum to an energy minimum. This does

not only prevent the atom from rolling down the back-side of the hill just climbed and thereby from recovering its kinetic energy. It has to climb yet another hill, to boot! The fate of the particle is analogous to the punishment of Sisyphus [6]. For an atom following a level of type 2, the situation is similar: When the probability for a transition to a level of type 1 is a maximum, now at field nodes, the decay brings the system from a hill top to a valley bottom. The repeated transitions, predominantly from higher to lower positions on the energy curves, are compensated by a gradual decrease in kinetic energy of the atom.

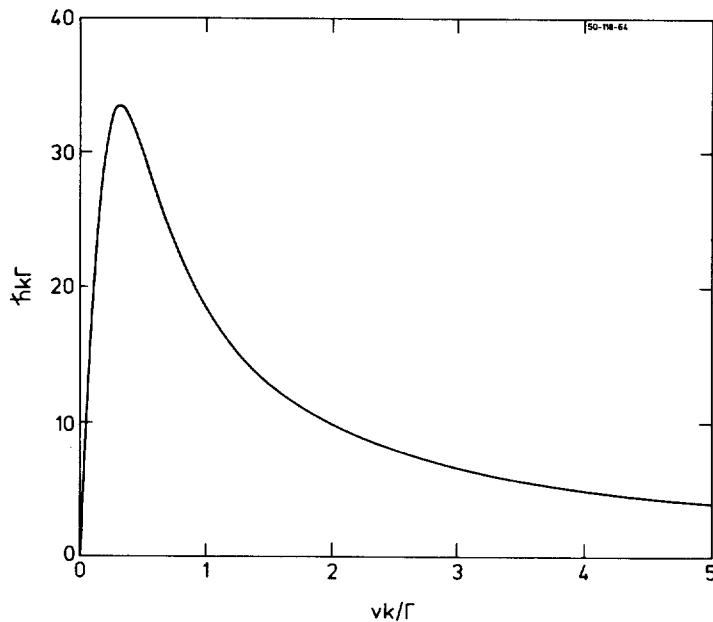
A change of sign on the detuning  $\delta$  leads to an interchange of the functions C and S in Eq. (47). For negative detuning, the spontaneous events therefore, in the main, bring the system upwards on the energy curves, and the atom is accelerated.

In the high velocity limit,  $v > \Gamma/k$ , the system on the average follows an energy curve over a large number of wavelengths of the laser field before decaying. This number is proportional to  $v$ , and the average force, equal to the energy change per unit path length, therefore becomes inversely proportional to  $v$ . Since the average force was found to be velocity proportional in the opposite limit  $v < \Gamma/k$ , we expect a maximum at velocities of the order of  $\Gamma/k$ . This expectation is borne out by the results in Fig. 2 which shows the magnitude of the average dipole force,

$$\bar{F}(v) = -\frac{\hbar}{2} \langle (\Pi_1(z, v) - \Pi_2(z, v)) \partial \Omega / \partial z \rangle \vec{\epsilon}_z . \quad (53)$$

Here, the brackets indicate an average over position. The populations were obtained from numerical solution of Eq. (52) under steady state conditions, i.e., corresponding to populations which have attained the periodicity of the field energy. With the parameter values applied,  $\omega_R(0)/\Gamma=1000$  and  $|\delta|/\Gamma=200$ , the maximum average force equals  $33 \mu k\Gamma$ , which may be compared with the maximum scattering force  $\hbar k\Gamma/2$  in a travelling wave. At a fixed value of the ratio  $\delta/\omega_R(0)$ , the dipole force is proportional to the Rabi frequency  $\omega_R(0)$  in the high velocity region,  $v > \Gamma/k$ .

Figure 3 shows the result of a Monte Carlo simulation of longitudinal cooling of a velocity distribution with a similar width as in an experiment (Ref. 12) to be discussed below [7]. The dressed atom was allowed to jump back and forth between upper and lower energy levels according to the position dependent transition probabilities. The detuning and the strength of the standing laser field had similar values as in Fig. 2, and the interaction time was  $17.2 \mu s$ , corresponding to the passage of a neon atom with an energy of 200 eV through a distance of 75 cm. In the experiment, linearly polarized light induced transitions between lower and upper atomic states with angular momenta  $J_g=2$  and  $J_e=3$ , and it was therefore necessary to generalize the dressed-atom description from a two-level system to a multilevel



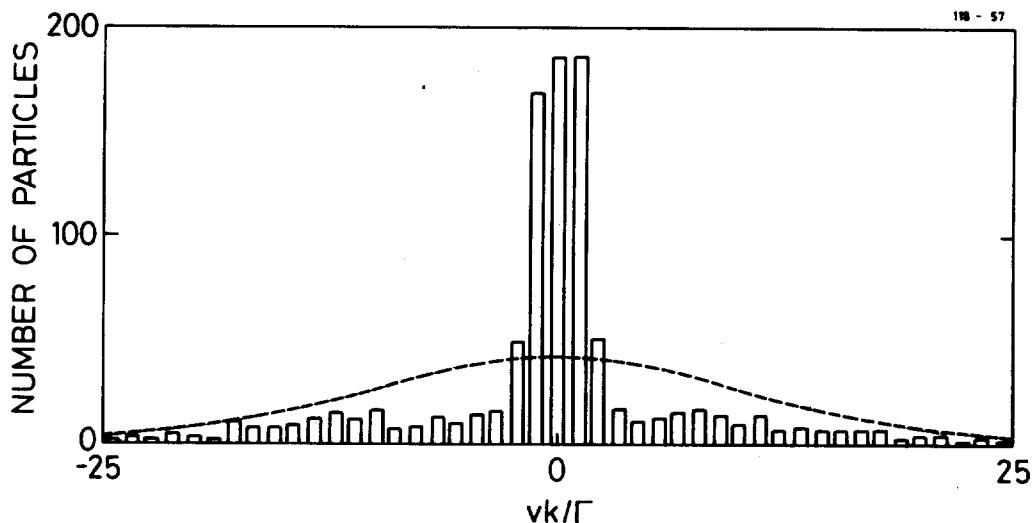
**Fig. 2:** Absolute value of the average dipole force, in units of  $Mk\Gamma$ , as a function of velocity, in units of  $\Gamma/k$ , at a Rabi frequency  $\omega_R(0)=1000\Gamma$  and a detuning  $|\delta|=200\Gamma$ . The force is retarding for  $\delta>0$  and accelerating for  $\delta<0$ .

system, in which the different members of a Zeemann multiplet give rise to different Rabi frequencies. The field strength applied in the calculation corresponded to a standing wave of cross sectional area  $4.2 \cdot 10^{-3} \text{ cm}^2$ , maintained in a resonator by means of two 0.5 W cw lasers. The loss per round trip for a light beam in the resonator was assumed to be 5%. From Fig. 3, very significant cooling is expected under these conditions.

#### 4.2 Nonadiabatic transitions

Spontaneous emission leads to transitions between dressed-atom states  $|i,n;Z\rangle$  and  $|j,n-1;Z\rangle$ . At finite atomic velocity, transitions can also occur between the different dressed states belonging to the same value of  $n$ . These nonadiabatic processes are analogous to Landau-Zener transitions between molecular states applied in the treatment of colliding atoms.

The dressed system is described by a Hamiltonian of the type shown in Eq. (35), with only a single value for the indices  $\vec{k}\mu$  and with the electron coordinates inside the square brackets replaced by the atomic position, as in Eq. (37). When the atom moves along a classical trajectory, the Hamiltonian becomes a function of time. To solve the corresponding time-dependent Schrödinger equation, it is convenient to expand the wave function  $\psi(t)$  for the dressed system in the eigenstates  $|i,n;t\rangle$  of the total Hamiltonian,



**Fig. 3:** Monte Carlo calculation of the transformation of an initial Gaussian distribution in atomic velocity as a result of interaction with a standing laser field during a time  $886\Gamma^{-1}$ . Linearly polarized light induced transitions between lower and upper atomic states with angular momenta  $J_g=2$  and  $J_e=3$ . Quite similar results would be obtained for a two-level system with Rabi frequency  $\omega_R(0)=800\Gamma$  and a positive detuning  $\delta=200\Gamma$ . In the experiment in Ref. 12, the velocity  $\Gamma/k$  equals 5.3 m/sec.

corresponding to the atomic position at time  $t$ . The transitions to be discussed do not change the quantum number  $n$ , and we may therefore limit ourselves to functions of the form

$$\psi(t) = \sum_{j=1}^2 a_j(t) e^{-i \int_0^t E_j(t') dt' / \hbar} |j;t\rangle, \quad (54)$$

where the explicit reference to the fixed value  $n$  has been omitted in the symbols for the dressed states and the corresponding energies. For simplicity, we again consider a two-state atom. Inserting Eq. (54) into the time-dependent Schrödinger equation, we find for the expansion coefficients  $a_k(t)$ ,

$$\frac{da_k(t)}{dt} = - \sum_{j=1}^2 \langle k;t | \frac{\partial}{\partial t} | j;t \rangle a_j(t) e^{-i \int_0^t \omega_{jk}(t') dt'}, \quad (55)$$

where the frequency  $\omega_{12}(t) = (E_1(t) - E_2(t)) / \hbar$  equals the generalized Rabi frequency  $\Omega(t)$ . Both this quantity and the matrix element  $\langle k;t | \partial/\partial t | j;t \rangle$  are essentially unchanged during a time interval  $\Delta t$  short compared to the time  $\sim (kv)^{-1}$  required for travelling a quarter of a wavelength. If the product  $\Omega \Delta t$  is large compared to unity, we then obtain for the probability of

finding the system in state  $j$  at time  $t+\Delta t$ , provided it was in state  $i$  at time  $t$  [4]

$$|a_j(t+\Delta t)|^2 \approx 2|\langle j; t | \frac{\partial}{\partial t} |i; t\rangle|^2 / \Omega^2(t), \quad \Delta t < (kv)^{-1}. \quad (56)$$

Here, a factor  $\sin^2(\Omega\Delta t/2)$  has been replaced by its average value 1/2. At very high velocities,  $kv > \Omega$ , the exponential in Eq. (55) may be approximated by unity, and the formula (56) represents an upper limit since the time  $\Omega^{-1}$  should be replaced by the shorter time interval  $\Delta t$ . To evaluate the time derivative of a dressed state in Eq. (56), we first note that according to their definitions in Eq. (47), the expansion coefficients  $C$  and  $S$  in Eq. (46) may be written in the form [4]

$$C = \cos \theta \quad \text{and}$$

$$S = \sin \theta, \quad (57)$$

where the angle  $\theta$  is given by the expression

$$\cos 2\theta = -\delta/\Omega. \quad (58)$$

The dressed states therefore fulfil the equations

$$\begin{aligned} \frac{\partial}{\partial t} |1; t\rangle &= |2; t\rangle \frac{d\theta}{dt} \quad \text{and} \\ \frac{\partial}{\partial t} |2; t\rangle &= -|1; t\rangle \frac{d\theta}{dt}. \end{aligned} \quad (59)$$

The derivative  $d\theta/dt$  is determined through differentiation of Eq. (58). The generalized Rabi frequency is of the form

$$\Omega(t) = (\delta^2 + \omega_R^2(t))^{1/2} = (\delta^2 + \omega_R^2(0)\cos^2(kvt))^{1/2}, \quad (60)$$

and we find for the probability in Eq. (56),

$$|a_j(t+\Delta t)|^2 \approx \frac{1}{2} \left( \frac{\delta}{\Omega^3(t)} \frac{d\omega_R}{dt} \right)^2 = \frac{1}{2} (\delta \omega_R(0)kv)^2 \frac{\sin^2(kvt)}{(\delta^2 + \omega_R^2(0)\cos^2(kvt))^3}. \quad (61)$$

As expected, the probability for a nonadiabatic transition is largest around the electric field nodes where the distance between the dressed energy levels is a minimum. The condition for neglecting the nonadiabatic transitions in the evaluation of the average dipole force is that the maximum value in Eq. (61) be small compared to the probability for spontaneous emission during the passage of half a wavelength of the laser field. Since this probability is of the order of  $\Gamma\lambda/(2v)$ , the velocity of the atom should be lower than the critical value [4]

$$v_c = (2\pi\delta^4/(\Gamma\omega_R(0))^2)^{1/3} \Gamma/k . \quad (62)$$

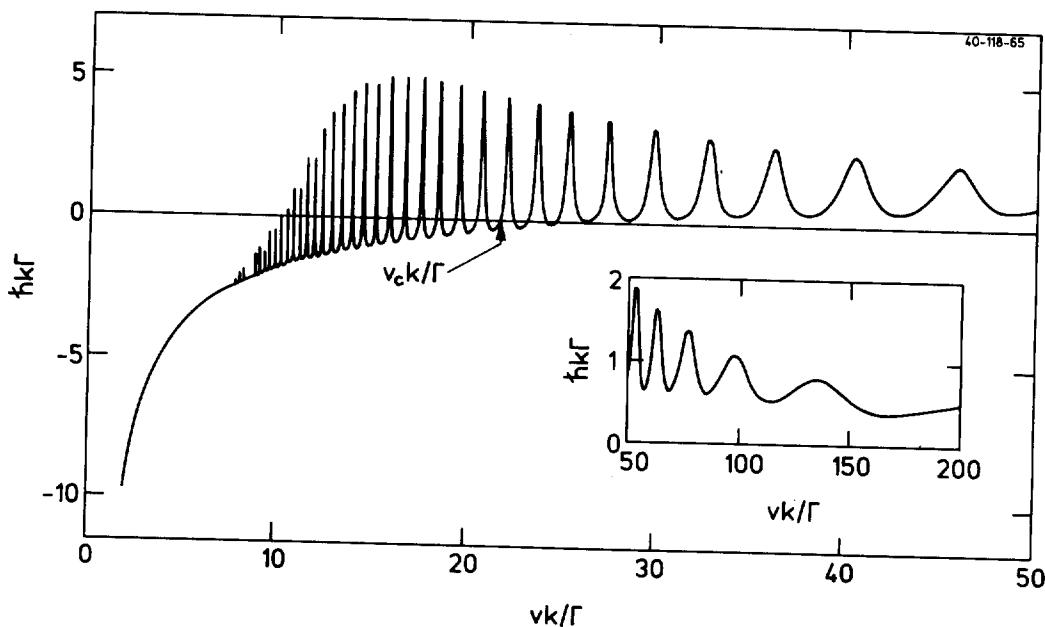
Nonadiabatic transitions become more important for decreasing minimum distance  $\hbar|\delta|$  between energy levels of type 1 and type 2 and for increasing rate of change of the dressed states  $|j;t\rangle$ , i.e., for increasing value of the ratio  $\omega_R(0)/|\delta|$ . In the examples shown in Figs. 2 and 3, the velocity  $v_c$  is  $\sim 20\Gamma/k$ .

When nonadiabatic transitions must be included, together with spontaneous events, the evaluation of the average force is normally based on the solution of the so-called Optical Bloch Equations which govern the time evolution of the density matrix connected with the atomic states  $|g\rangle$  and  $|e\rangle$ , and not with the dressed states [8]. However, also at these higher velocities, the force may be obtained within the dressed-atom picture [9], [10]. The populations and the off-diagonal elements of the density matrix in the basis of dressed-atom states are now coupled, and the interpretation of the average force becomes somewhat less straightforward than in the adiabatic case. We therefore only indicate the results. From Fig. 4 it is observed that as a function of velocity, the average force changes sign after the passage of a transition region in the vicinity of  $v_c$  with strong and rapid oscillations. These features result from interference between amplitudes for nonadiabatic transitions at successive field nodes. Beyond the critical velocity  $v_c$ , the force again becomes quite appreciable, provided the Rabi frequency exceeds the detuning, corresponding to saturation of the atomic transition. In fact, in the entire velocity range  $v_c < v < |\delta|/k$  it is stronger than the saturated scattering force, and it is given by the expression  $\frac{1}{2}\hbar k\Gamma|\delta|/(kv)$ , apart from modulations.

#### 4.3 Experiments

The change in direction of the force in a strong standing wave above the velocity  $v_c$  has been verified in two very recent experiments. The first measurement [11] was on a thermal beam of sodium atoms with an average velocity of  $\sim 500$  m/sec and a velocity spread of  $\sim \pm 200$  m/sec. The field was standing in the laboratory frame and virtually all of the particles had an initial velocity higher than  $v_c$ . The detuning was negative, and after penetration of the wave over a distance of 27 cm, the width of the distribution in longitudinal velocity had, in one case, diminished to 12 m/sec (FWHM). The laser power applied was 0.2 Watt.

In the second experiment one was directly concerned with the problems associated with laser cooling of a fast beam [12]. The laser field must now be standing in a frame moving with the average particle velocity. To obtain significant velocity changes in a single-pass experiment with an interaction time two orders of magnitude shorter than in Ref. 11, it was necessary to



*Fig. 4: Average force, in units of  $\hbar k/\Gamma$ , as a function of velocity, in units of  $\Gamma/k$ , at a Rabi frequency  $\omega_R(0)=1000\Gamma$  and a positive detuning  $\delta=200\Gamma$ . The critical velocity  $v_c$  is given by Eq. (62).*

increase the energy density in the laser wave by means of a resonator. In the neon beam applied, the longitudinal velocity spread was on the order of the critical velocity  $v_c$ , owing to the velocity compression upon acceleration to an energy of 1 keV. In Fig. 5 are shown the velocity distributions before and after interaction with the field, as well as the difference between the two. Since the detuning is now positive, low and high velocity particles are forced towards the centre and towards the tails, respectively. The changes observed are in qualitative agreement with estimates based on a transformation of the initial distribution according to the calculated velocity-dependent average force. The intention is to carry the experiment to somewhat higher field strengths for a constant value of the ratio  $\delta/\omega_R(0)$ . For a positive detuning, this will lead to longitudinal cooling of essentially the entire beam, as in the calculation presented in Fig. 3.

## 5. SUMMARY

For atoms or ions in a beam, we have discussed the force from a strong collinear laser beam which forms a standing wave in a reference frame moving with the average velocity of the particles. In an experiment, the standing wave in the moving frame is obtained as the superposition of two travelling waves which are propagating in opposite directions and which have different frequencies in the laboratory.

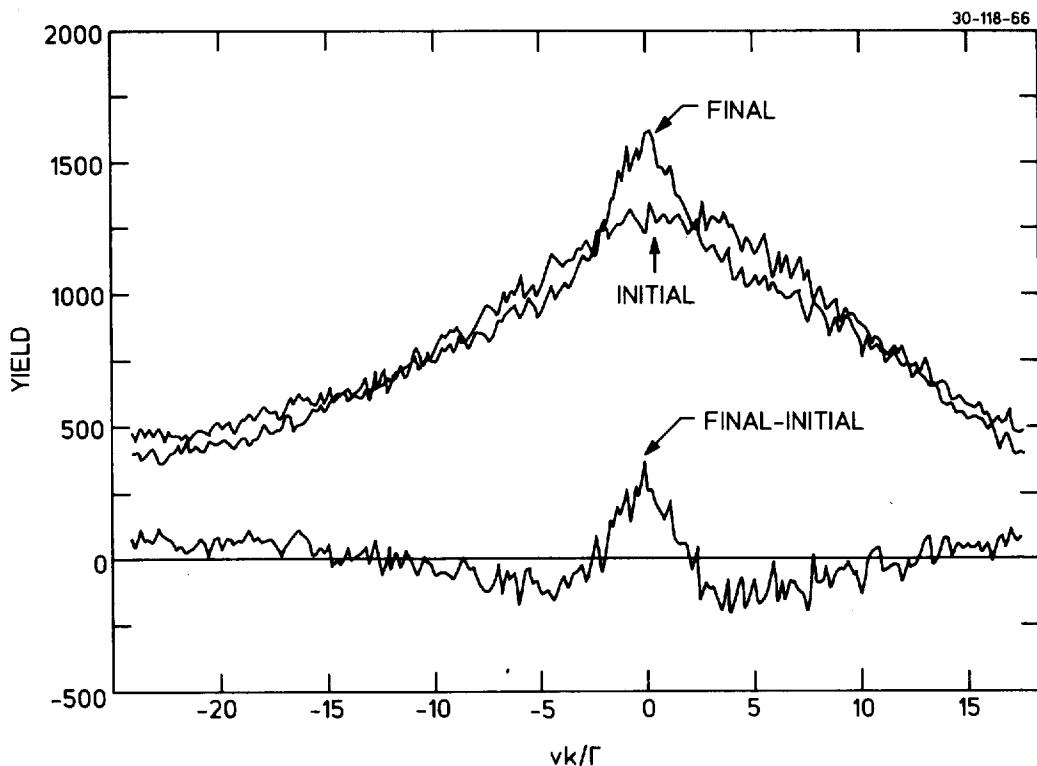


Fig. 5: Measured longitudinal velocity distribution for 1 keV neon atoms before and after interaction with a standing laser field during  $6.5 \mu\text{sec} = 330\Gamma^{-1}$ . The atomic states have angular momenta  $J_g=2$  and  $J_e=3$ , and the light is linearly polarized and has a positive detuning of  $60\Gamma$ . A closely equivalent two-level system has a Rabi frequency of  $\omega_R(0)=275\Gamma$ . The velocity  $v_c/2$  equals  $5.1 \Gamma/k$ , where  $\Gamma/k=5.3 \text{ m/sec}$ . The distributions are inferred from the Doppler shift of a transition induced by a probe laser, and the resolution is  $\sim \pm 2\Gamma/k$ . The yield refers to spontaneous emission following the excitation. Published with kind permission of the author of Ref. 12.

When the Rabi frequency  $\omega_R(0)$  becomes large compared to the detuning  $|\delta|$ ,  $\omega_R(0) > |\delta| > \Gamma$ , the laser intensity is sufficient to saturate the atomic transition under consideration. In these circumstances, particles with velocities relative to the standing wave in the range  $0 \leq v \leq |\delta|/k$  experience a force which is stronger than the maximum scattering force from a travelling wave. As an example, we may choose the detuning  $\delta$  such that the upper velocity limit  $|\delta|/k$  equals 1000 m/sec. In a typical case with a photon energy  $\hbar\omega=2 \text{ eV}$  and an atomic lifetime  $\Gamma^{-1}=10^{-8} \text{ sec}$ , saturation is then obtained when the two travelling waves composing the standing wave have intensities of the order of 1 Watt/mm<sup>2</sup>.

The average force changes sign with the detuning, and for given values of the detuning and the field strength, the force directions are opposite in

the two velocity regimes below and above a complicated transition region around the critical velocity  $v_c$ , given in Eq. (62). For red detuning, the force is retarding above  $v_c$  which is of the order of 100 m/sec in the example considered above, provided the saturation is not too strong. Apart from modulations, the force equals  $\frac{1}{2}M\Gamma|\delta|/v$  in the velocity region  $v_c \leq v \leq |\delta|/k$ , and the time required for the average force to bring an atom of mass M from one end of the velocity region to the other is therefore approximately equal to  $M|\delta|/(M\Gamma k^2)$ . In the example, we obtain a time of 0.3 msec for atomic mass 20. After the change of direction around  $v_c$ , the average force again becomes inversely proportional to velocity, and it increases linearly with the Rabi frequency when the ratio  $\delta/\omega_R(0)$  remains constant. As is observed from Fig. 2, the maximum force, attained at a velocity less than  $\Gamma/k$ , becomes quite strong for realistic field strengths.

In conclusion it would appear that for ions circulating in a storage ring, a strong standing laser field might be useful in a second cooling stage when a very accurate definition of longitudinal velocity is desired. First, electron cooling and/or stochastic cooling should reduce the velocity spread to a value of the order of a thousand meters per second. Then a millisecond of interaction with a laser field with red detuning could transform the velocity distribution into two peaks around the velocities  $\pm v_c$  with respect to the standing wave. Finally, a change of sign of the detuning together with a moderate increase in its absolute magnitude would bring the particles into the low velocity region with a strong dipole force, and a rapid cooling to essentially vanishing velocity spread would result.

#### ACKNOWLEDGMENTS

I gratefully acknowledge help and constructive criticism from J.U. Andersen, P. Jessen, K. Mølmer, and O. Poulsen.

\* \* \*

#### REFERENCES

- [1] For instance, J. Slater: Quantum Theory of Matter. McGraw-Hill (1975), Chapter 13.
- [2] Reproduced from lecture notes by P. Kristensen, Aarhus University, 1981.
- [3] A. Yariv: Quantum Electronics. John Wiley (1975), p. 90.
- [4] J. Dalibard and C. Cohen-Tannoudji, J.Opt.Soc.Am. B 2 (1985) 1707.
- [5] C. Cohen-Tannoudji, in: Frontiers in Laser Spectroscopy, Vol. 1 (Les Houches, Session XXVII, 1975) North-Holland (1977), ed. by R. Balian, S. Haroche, and S. Liberman.

- [6] A. Aspect, J. Dalibard, A. Heidmann, C. Salomon, and C. Cohen-Tannoudji, Phys.Rev.Lett. 57 (1986) 1688.
- [7] E. Bonderup and K. Mølmer, J.Opt.Soc.Am. B 6 (1989) 2125.
- [8] For instance, J. Dalibard, Thèse, Université de Paris, 1986.
- [9] A.P. Kazantsev, V.S. Smirnov, G.I. Surdutovich, D.O. Chudesnikov, and V.P. Yakovlev, J.Opt.Soc.Am. B 2 (1985) 1731.
- [10] K. Mølmer, to be published.
- [11] M. Prentiss and A. Cable, Phys.Rev.Lett. 62 (1989) 1354.
- [12] P. Jessen, Thesis, Aarhus University, 1989.



## SMALL RING LATTICE PROBLEMS

*E.J.N. Wilson*

CERN, Geneva, Switzerland

### ABSTRACT

This is a review of the special problems in particle dynamics lattice design, and magnet construction and measurement which should be considered in the design of small synchrotrons and storage rings. The CERN Antiproton Accumulator is used as an illustration.

### I. INTRODUCTION

This report sets out the practical problems which face the designer of a small storage ring : the CERN Antiproton Accumulator. The designers of other small rings for electrons and high intensity proton synchrotrons will encounter similar problems. Their solutions may differ according to the application but the example will be instructive

Synchrotrons built for the highest energy electrons and protons are built of hundreds and even thousands of small aperture magnets several meters in length. There are usually several hundred regular periods and enough space in the six or eight insertions to string together a series of purpose built sections for injection, ejection, dispersion correction and low beta matching. In a small ring the number of magnets and periods is severely restricted and one must often exercise considerable ingenuity to arrive at a design which satisfies all the requirements necessary to arrive at the desired performance.

The ends of the long high - energy - machine magnets constitute only a small fraction of the integrated field seen by a circulating beam and the fact that the fields in the ends are three - dimensional can usually be ignored. Indeed many of the computer programs developed for the design of large rings contain approximations which treat the magnets as pure, two - dimensional fields and ignore the small curvature of the central trajectory through the end field. Such approximations must be reviewed critically by the designer of a small ring where the magnet aperture can become comparable to the length and where much of the magnet's effect comes from a fringe field in which the particle is deflected with a radius of curvature comparable to the magnet's length.

In this report we first discuss how one may satisfy a number of design constraints in a small ring like the Antiproton Accumulator, how measurement and correction of the end field can be combined with Q measurements on the finished machine to correct effects introduced by short magnets of large aperture, and finally we consider a class of end effects which are not normally included in even the most rigorous lattice programs.

## 2. THE EXAMPLE – AN ANTIPROTON ACCUMULATOR

In the photograph taken though a "fish - eye" lens, we see the large bending dipoles and focusing quadrupoles of the Antiproton Accumulator which has a mean radius of 25 m. The objects to be seen in Figure 1, wrapped in shiny aluminium bake - out jackets and installed in between some of the magnets are pick - ups and kickers which are mainly used for stochastic cooling and do not concern us for the purposes of this talk.

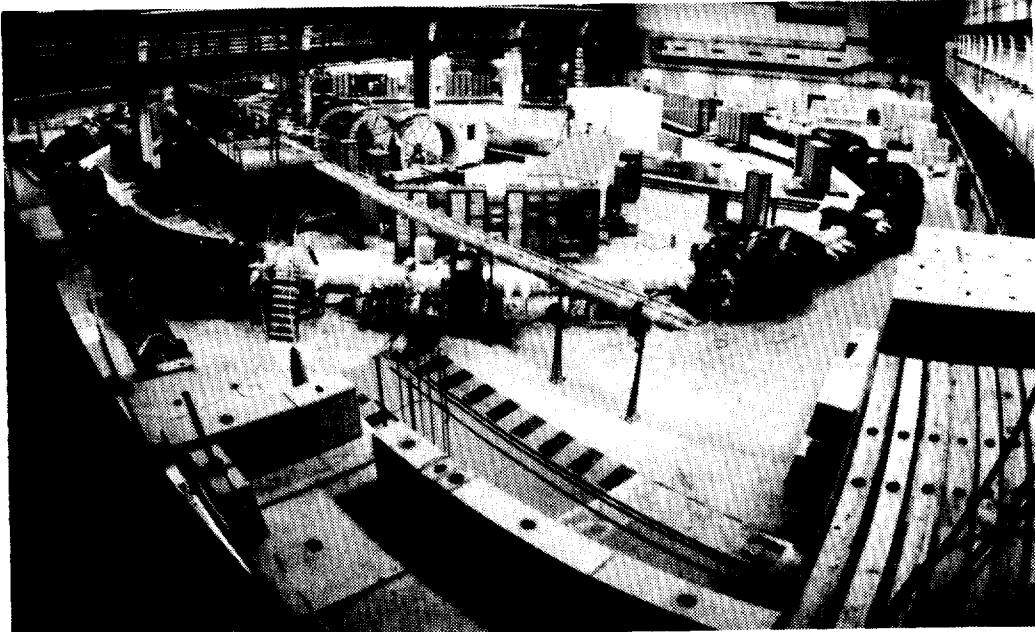


Figure 1: A Fish - eye view of the Antiproton Accumulator

In the bottom right hand corner of the plan view shown in Figure 2, we see the target where antiprotons are produced and a short transport line which brings them to a point, at 12 o'clock on the "dial" of the ring, where they are bent by a septum magnet to join the circulating beam. A little further around, at about 2 o'clock on the dial, is a kicker magnet which inflects them onto the injection orbit. At 10 o'clock is an ejection kicker which uses the same septum to extract the beam which has been accumulated and stored.

One of the constraints on the design is that these kickers should be located at a particular horizontal betatron phase advance from the septum. This should be rather close to an odd multiple of 90 degrees. We shall see that such injection and ejection details often have quite strong influence on the lattice design and should not be left until after the major parameters have been frozen.

One of the problems which arises naturally when the ring is small is that there is not enough room for all the components and this leads to a great shortage of space between the magnets. In Figure 3 we see a typical gap between quadrupole and bending magnet. We must be careful to ensure that their fields do not interfere or, if they do, that we know from measurements with the neighbouring magnet in place, how the effective length of the magnet and quadrupole are affected by the other's presence. We can see in Figure 4 some early measurements of quadrupole gradient along a line parallel to the axis which were made in the 1950's [1], that a steel plate acting as a mirror to simulate the presence of another magnet has a significant effect on the effective length and central field gradient.

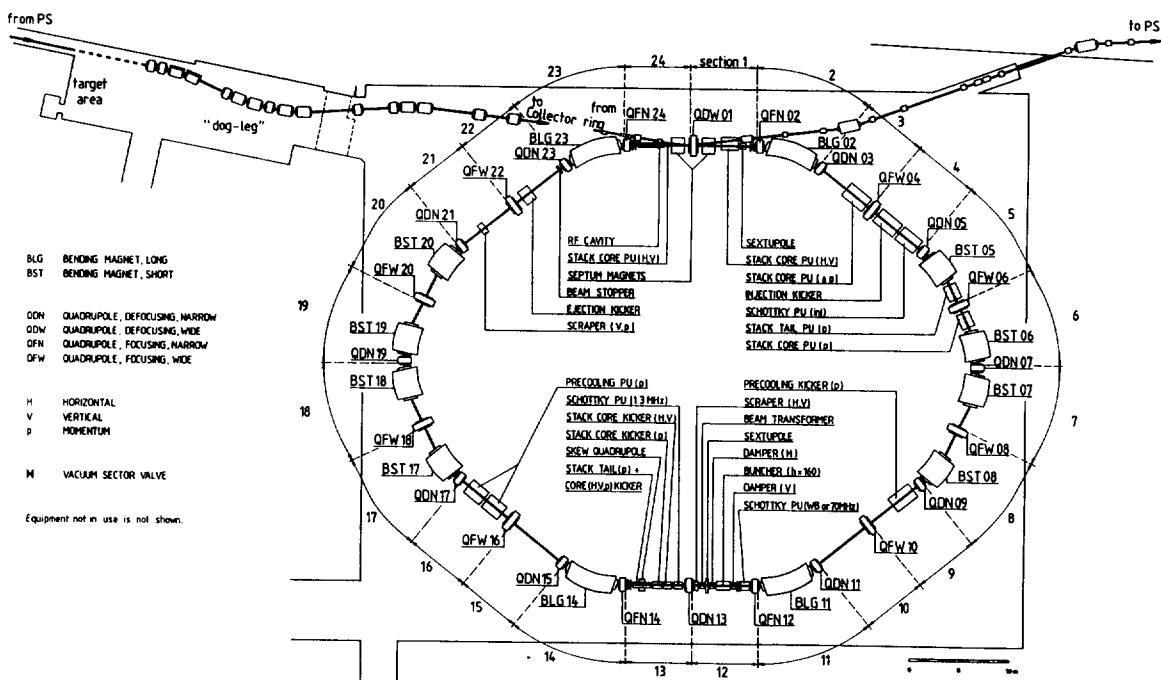


Figure 2: General layout of the Antiproton Accumulator

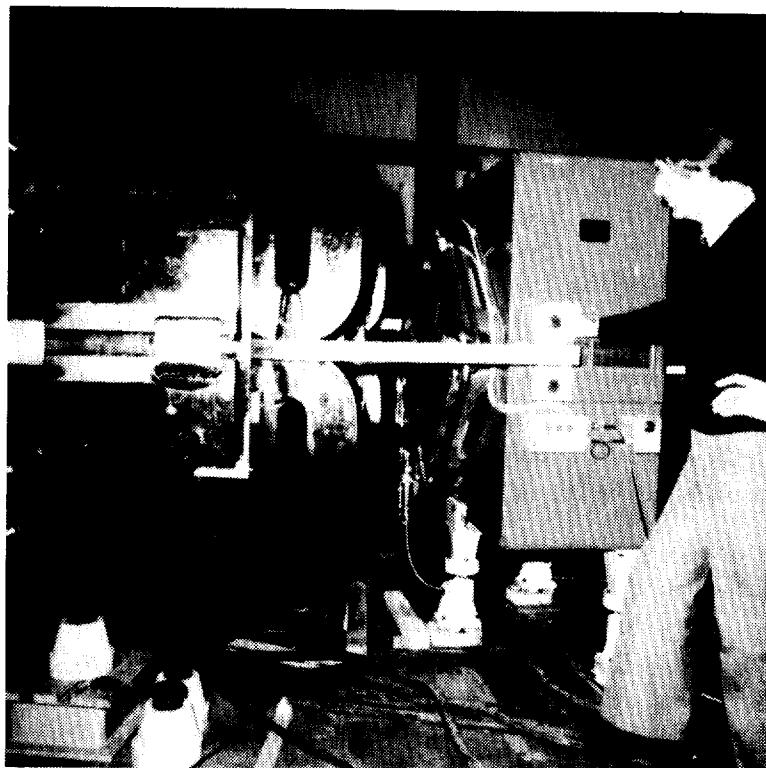


Figure 3: Gap between a dipole and a quadrupole in the AA.

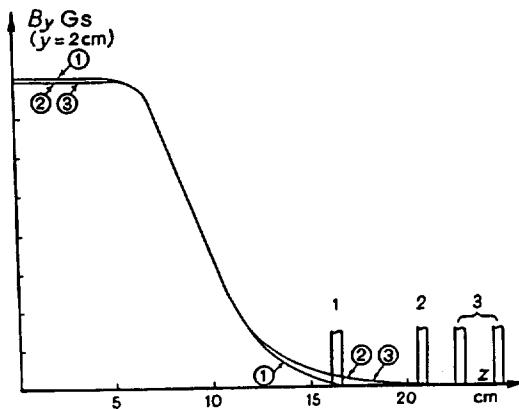


Figure 4: End field shape for a quadrupole with a steel mirror plate

In arriving at a suitable lattice design for this ring we first chose the energy to be 3.5 GeV since this is the energy at which the antiprotons are most abundant and we tried to make the radius 25m, so that its circumference is exactly one quarter of that of the PS. The bunches of protons which produce the antiprotons originate in the PS and it is this machine which will later have to re-accelerate the antiprotons.

It is natural to first try to design around a FODO pattern of quadrupole lenses since this is the simplest from many points of view. Of course, designers of electron rings for synchrotron radiation usually choose another kind of lattice, but the problems we shall discuss tend to be common to any lattice configuration.

This particular ring has a specific requirement imposed by the cooling system which demands that the spread of revolution frequencies, which stems from the momentum spread and which is determined by

$$\eta = \frac{1}{\gamma^2} - \frac{1}{\gamma_{tr}^2} = \frac{p}{f} \frac{df}{dp}, \quad (1)$$

is within rather close limits. This fixes  $\gamma_{tr}$ . Other machine designers will encounter different constraints on this parameter. In high intensity boosters it is often thought a good idea not to include  $\gamma_{tr}$  in the energy range of the synchrotron. In electron rings, although one is far above transition, the momentum compaction which is linked to the energy damping time places a similar constraint on the designer.

The second term in the expression for eta is equal to the momentum compaction function divided by bending radius. It is determined by the change in circumference with momentum about the equilibrium momentum and is just the average value of the dispersion function D(s) around the ring:

$$\frac{1}{\gamma_{tr}^2} = \frac{\alpha_0}{R} = \frac{1}{2\pi R} \int \frac{D(s)}{\rho(s)} ds \quad (2)$$

We shall see that this parameter is controlled by the choice of Q value and indeed, for a proton machine transition, is roughly equal to Q. The Q value is in turn closely tied to one quarter of the number of FODO periods since one tries to choose a betatron phase advance of 90 degree or perhaps 60 degrees but rarely outside this range. Since the desired  $\gamma_{tr}$  for the AA was 2.4, the Q was chosen to be 2.3 and therefore the number of periods had to be approximately 10.

If the number of periods is too large the space will be chopped up into too many small pieces; there will be too many components and too many wasteful ends to them. Furthermore, the AA ring has a natural symmetry of four and we are left with only 8, 12 and 16 as possible numbers of periods. When we come to examine the relative merits of the few remaining options we find that  $N = 8$  would suggest a smaller Q value and hence beta, which is roughly  $R/Q$ , becomes uncomfortably large. When we considered the acceptance necessary to collect enough antiprotons the apertures of the magnets are large even with  $N = 12$ . The larger beta of  $N=8$  would make matters worse. On the other hand 16 periods would result in too many small components.

Once the period number is fixed, the position of all the quadrupoles is determined and the Q value is fixed within narrow limits. The length of the quadrupoles is also determined by the peak field one may allow on their pole pieces which must inscribe an ellipse or rectangle sufficient to accommodate the beam. The lattice functions and the required emittance and momentum spread ( 100 pi mm. mrad and  $\pm 3\%$  in the case of the AA ) are now pretty well defined and it is merely a matter of juggling to make the exact numbers from a lattice program consistent.

We still have to decide on the layout of bending magnets within the ring and the positions of injection, ejection and ( in our example) the cooling systems. The kickers must be 90 degrees (or some odd multiple of this) from the septum and this determines which of the half periods they lie in. The AA ring has a rather special requirement that there will be two "beams" circulating which differ only in momentum as shown in Figure 5, one is the "stack" were antiprotons are accumulated and the other the injected beam which must be physically separated from the stack at the injection kicker. This magnet, which encloses only the injected beam has a movable ferrite shutter which closes the magnet aperture on the stack side to prevent the stack from being disturbed by the firing of the kicker. The two beams must be separated by enough space for this shutter.

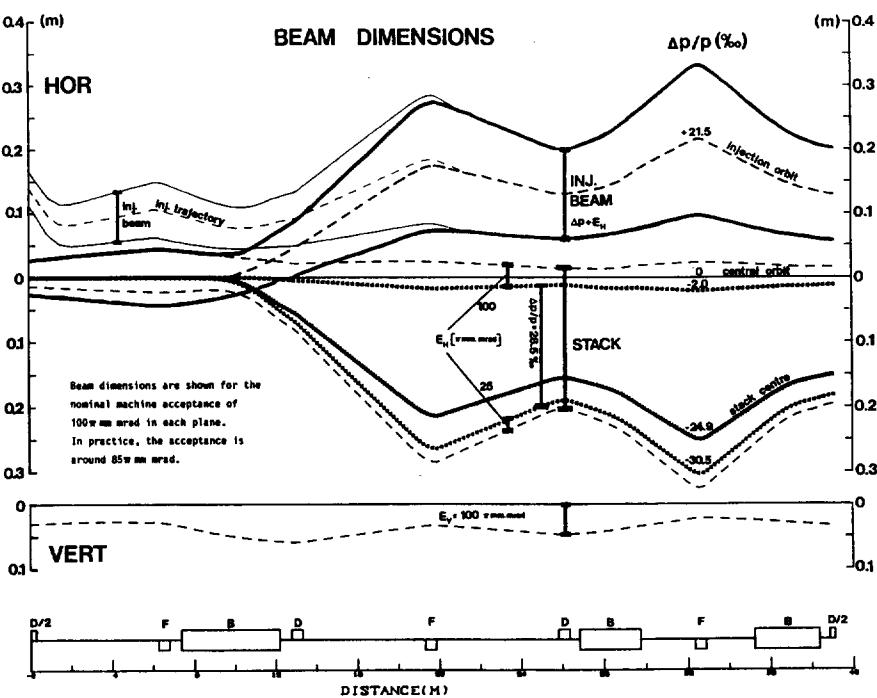


Figure 5: Beam dimensions for one quadrant of the AA.

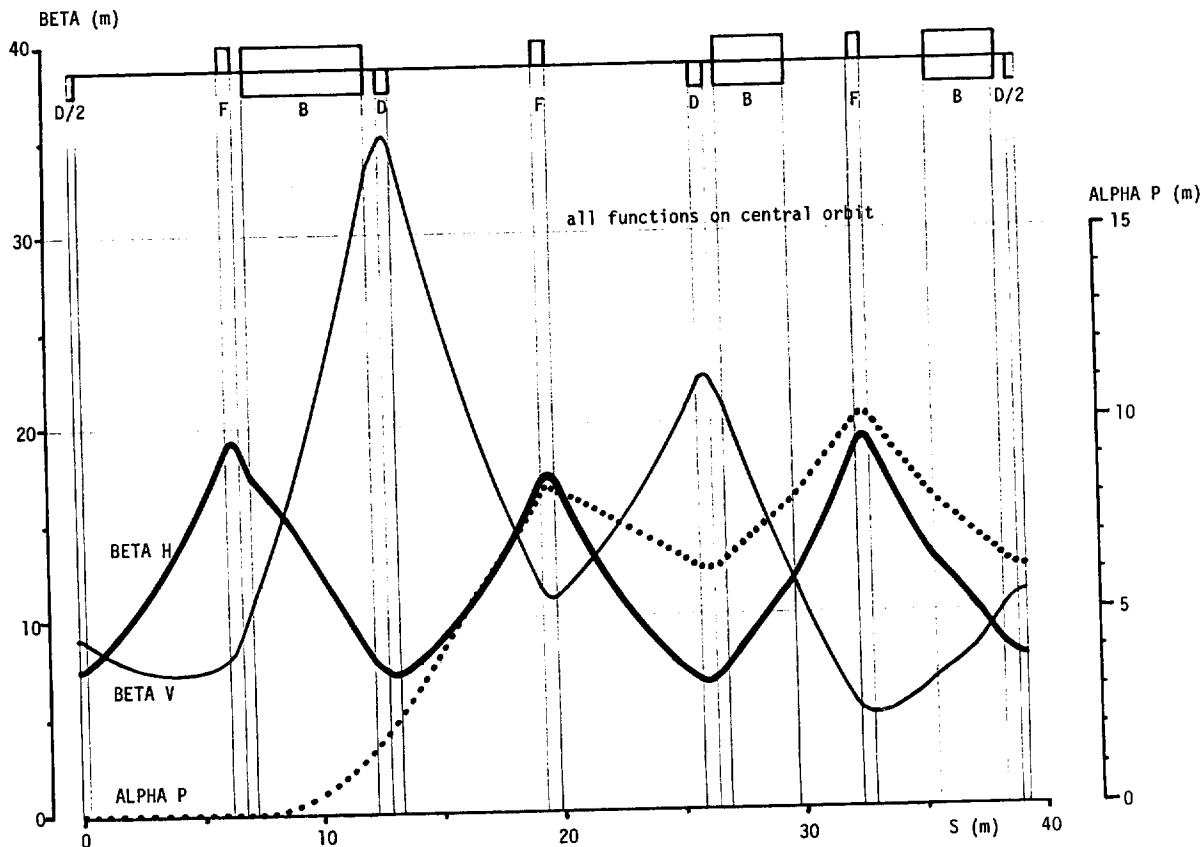


Figure 6: Beta functions and dispersion for one quadrant of the AA.

Elsewhere both beams must pass through a narrow cooling pick - up at 12 and 6 o'clock and there must be a steep rise in the dispersion function between the septum and kicker as can be seen in Figure 6.

This rapid change in dispersion is achieved by choosing to make the lattice symmetric about a D quadrupole at 6 o'clock. We also make use of the horizontal defocusing properties of the end field of a dipole whose faces are tilted horizontally to make an angle with the beam. Both these features help to raise the dispersion as fast as possible immediately after the first dipole encountered by the incoming beam and then cause it to roll over into a high sustained value in the rest of the superperiod [2]. The result is to be seen in Figure 6. I explain this rather peculiar feature only to complete the AA story. Other machines will no doubt have their own special reasons for choosing a particular symmetry or a particular kind of dipole end design.

Once the lattice designer has reached this advanced stage in fixing the design it is high time that he checks with the designers of all the other components of the ring that there is enough room for them to be installed. Some of these will be best placed where the dispersion is zero. In an electron ring this is true of RF cavities and in our example it is the betatron cooling kickers that have this preference. There may be other components, momentum cooling pick - ups in the AA but momentum scrapers in synchrotrons, which need to be where the dispersion is greatest. All these must have enough free length for their function and still leave room for diagnostic pick - ups.

### 3. MAGNET DESIGN, MEASUREMENT AND CORRECTION

The magnets of small rings tend to be short when measured in numbers of gaps or aperture radii. The end fields of such magnets must be carefully considered because no longer will the shape of the poles in the body of the magnet alone fix the field purity. This is particularly so in the case of the AA magnets which are not only short but have huge apertures in order to accept enough beam Figure 7. Each of the dipoles of a small ring with only a few periods will bend through an appreciable fraction of a circle. In the case of the AA the average entry angle is more than 20 degrees.

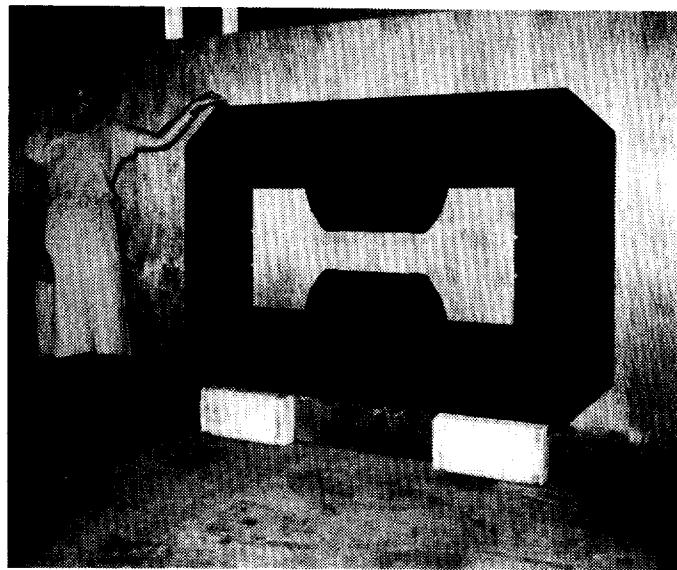


Figure 7: One lamination of an AA dipole magnet

We must decide whether we should make the magnet curved so that the end faces are normal to the entering and exiting beam or whether the magnets will be just simply stacked from parallel laminations so that their ends are parallel and present, in this case, an angle of 20 degrees to the beam. The focusing properties of the two kinds of magnet are quite different. The transport matrices in the horizontal and vertical planes for a magnet which is curved to ensure that the beam enters and leaves normally are given by :

$$M_H = \begin{pmatrix} \cos\theta & \rho\sin\theta & \rho(1-\cos\theta) \\ -\frac{\sin\theta}{\rho} & \cos\theta & \sin\theta \\ 0 & 0 & 1 \end{pmatrix} \quad (3)$$

$$M_V = \begin{pmatrix} 1 & \rho\theta & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (4)$$

We can see that the (2,1) element of the horizontal matrix, which expresses the focusing power, is finite. In contrast, in the vertical plane there is no focusing action. On the other hand you can change this by tilting the end faces by angle  $\epsilon$  as shown in Figure 8. To first approximation, the effect of this tilt may be expressed by thin - lens matrices at each end :

$$M_H = \begin{pmatrix} 1 & 0 & 0 \\ \frac{\tan\epsilon}{\rho} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (5)$$

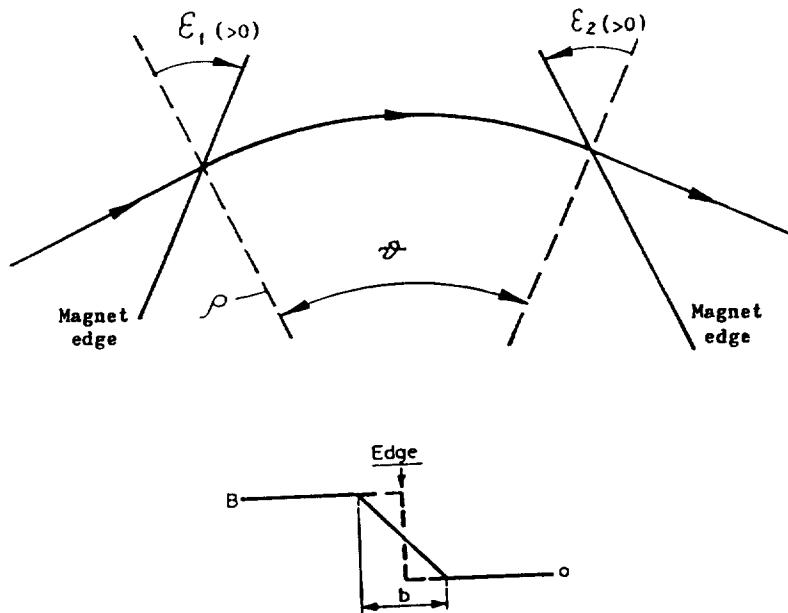


Figure 8: Geometry of a bending magnet which is not a sector.

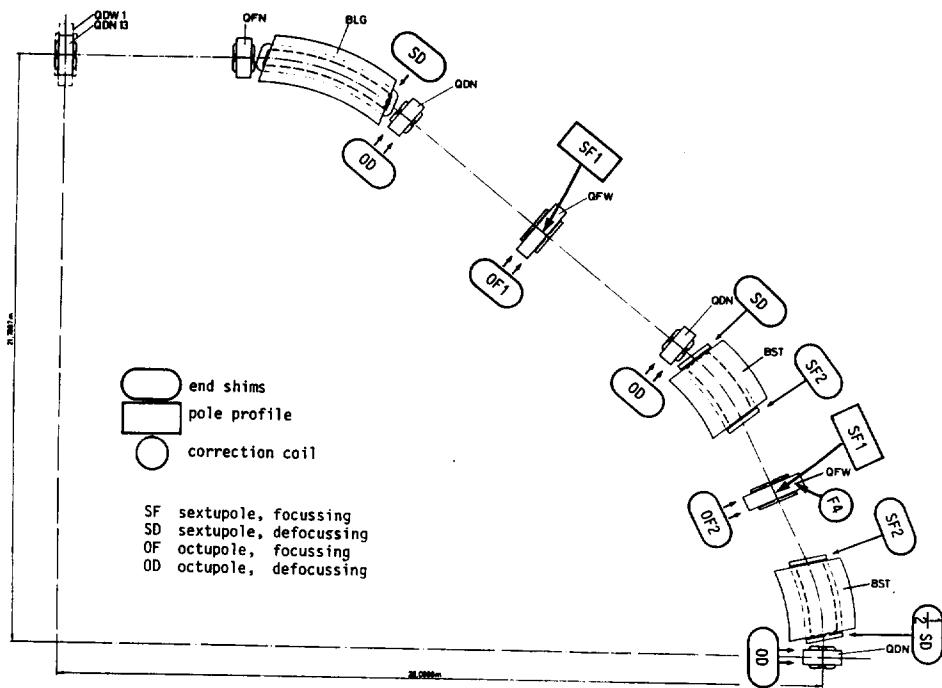
$$M_V = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{\rho} \left( \frac{b}{6\rho \cos \epsilon} - \tan \epsilon \right) & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (6)$$

For the special case of the parallel ended magnet  $\epsilon = \theta/2$  the (2,1) elements become  $\frac{\pm \tan(\theta/2)}{\rho}$ . If you multiply the matrices together you will find that the focusing power in the horizontal plane is cancelled by these thin lenses and instead, the magnet focuses vertically. Either way one cannot avoid considerable beating of the beta function in one or the other of the transverse phase planes. We see from Figure 6 that we chose to let this happen in the vertical plane of the AA, where apertures were not so huge, rather than adopt complicated special quadrupole arrangements to match it out.

The optical properties are becoming quite precise at this stage and one should recalculate Figure 5 to check aperture dimensions and design the vacuum chamber. We can see that some of the magnets have to house a beam 60 cm wide. This is a very extreme case which illustrates admirably some of the effects which one must be aware of in small ring even if, when calculated for smaller apertures, they prove not to be as important as in the AA.

#### 4. MULTipoles IN THE MAGNET DESIGN

One of the difficulties we encounter in an analysis of beam dynamics for such a machine is that there are non-linear terms in the focusing which have to be corrected. These are important for they modify the chromaticity or introduce a variation with momentum in the dispersion of the machine. One must compensate the chromaticity rather precisely in a storage ring and this may be done in a large ring with sets of sextupole magnets. In a small ring there is often too little space for special correction magnets but multipole correction fields may be incorporated in the magnet design. This may



*Figure 9:* Multipole corrections applied to one quadrant of the AA

either be done by shaping the pole pieces or by shimming their ends with lumps of steel bolted onto the pole pieces. For example we may design the main quadrupoles with poles having a sextupole asymmetry which causes a variation of gradient so that it is no longer the same over the whole width but has a slope so that it is stronger on the outside of the machine and therefore compensates the variation of  $Q$  with momentum : the chromaticity. The corrections labelled SF1 in the plan of one quadrant of the AA, Figure 9 are of this type. The left right symmetry of the quadrupole is broken to do this. At the same time any curvature in the variation of  $Q$  as a function of momentum may be corrected.

We can see in Figure 10 how the integrated gradient of the quadrupole, designed with a magnet field mesh program, has a linear variation with horizontal displacement plus a quadratic term to match the chromatic curvature. The linear term is sextupolar and the curvature is octupolar (labelled OF, and OD in Figure 9). The same Figure shows correction SF2 and SD which are sextupole terms applied by end shims to the dipoles.

In the AA we designed all these corrections in at the beginning but also provided the means to make adjustments to the radial field variation afterwards with packs of washers mounted on studs protruding from the pole ends. In Figure 11 we see a wide quadrupole magnet being measured and the studs with their washers are clearly seen. The long coil integrates the gradient along a paraxial line. By controlling the number of washers on each stud we can shape the integral of the gradient as a function of horizontal position without using any power or taking up any of the circumference.

In order to calibrate the effect of washers in the 16 different stud positions seen in Figure 12 a pack of six washers was placed on each stud in turn and the change in integrated gradient compared with an unshimmed magnet measurement [3].

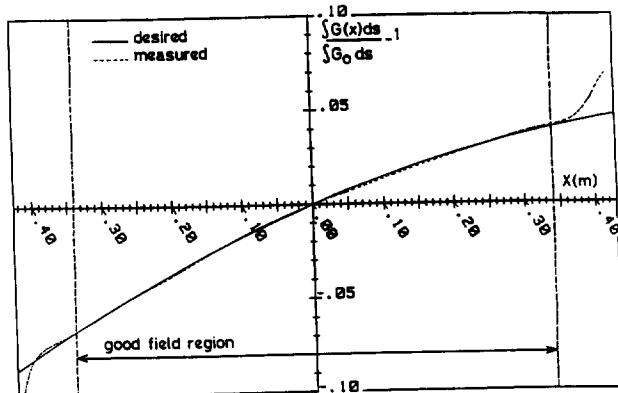


Figure 10: Comparison between required and measured quadrupole gradient

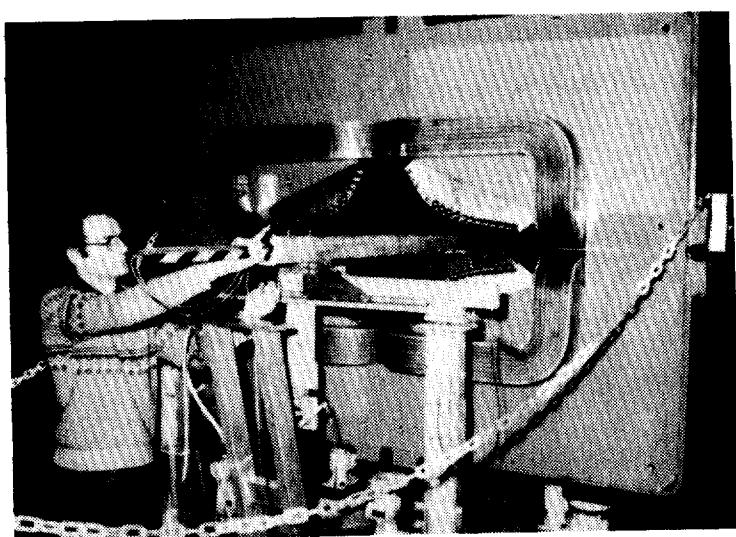


Figure 11: Measurements being made on an AA wide quadrupole

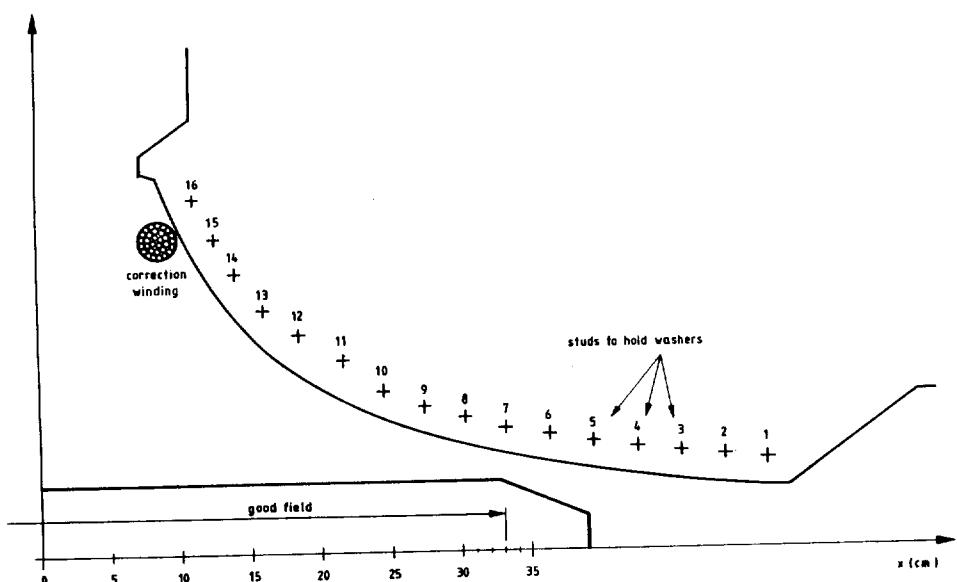


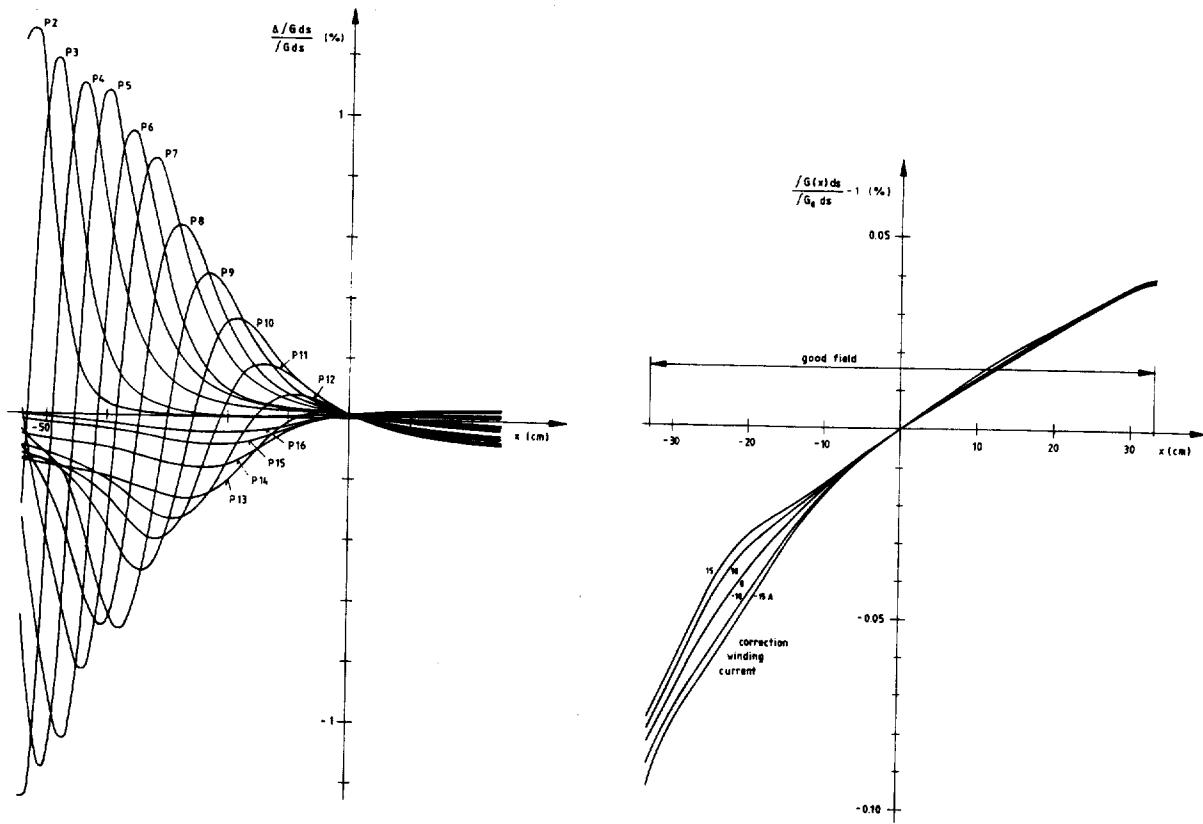
Figure 12: Stud positions and correcting winding on an AA wide quadrupole

The results of this calibration are shown Figure 13 where each curve shows the change in gradient due to one pack of six on each of the washer positions. A program was written to combine combinations of washers to produce any desired change in field gradient.

It proved perfectly practical to shim field shapes empirically on the basis of Q measurements with these washers but the process took several hours. To modify field shapes on - line and with the beam circulating we installed single - turn correction windings which can be seen in Figure 12 mounted on the pole. Their effect was also calibrated by field measurement on a prototype and the results for a range of currents can be seen in Figure 14. Of course such correction windings lead to power supply complications in a pulsed machine

Now we come to consider how successful all this shimming was in correcting the Q variation with momentum. We see in Figure 15 the residual variation in Q when the AA was first switched on [4]. Although every care had been taken there is still a Q variation in the horizontal plane which is larger than the space between the one - third and one - quarter integer resonances.

Adjustments to the end shims were calculated and applied to reduce this by almost an order of magnitude (Figure 16) and when the same points are plotted on a  $Q_h$ ,  $Q_v$  diagram (Figure 17) we find that all sum resonances up to 11th order are avoided. The correction was applied in two iterations.



*Figure 13:* The effect of a six - pack of washers on each stud position

*Figure 14:* Effect of current in the correction winding of an AA quadrupole

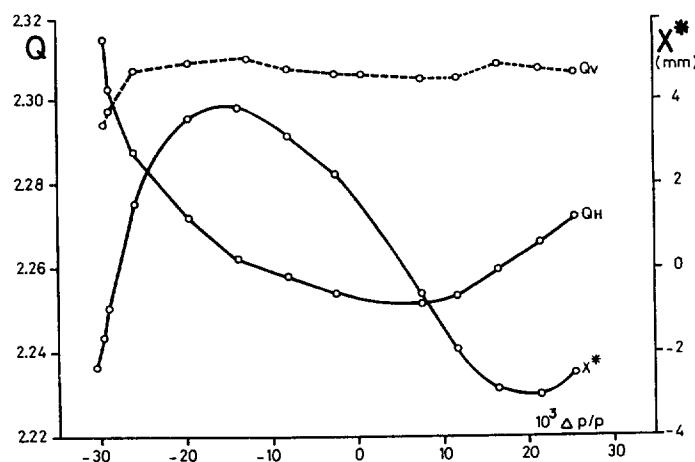


Figure 15: Variation of  $Q$  and mid-straight dispersion before correction

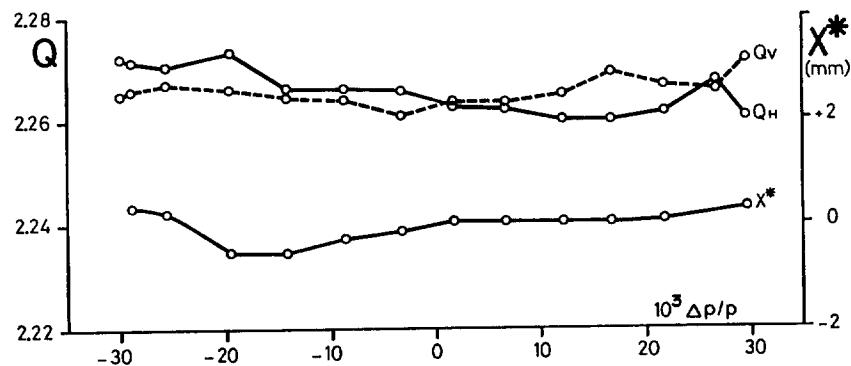


Figure 16: Variation of  $Q$  and mid-straight dispersion after correction

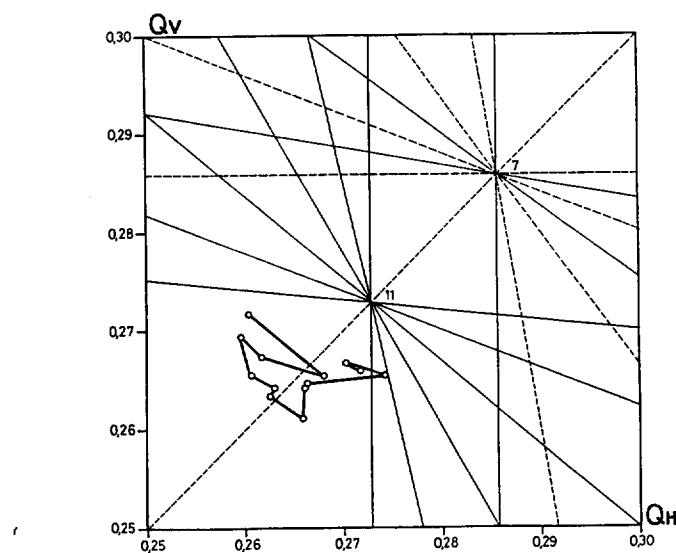


Figure 17: The working line of the AA after correction

## 5. TWO-DIMENSIONAL FIELD INTEGRALS

And now we should turn to some of the theory behind the description of magnets and how this must be modified when the magnets are only a few gap dimensions long. If a magnet's ends are identical, the field in the plane of symmetry mid way between the ends will be "two dimensional". Symmetry dictates that there can be no axial component,  $B_r$ . Laplace's equation for the scalar magnetic potential,  $\phi$  :

$$\frac{\partial^2 \phi(x,y,z)}{\partial x^2} + \frac{\partial^2 \phi(x,y,z)}{\partial y^2} + \frac{\partial^2 \phi(x,y,z)}{\partial z^2} = 0 \quad (7)$$

reduces to just the outer two terms [5]. The two-dimensional equation has the very attractive property that the solution is a harmonic series

$$\phi = \phi_n r^n \sin(n\theta) \quad (8)$$

Each term in this series corresponds to a magnet with a different number of poles. The index  $n$  of the  $n$ 'th term is just half the number of poles, pure multipole would have which produced the  $n$ 'th term in the series. It is also associated with the order of the non-linear resonance which this multipole can produce. For example if  $n$  is 2 we obtain the field produced by a quadrupole. The potential must reverse as we describe a circle of constant radius to have two positive and two negative excursions corresponding to the four poles. Of course, in this case, the motion is linear and any resonant condition has order 2.

This is a useful simplification when we consider the dynamics of a beam passing through a long magnet where most of the field has this two-dimensional property. We can even preserve this simplification when we include the end fields of a short magnet provided the deflection in the magnet is small, as is the case in high energy synchrotrons with large radius. We can then approximate the particles equilibrium orbit to a straight line paraxial to the centre line of the magnet. The integral of the potential along such a line, in other words, the average potential, has the same two-dimensional properties. This may be proved by integrating the three terms of Laplace equation along such a line. The middle term becomes:

$$\int_{-y_0}^{y_0} \frac{\partial^2 \phi(x,y,z)}{\partial y^2} dy = \left[ \frac{\partial \phi(x,y,z)}{\partial y} \right]_{-y_0}^{y_0} \quad (9)$$

Clearly if  $y_0$  is taken arbitrarily far away from the ends the expression will be zero at each limit. Each of the other two terms becomes a simple average of the transverse second derivatives of potential. For example:

$$\int_{-y_0}^{y_0} \frac{\partial^2 \phi(x,y,z)}{\partial z^2} dy = \frac{\partial^2}{\partial z^2} \left[ \frac{1}{2y_0} \int_{-y_0}^{y_0} \phi(x,y,z) dy \right] = \frac{\partial^2 \bar{\phi}}{\partial z^2} \quad (10)$$

and the two-dimensional Laplace Equation as well as its trigonometrical solutions apply exactly to the averages of the potential along paraxial lines. The average transverse fields seen by the particles are just the derivatives of the average potential and are the familiar trigonometric series with  $n$  being half the number of poles in the multipole associated with each term.

It is of course the vector potential,  $\mathbf{A}$ , which is more commonly used in the study of beam dynamics since it describes the effect of magnetic field in the Hamiltonian of the motion. In the two-dimensional

sional case  $A_x$  and  $A_z$  are zero and the axial component obeys Laplace's equation in two - dimensions producing a harmonic series solution for the field expansion.

The two - dimensional solution of  $A_y$  may also be expressed in Cartesian coordinates as a polynomial:

$$A_y = \sum_{n=1}^{\infty} A_n f_n(x, z) \quad (11)$$

where :

$$f_n(x, z) = (x + iz)^n . \quad (12)$$

For a quadrupole this is:

$$f_2(x, z) = (x^2 - z^2) + i(2xz) . \quad (13)$$

We can derive the transverse field components:

$$B_z = \frac{\partial A_y}{\partial x} = \sum_{n=1}^{\infty} n A_n x^{(n-1)} \quad (14)$$

when  $z = 0$ . And we can deduce that the imaginary terms in the polynomial correspond to the skew orientation of multipoles while the real term in our example could represent a normal lattice quadrupole. A real cubic term would be a sextupole with a vertical plane of symmetry while the imaginary cubic terms stem from a skew sextupole, one that is, which has been rotated about its axis by  $1/12$  of a revolution.

We find, therefore that there is a simple association between the terms in the Cartesian expansion and the order of multipole, just as there was in the polar case. We can even go further to associate each term with a term in the Taylor expansion of the field about the axis of the magnet:

$$B_z = \sum_{n=1}^{\infty} \frac{1}{(n-1)!} \frac{dB_z^{(n-1)}}{dx^{(n-1)}} z x^{n-1} \quad (15)$$

and thus a quadrupole produces a gradient, a sextupole a second derivative and higher multipoles produce successively higher derivatives.

All of this applies equally to particles whose path may be approximated by paraxial rays and to measurements made with long paraxial search coils which extend clear of the magnetic field at each end. One can often identify the multipole content of the magnet simply by inspecting the transverse dependence of the field measured by such a coil. A quadratic variation indicates a sextupole and a cubic points to an octupole error term.

Just to complete this two - dimensional treatment, which I repeat is an approximation valid only for paraxial trajectories, let us recall the simple Hamiltonian:

$$H = \frac{p_x^2}{2} + \frac{p_z^2}{2} - \left(\frac{e}{p}\right) A_s . \quad (16)$$

The vector potential term can, in the paraxial approximation , be written:

$$A_s \approx \sum_{n=1}^{\infty} A_n f_n(x, z) \quad (17)$$

and we can conveniently associate each multipole with a term in the Hamiltonian. So for our example of a quadrupole we obtain for the motion in the x direction:

$$H = \frac{p_x^2}{2} - \frac{kx^2}{2} \quad (18)$$

where k is the normalized gradient  $B'/(B\rho)$ . Application of Hamilton's equations leads rapidly to the familiar Hill's equation of motion. This is using a sledgehammer to crack a nut but when we come to introduce higher order terms which are non-linear the power of the Hamiltonian produces an overall simplification.

## 6. CURVILINEAR COORDINATES

The above treatment is all well and good for large synchrotrons but breaks down seriously if the dipoles of the ring bend particles significantly from a paraxial path. This is much more likely to be the case in a small ring with a few focusing periods. In such a case we may consider a model which I will refer to as the "sliced loaf" model in which the end field is broken up into elementary slices each of which may be represented as one of the multipole shaped fields of the two-dimensional model. We would expect such a model to be capable of telling us, for instance, how bad the effect of correcting an error in the field due to the pole profile by putting shims on the end of the magnet. The two fields would not quite compensate for a particle that moved its position with respect to the magnet axis as it entered the fringe field.

To make this model work we must be very careful with the curvature of the trajectory. Bengtsson [6] has recently made an elegant analysis of this using the metric tensor which defines the relation between differential changes in the curvilinear system to the rectilinear Cartesian system:

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -(1+hx)^2 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad (19)$$

where h is the curvature  $1/\rho$ . Using the four-vector coordinates of special relativity he arrives at a differential equation of motion:

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{\nu\lambda} \frac{dx^\nu}{d\tau} \frac{dx^\lambda}{d\tau} = \frac{e}{m_0 c^2} F^\mu_\nu \frac{dx^\nu}{d\tau} \quad (20)$$

where  $F^\mu_\nu$  is the electromagnetic field tensor. This is really just an up-market version of:

$$\frac{d^2x}{dt^2} = \frac{e}{m_0} \mathbf{v} \times \mathbf{B} \quad (21)$$

and we can see that the middle extra term contains all the information on the transformation to curvilinear coordinates. He derives explicit equations of motion from this expressed in terms of the two-dimensional harmonic field coefficients. Later he derives a Hamiltonian in terms of the normalized quadrupole and sextupole strengths, k and m, and including the effect of dispersion, D, on a particle with a momentum defect  $\delta = dp/p$  :

$$\begin{aligned}
 H = & \frac{1}{2} \{ h^2 - k - \delta[(m + 2hk)D + h^2 - k] \} x^2 \\
 & + \frac{1}{2} \{ k + \delta[(m + hk)D - k + h'D'] \} z^2 \\
 & - \frac{1}{6} (m + 2hk)x^3 + \frac{1}{2}(m + hk)xz^2 \\
 & + \frac{1}{2}(1 + \delta h D)p_x^2 + \frac{1}{2}(1 + \delta h D)p_z^2 \\
 & + \delta h D'xp_x + \frac{1}{2}h'z^2p_x + \frac{1}{2}hxp_x^2 + \frac{1}{2}hxp_z^2
 \end{aligned} \tag{22}$$

Inspecting this Hamiltonian we first should note the appearance of the focusing strength,  $k$ , alone in the first two lines. This is just like the simple two - dimensional Hamiltonian from which one may derive linear betatron motion. We can see that the focusing will be modified by the first term  $h^2$ . This term will be familiar to those who remember combined function machines. The square brackets in the first two lines contain terms which affect the off - momentum behaviour. The simple focusing terms  $k$  and  $h^2$  appear again here to express chromaticity. The sextupole strength,  $m$ , also appears as a coefficient of  $\delta D$  and this must describe the effect of sextupoles in modifying the chromaticity. Terms which contain the product,  $hk$ , must be peculiar to combined function lattices where curvature and focusing occur in the same element but we should not ignore the product  $h'D'$  in the second term which tells us that the shape of the curved orbit can modify vertical chromaticity.

Turning to the third line we have two terms which, in a separated function lattice in which  $hk = 0$  just describe the effect of sextupoles on the betatron motion. These terms occur in our simple two - dimensional description. One can derive the strength of non - linear resonances from these terms. The fourth and fifth lines contain momentum dependent terms in which the curvature,  $h$ , is present and the last three terms which modify the betatron motion for on momentum particles where there is finite curvature and, presumably can contribute to non - linear resonance width.

If we follow the analysis of the "sliced bread" model of Bengtsson and which is also embodied in computer programs like ORBIT [7], MAD [8], we go a long way towards an exact description of the end field of magnets. The lattice functions, dispersion and chromaticity will turn out close to reality. However there remains one more effect which is less well known and which to the best of the author's knowledge is correctly embodied in only one computer program MIRKO [9]. We shall call it Electron Microscope Distortion since it is well known in that field.

## 7. EXCURSIONS WITHIN THE END FIELD

Suppose we return to the model of paraxial trajectories and ignore curvature for the moment. We have shown that the integrated field is two - dimensional but we know that locally this is not the case. The paraxial particle may see, say, a left hand deflecting field from the three dimensional nature of the field as it enters the end field region but our paraxial theorem tells us this is exactly cancelled by an equal and opposite right hand deflection somewhere else in the end field. Now suppose the particle is following a line which is not paraxial because the beta function is varying in the end field or perhaps it is an off - momentum particle whose displacement follows the dispersion function and is not therefore paraxial. The two perturbations will now be different because the paraxial theorem relies on the particle staying the same distance from the axis through the end field and this is no longer the case. In order

to understand such effects we must have a model for the end field which expresses how the multipole coefficients vary as a function of distance along the beam axis. There is a complete polynomial expression for the three - dimensional end field to be found in a text by Glaser on electron optics [10] and we reproduce this polynomial for the scalar potential as he wrote it.

$$\begin{aligned}
 \Phi_m(x, y, z) = & \Phi_m - Gx - Hy - \frac{1}{4}(\Phi_m'' - \Delta)x^2 + Qxy - \frac{1}{4}(\Phi_m'' + \Delta)y^2 \\
 & + \frac{1}{3}(\frac{1}{2}G' + G_1)x^3 - H_1x^2y - G_1xy^2 + \frac{1}{3}(\frac{1}{2}H' + H_1)y^3 \\
 & + (\frac{1}{64}\Phi_m^{(4)} - \frac{1}{48}\Delta'' + \Delta_1)x^4 - (\frac{1}{12}Q' - 4Q_1)x^3y \\
 & + (\frac{1}{32}\Phi_m^{(4)} - 6\Delta_1)x^2y^2 - (\frac{1}{12}Q' + 4Q_1)xy^3 \\
 & + (\frac{1}{64}\Phi_m^{(4)} + \frac{1}{48}\Delta'' + \Delta_1)y^4.
 \end{aligned} \tag{23}$$

The coefficients G, H, Q refer to vertical and horizontal dipole and quadrupole fields respectively. Other coefficients contain the axial derivatives of these quantities. We can still identify most of the multipole coefficients although in Glaser's notation we must remember that x and y are transverse and z is axial.

Table 1

Terms in Glaser's expansion according to multipole symmetry

NAME	$\Phi_m$	$B_y$	$B_x$	$dB_y/dx$	$dB_x/dy$
SOLENOID	$\phi_m - \phi r^2/4 + \phi^{(IV)}r^4/64$	$-\phi^{(IV)}(x^2y + y^3)/16$	$-\phi^{(IV)}(y^2x + x^3)/16$	$-\phi^{(IV)}xy/8$	$-\phi^{(IV)}xy/8$
H BENDING	$-Hy$	$H$	$0$	$0$	$0$
V BENDING	$-Gx$	$0$	$G$	$0$	$0$
QUADRUPOLE	$Qxy$	$-Qx$	$-Qy$	$-Q$	$-Q$
SKEW QUADRUPOLE	$\Delta(x^2 - y^2)/4$	$\Delta y/2$	$-\Delta x/2$	$0$	$0$
SEXTUPOLE	$H(y^3 - 3x^2y)/3$	$H_1(x^2 - y^2)$	$2H_1xy$	$2H_1x$	$2H_1x$
SKEW SEXTUPOLE	$G(x^3 - 3xy^2)/3$	$2G_1xy$	$G_1(y^2 - x^2)$	$2G_1y$	$2G_1y$
OCTUPOLE	$\Delta(x^4 - 6x^2y^2 + y^4)/4$	$4\Delta_1(3x^2 - y^3)$	$4\Delta_1(3xy^2 - x^3)$	$24\Delta_1xy$	$24\Delta_1xy$
SKEW OCTUPOLE	$4Q_1(xy^3 - x^3y)$	$4Q_1(x^3 - 3xy^2)$	$4Q_1(3x^2y - y^3)$	$12Q_1(x^2 - y^2)$	$12Q_1(x^2 - y^2)$
END QUADRATIC	$(G'x^3 - H'y^3)/6$	$H''y^2/2$	$-G''x^2/2$	$0$	$0$
END CUBIC	$-Q(x^3y + xy^3)/12$	$Q(x^3 + 3xy^2)/12$	$Q(3x^2y + y^3)/12$	$Q(x^2 + y^2)/4$	$Q(x^2 + y^2)/4$

In Table 1 we can see the same terms sorted according to the familiar multipoles and shown in the various columns are the expression for the fields and their gradients. Near the bottom of the table are two lines which we name "end quadratic" and "end cubic" which do not fit into our multipole description. The "end cubic" looks at first sight like an octupole but when its field gradient or focusing effect is plotted as a function of the transverse (x, y) in Figure 18 we see it is like a hammock while an octupole Figure 19 is a saddle shape. Anyone who has tried to sleep in a saddle or ride in a hammock will avow to the different topology of these functions.

If we look carefully at the symmetry of the end field cubic we find that while it has the symmetry of a quadrupole in  $\theta$  its radial dependence is that of an octupole. It is a characteristic of the end field of a quadrupole and proportional to the slope of the main gradient term, Q. Such terms must be included explicitly in any simulation of the end field shape although their effect in one plane can be thought of as a simple multipole.

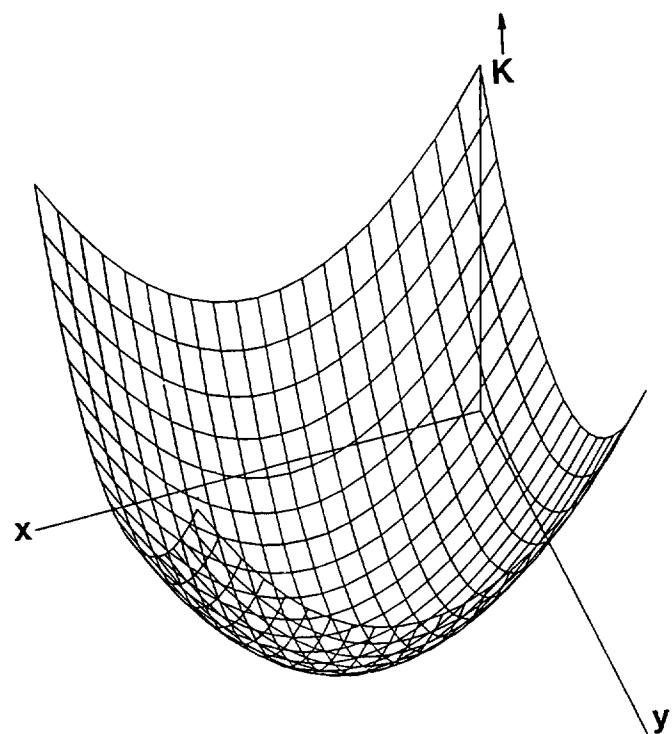


Figure 18: End quadratic potential.

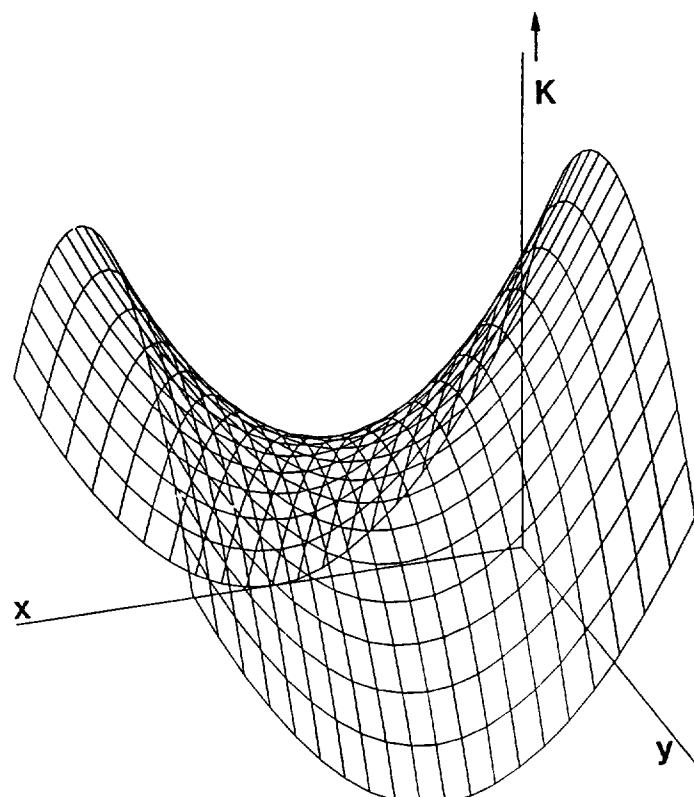


Figure 19: Octupole potential.

Because  $Q'$  reverses sign as we pass through the fringe field the paraxial integral of these terms is zero but they have an effect which is proportional to the slope of the betatron function,  $\alpha$ , and, in the case of an off - momentum particle, to the slope of the dispersion function [11].

At the time that the AA ring was designed this effect had not been discovered and it is still not incorporated in the standard lattice programs. When it was simulated with the MIRKO program, it exactly matched the curvature in  $Q$  versus momentum (Figure 15) before empirical shim correction was applied. Simulation has also shown that it can drive fourth order resonances. Clearly both of these effects become significant only in small rings of short magnets where the emittances and momentum spread are large.

## 8. CONCLUSIONS

In this review of the special features of small rings we have re - examined some of the approximations which are normally built into the analysis and the computer programs which describe transverse motion in today's large synchrotrons. The accelerator designer should be particularly careful of these approximations when the ring he is designing has magnets which bend through an angle of a few degrees or when the magnet length, measured in number of gap widths, is small.

The practical problem of having so many ends of quadrupoles, dipoles and other equipment in a small ring yet finding enough room for the components themselves is always severe but can be made less difficult if corrections are built into the magnet design and tuned by modifications to the ends of poles. One may even save the space for closed orbit dipoles by installing remotely controlled jacks to support the quadrupoles. This solution proved perfectly successful in the design of the AA and later the Antiproton Collector ring.

Finally, although there has been no time in this review to apply these lessons to small electron ring, many of the considerations are equally valid [13]. However the reader can readily appreciate that some of the correction methods may not be possible in a fast cycling electron ring and the acceptances may not be large enough for the finer points of the end field shape to be an important consideration. Nevertheless, the scenario for the design of the AA is a good vantage point to scan the horizon of possible pitfalls.

\* \* \*

## REFERENCES

1. Grivet and Septier. *Les Lentilles Quadrupolaire Magnétiques* CERN 58-25
2. B.Autin *Dispersion Suppression with Missing Magnets in a FODO Structure Application to the CERN Antiproton Accumulator* Proceedings of the 1979 Particle Accelerator Conference, San Francisco, IEEE Transactions on Nuclear Science Vol.NS-268, No.3, pp.3493-5, (1979)
3. R.Brown,T.Dorenbos,M.Frauchiger,C.D.Johnson,L.Petty *An Automated Measuring System for the Antiproton Accumulator Magnets* Proceedings of the 7th. International Conference on Magnet Technology, Karlsruhe, IEEE (1981)

4. B.Autin,R.Billinge,R.Brown,G.Carron,C.Johnson,E.Jones,  
H.Koziol,C.Leeman,T.R.Sherwood,S.van der Meer,E.J.N.Wilson *Beam Optics Studies on the Antiproton Accumulator* Proceedings of the 1981 Particle Accelerator Conference, Washington, IEEE Transactions on Nuclear Science Vol.NS - 28,No.3,pp.2055 - 7,(1981)
5. W.C.Elmore and M.W.Garret *Measurement of Two - Dimensional Fields. PartI:Theory Review* of Scientific Instruments, Vol 25, No.5 ,pp.482 - 483 (1954)
6. J.Bengtsson *Non - Linear Transverse Dynamics for Storage Rings with Applications to the Low - Energy Antiproton Ring (LEAR) at CERN* CERN 88 - 05
7. B.Autin and M.Bell *ORBIT - A Computer Program for Circular Machines* Users Guide (available from authors, CERN, Geneva)
8. F.C.Iselin and J.Niederer *The MAD Program - Users Reference Manual* CERN LEP - TH 88 - 38
9. B.Franczak *MIRKO - An Interactive Program for Beam Lines and Synchrotrons Computing in Accelerator Design and Operation*, Lecture Notes in Physics No.215 p.170 ,pub: Springer Verlag (1974)
10. W.Glaser *Elektronen &hyph und Ionenoptik* Handbuch der Physik, Band XXXIII, Springer Verlag (1956),p.166.
11. E.J.N.Wilson *Strange Multipoles in Magnet Ends* CERN PS/ACOL/Note 84 - 3
12. P.Krejcik *Non - Linear Quadrupole End - Field Effect in the CERN Antiproton Accumulator* Proceedings of the 1987 IEEE Particle Accelerator Conference, Washington, IEEE Catalog No. 87C#2387 - 9,pp.1278 - 9
13. J.P.Delahaye,A Krusche *The Lattice Design of the LEP Electron Positron Accumulator (EPA)* Proceedings of the 1983 Particle Accelerator Conference, Santa Fe, IEEE Transactions on Nuclear Science, Vol NS - 30,No.4,(1983),p.2050



## ANALYSIS OF PARTICLE TRACKING DATA

F. Willeke

Deutsches Elektronen-Synchrotron (DESY), Hamburg, Fed. Rep. Germany

### 1. INTRODUCTION

Nonlinear forces introduce a limit for stable betatron oscillation amplitudes of the particles in an accelerator beam. Beyond this limit, the oscillation amplitudes grow until the particles reach the wall of the beam pipe and a loss occurs. The issue of nonlinear forces in accelerators has become more and more important with the construction of slowly cycling synchrotrons and large storage rings, especially for hadron beams. In accelerators with superconducting magnets, which exhibit large systematic and nonsystematic nonlinear field errors of the focussing and bending magnets, it becomes one of the most crucial beam dynamics issues.

In general, analytical methods fail to describe the impact of nonlinear forces on the motion of an accelerator particle. Single strong resonant effects are exceptions which can be described well by perturbation theory but which are relevant only for special operating conditions such as slow resonant extraction. A considerable effort has been made to overcome this problem. However no reliable analytical procedure to determine the beam stability has yet been made available.

Instead, particle tracking has been developed as the most important tool for investigating the stability of the motion of single particles in accelerators. Design decisions for large accelerators (like HERA), such as specification of maximum tolerable field errors, chromaticity correction schemes, magnet apertures and optimization of linear lattice design, have been based to a considerable extent on the results of numerical particle tracking. Several tracking algorithms and corresponding computer codes have been developed in the last few years. However it turns out that not so much effort has been invested to provide tools for analysing the results which are produced by these computer codes.

What tracking programs calculate is conceptionally quite simple. The equation of motion for particles travelling through accelerator components is solved and the particle coordinates at the end of a beam transport element are used as initial conditions for the motion through the next element. This process is repeated over and over again and particles are tracked many times around the accelerator. A limitation of the number of turns tracked is given only by round-off errors due to the finite resolution of digital numbers. Depending on the size of the accelerator considered, this limit is found between  $10^5$  and  $10^7$  turns when one is operating with double precision numbers. Particles are considered stable if the oscillation amplitudes never exceed a certain specified value. It may be chosen to be the beam pipe radius or an artificially large aperture.

The results of tracking, the maximum stable amplitudes, depend on many control parameters, for example on tunes, chromaticity settings, momentum amplitudes, magnitude and distribution of nonlinear field errors, average of nonlinear field errors, presence of closed orbit errors, linear distortions, and linear coupling. Usually, the dependence of tracking results on these parameters is quite complicated and inconclusive. One reason is that for "economical" reasons, it is usually not possible to perform a complete scan of the parameters. The other reason is the complicated nature of the nonlinear dynamics which makes it difficult to draw conclusions from tracking data.

What is needed to improve this situation are models which provide a qualitative understanding of tracking data. Another need is to find ways to display the data appropriately in order to elucidate the information which they contain. Models enable us to derive numbers from the tracking data which may be used to parameterize the impact of nonlinearities on the stability of the motion. Such parameters make quantitative comparisons between different designs possible.

In this lecture, examples will be given of the analysis of tracking data using the familiar models of nonlinear accelerator theory. In particular, the concept of nonlinear distortions, the concept of nonlinear resonances, the concept of detuning, and the concept of chaotic motion will be used and discussed.

Collective effects will not be considered. Thus we are only concerned about single particle phenomena. Another important aspect of nonlinearities in accelerators which cannot be taken into account here is the impact of external excitation and noise. Because of their importance and subtlety, these subjects deserve separate lectures.

The outline of the lecture is as follows. After this introduction, a short review of the different tracking procedures and techniques will be given. This is followed by some general remarks about the performance of tracking calculations. The final section contains analysis of the tracking data as described above.

## **2. TOOLS FOR PARTICLE TRACKING**

One may distinguish two classes of tracking procedures and codes used in the accelerator community. The first class will be referred to as "successive integration method", because the equation of motion is solved piecewise, element by element, and the particles are tracked element by element around the machine. The second class of procedures uses a single one-turn transfer map which is obtained by concatenation of single element transfer maps. Both methods have their particular advantages and disadvantages which will be discussed below. What all the methods should have in common is the symplectic nature of the mapping used. This is a somewhat controversial topic, since symplecticity is sometimes considered to be of solely academic interest. Therefore, in the next section, the practical aspect of the symplectic condition in the context of tracking programs is discussed briefly.

### **2.1 The Symplectic Condition**

The motion of charged particles in accelerators is in general quite complicated. However, the full complexity is not always needed to describe the relevant effects in accelerators. Therefore it is justified to make approximations resulting in models of reduced complexity. The simplest model is the linear accelerator theory which is an analytic description of the particle motion. However, if one starts simplifying the equations of motion in a straightforward manner, there is a certain danger that nonphysical effects such as artificial dissipation [1] will be introduced into the model. These effects, even if they are very small, can lead to confusion in the interpretation of single particle tracking data, since nonlinear forces lead to effects which look very similar but which are of a quite different nature. Moreover, dissipative effects interfere strongly with nonlinearities (see for example ref.[23]) which can be studied only if dissipation can be controlled in the model. Therefore it is advantageous to start from a dissipation free, Hamiltonian model to derive the equations of motion

$$\dot{q} = \frac{\partial H}{\partial p}; \quad \dot{p} = -\frac{\partial H}{\partial q} \quad (1)$$

which may be written in vector form

$$\frac{d}{dt}\vec{z} = \underline{S} \nabla H(\vec{z}) \quad (2)$$

where  $\vec{z}$  is the phase space vector of a particle and  $\underline{S}$  is the symplectic form

$$\underline{S} = \begin{pmatrix} S_2 & 0 \\ 0 & S_2 \end{pmatrix}; \quad S_2 = \begin{pmatrix} 0 & +1 \\ -1 & 0 \end{pmatrix}. \quad (3)$$

The solutions of Hamilton's equation of motion have the remarkable feature that the density of particles in phase space is conserved during the motion (Liouville's theorem). This corresponds to the symplectic property of the Jacobian  $\underline{J}$  of the map which transforms initial coordinates into final coordinates

$$\underline{J} = \frac{\partial \vec{z}(t)}{\partial \vec{z}(0)}. \quad (4)$$

The time dependence of the Jacobian is given by

$$J_{ik} = \frac{\partial \dot{z}_i(t)}{\partial z_k(0)} = \frac{\partial}{\partial z_k(0)} \sum_l S_{il} \frac{\partial H}{\partial z_l} = \sum_n \frac{\partial z_n(t)}{\partial z_k(0)} \sum_l S_{il} \frac{\partial^2 H}{\partial z_l \partial z_n} \equiv J_{nk} \sum_{ln} S_{il} H_{ln} \quad or \quad \underline{J} = \underline{J} \underline{S} \underline{H} \quad (5)$$

Since  $\underline{S}^2 = -1$ ,  $\underline{S}^+ = -\underline{S}$  and  $\underline{H}_{nl} = \underline{H}_{ln}$  it follows that  $\underline{J}$  is a symplectic matrix

$$d/dt(\underline{J}^+ \underline{S} \underline{J}) = 0 \rightarrow \underline{J}^+ \underline{S} \underline{J} = \text{const.} = \underline{S}. \quad (7)$$

Maps which produce symplectic Jacobians are called symplectic maps. Solutions of Hamilton's equations are generated by symplectic maps. The determinant of a symplectic matrix is equal to unity because

$$\begin{aligned} \text{Det} \underline{J}^+ \text{Det} \underline{S} \text{Det} \underline{J} &= \text{Det} \underline{S} \\ \text{Det} \underline{S} &= 1 \rightarrow \text{Det} \underline{J} = \pm 1. \end{aligned} \quad (8)$$

Therefore the volume in phase space occupied by particles is invariant during the motion of the particles:

$$\begin{aligned} &\int \dots \int dq_1(0) dp_1(0) \dots dq_N(0) dp_N(0) \\ &= \int \dots \int dq_1(t) dp_1(t) \dots dq_N(t) dp_N(t) \text{Det} \underline{J}(0, t) \\ &= \int \dots \int dq_1(t) dp_1(t) \dots dq_N(t) dp_N(t) \end{aligned} \quad (9)$$

The positive sign of  $\text{Det} \underline{J}$  follows from the fact that the determinant is +1 for  $t = 0$ . This is Liouville's theorem. In accelerator language it means that there is no emittance blow up or damping during the motion around the accelerator if it is described by a symplectic map. All effects which appear to be dissipative must then be produced by nonlinearities owing to filamentation effects which may lead to considerable increase of the "effective" volume in phase space occupied by the beam. Tracking codes used should therefore generate symplectic maps, otherwise it will be very difficult to study weak nonlinear instabilities and their impact on the machine performance.

## 2.2 The Method of Successive Integration

The method of successive integration consists of tracking the particles element by element using analytic formulae for the transfer map through each element. Since the exact expression for the particle transfer through a nonlinear element of finite length  $l$  is very complicated and its evaluation would require a lot of computing time, one is forced to make a thin lens approximation. The nonlinear force is concentrated in a point

$$\vec{F}(s) \rightarrow \sum_i \int_{s_i - l/2}^{s_i + l/2} ds' \vec{F}(s') \delta(s_i - s). \quad (10)$$

Then the change of the particle position during the action of the nonlinear force is neglected and only a change in transverse momentum (kick) occurs. Corresponding computer codes are called kick codes.

The usual kick code starts from the solution of the linearized equation of uncoupled transverse motion (linear optics) for fixed momentum  $p = p_0 \cdot (1 + \frac{\Delta p}{p_0})$

$$\begin{aligned} x'' + (k + G^2)x &= \delta \\ y'' - k + y &= \delta \\ G &= \frac{\epsilon}{p} B_y; \quad k = \frac{\epsilon}{p} \partial B_{y,x} / \partial x, y; \quad \delta = \frac{\Delta p}{p} G \end{aligned} \quad (11)$$

( $x$  is the horizontal or radial and  $y$  is the vertical coordinate in the usual accelerator coordinate system and  $B_{x,y}$  are the transverse components of the component of the bending and focussing fields) by adding nonlinear thin-lens kicks in between linear beam transport elements.

$$\begin{aligned} \Delta x' &= -\Re \int dl G \sum_n (b_k + ia_k)((x + iy)/r_0)^n \\ \Delta y' &= +\Im \int dl G \sum_n (b_k + ia_k)((x + iy)/r_0)^n \end{aligned} \quad (12)$$

( $b_k$  and  $a_k$  are the normal and skew multipole components relative to the bending field which are measured at a radius  $r_0$ .) Single linear elements with no nonlinear kicks in between are concatenated to a block which is described by a single linear transfer matrix. Kick codes like RACETRACK [2,3] and PATRICIA [4] are examples for computer codes used in many laboratories.

There is no straightforward method for including longitudinal oscillations into such kick codes. The description of the coupling between transverse and longitudinal motion in a symplectic manner requires nonlinear terms in the transverse variables to be taken into account even for elements with solely linear

forces [5]. Furthermore it appears to be inconsistent to neglect all the nonlinear terms which arise from the relativistic nature of the motion and which are nonlinear in the transverse momentum, whereas the nonlinearities in the transverse coordinate are added later in the form of nonlinear kicks. The reason why this is justified in most cases is the different size of these two sources of nonlinearities in large accelerators. In very small accelerators, this is not necessarily the case.

To avoid such inconsistencies especially for small machines, another tracking algorithm has been developed which treats all elements as thin lenses [6]. In this case the nonlinearities in transverse momentum which also occur in the "linear" elements such as dipole and quadrupole magnets (see for example ref.[5,1]) can be treated the same way as the nonlinearities in the coordinates due to multipole fields. There are no complications if synchrotron oscillations are taken into account since thin lenses do not contribute to the path length of a particle trajectory and there are no problems in violating symplecticity. There is a computer code available called "TEAPOT" which is based on this philosophy. A certain inconvenience in using this procedure is that the linear optics is not exactly reproduced in the thin-lens description and one needs some readjustment of the strengths of linear elements (matching) in order to describe the linear lattice properly.

An alternative treatment of synchro-betatron coupling is obtained by consistently pursuing a Hamiltonian, symplectic description of the motion when making approximations. In this case the equation of motion for synchro-betatron oscillations then will always contain nonlinear terms even for magnets with linear fields. The differential equations however can be solved analytically [5]. The nonlinear terms are necessary to describe the feedback of the betatron amplitude on the path length which establishes synchro-betatron coupling. There is a computer code available called "SIXTRACK" [7] which is based upon this formalism. Because of the nonlinear terms, the particle coordinates are transformed element by element. The concept of concatenating "linear" elements has to be abandoned in order to preserve symplecticity when the synchrotron oscillations are included in the tracking. Thus computation speed is somewhat reduced with respect to the simple kick codes, which is the only disadvantage.

### 2.3 Mapping Methods

Transforming particle coordinates element by element using a kick approximation to take into account nonlinear terms in the equation of motion is a very straightforward and (computing) time-consuming way of particle tracking. More effective procedures would be very helpful in order to perform long time tracking for large accelerators. Attempts have been made to describe the motion through a whole section of the lattice, in particular the motion once around the accelerator, by a closed form map. The exact map could in principle be represented by a polynominal in the initial particle coordinates. However this polynominal would include very high powers (a lattice with 100 sextupoles is described by a polynominal of order 200) and it would be quite impractical. In order to make a progress, the map has to be terminated. The terminated map corresponds to a Taylor series expansion of the motion around the closed orbit. The approximate map can for example be generated numerically using the mapping provided by a kick code. One should mention at this point differential algebra methods which allow one to calculate exact Taylor expansion coefficients of any mapping up to arbitrary order [8,9,10]. Differential algebra is a technique to propagate functions and their derivatives through a series of mathematical operations using the rules of differentiation. Taylor expansion maps are in general not symplectic, so they cannot be used directly for tracking. A symplectification procedure is required in order to overcome this (see for example [9]).

Approximate maps can also be built up analytically. A very elegant method [11] uses Lie algebra techniques. In the following, a brief description of the basic ideas is given. A formal solution of Hamilton's equation of motion

$$\frac{\partial \vec{z}}{\partial t} = \underline{S} \nabla H(\vec{z}) = [\vec{z}, H] \equiv \mathcal{L} \vec{z} \quad (13)$$

(the brackets denote the Poisson bracket) is written as

$$\vec{z}(t) = e^{\int_0^t \mathcal{L} dt} \vec{z}(0) \equiv e^{\underline{L}} \vec{z}(0) = \left( [ , L] + \frac{1}{2} [[ , L], L] + \frac{1}{6} [[[ , L], L], L] + \dots \right) \vec{z}(0) \quad (14)$$

Conversely, a mapping generated by an operator  $e^{\underline{L}} = e^{[ , L]}$  where  $L$  can be any function of the initial phase space variables, is a solution of some set of Hamilton's equation and the map is a symplectic map

[12]. These operators are called "Lie operators". They have some interesting formal properties which make them useful for beam transport problems. Transformation of particle coordinates through beam transport elements can be described by a Lie operator. For example, the horizontal motion through a quadrupole is described by

$$\begin{pmatrix} x \\ p_x \end{pmatrix}_f = e^L \begin{pmatrix} x \\ p_x \end{pmatrix}_i = \begin{pmatrix} \cos \sqrt{kl} & \frac{1}{\sqrt{k}} \sin \sqrt{kl} \\ -\sqrt{kl} \sin \sqrt{kl} & \cos \sqrt{kl} \end{pmatrix} \begin{pmatrix} x \\ p_x \end{pmatrix}_i \quad (15)$$

where

$$L = \frac{\sqrt{kl}}{2} (x^2 \sqrt{k} + p_x^2 / \sqrt{k}), \quad \hat{L} = [ , L]$$

or a thin-lens sextupole kick is described by

$$L = \frac{ml}{6} x^3 \quad (16)$$

where  $x, p_x$  are initial values of the canonical variables in the accelerator coordinate system.

An important property of Lie operators is that they can be concatenated using

$$e^{\hat{L}} e^{\hat{K}} = e^{([\cdot, L+K] + \frac{1}{2} [\cdot, [L, K]] + \frac{1}{12} [\cdot, [L, [L, K]]] + \dots)} \quad (17)$$

(the Baker Campbell Hausdorff relation) and an exchange rule

$$e^{\hat{L}} e^{\hat{K}} = e^{\hat{K}} e^{\hat{G}L}, \quad \hat{G} = e^{[\cdot, -K]}. \quad (18)$$

The concatenated map can be reorganized and represented in the form

$$e^L = e^{\hat{L}_1} e^{\hat{L}_2} e^{\hat{L}_3} e^{\hat{L}_4} \dots \quad (19)$$

where  $L_m$  are homogeneous polynomials in the phase space variables. Since Lie operators always produce symplectic maps, the map can be truncated at any order  $m$  without violating the symplectic condition. Unfortunately a Lie operator of this form does not generally terminate if applied to a phase space vector. A more convenient form is obtained by identifying the mapping by a canonical transformation from initial to final phase space coordinates  $\tilde{z}_{initial} \rightarrow \tilde{z}_{final}$ . This transformation is generated by a generating function  $G(q_{initial}^i, p_{final}^i)$  which is mixed in initial and final coordinates. For every Lie operator  $e^L$  one can write down the corresponding generating function  $G[13]$ . The transformation equations which describe for example the motion once around the accelerator

$$\begin{aligned} x_{final} &= x_{initial} + \partial G / \partial p_{final} \\ p_{initial} &= p_{final} + \partial G / \partial x_{initial} \end{aligned} \quad (20)$$

have to be solved iteratively for example by Newton's method. The computer code MARYLIE is the result of the implementation of this procedure. For more details see references [14].

Both methods; successive iteration methods and mapping methods are only an approximate solution of the equation of motion. In both cases, the approximation can be refined; in the case of a thin-lens description by describing a thick element by more than one thin lens and in the case of truncated maps by subdividing the accelerator into more than one part described by a single truncated map. Such improvements will be limited by consumption of computing time. A certain advantage of the kick code is that we can easily see what the local effect of the approximation is. The impact over a large number of simulated turns is in both cases not transparent. Tests of the optimum degree of approximation for any method are necessary every time a new accelerator system is to be investigated. Several intermediate checks of the reliability are recommended during comprehensive systematic tracking studies. The analysis of particle tracking data proposed in this lecture does not depend upon a particular tracking algorithm. The data could be produced by any of the procedures described in this section.

### **3. SOME SUGGESTIONS HOW TO PERFORM PARTICLE TRACKING**

The following suggestions turned out to be useful in the performance of systematic particle tracking for large accelerators.

Suggestion 1. Before one starts to do tracking, one should have a good understanding of the properties of linear optics. There is an interference of linear optics properties and the impact of nonlinearities. The sensitivity of the linear optics to linear distortions should be explored in order to be able to understand the impact of distortions on the dynamic aperture. The same is also true for chromatic effects, physical acceptance limitations and linear coupling. Systematic tracking calculations should always be preceded by a systematic investigation of the linear optics properties.

Suggestion 2. The values of the phase space coordinates of the particles obtained for each turn make important information necessary for analysing the tracking results. This information needs to be preserved by recording the coordinates on every completed turn. Most of the kick codes do not provide this option and therefore need to be modified. Particle coordinates should be dumped on a file which can later be used by a postprocessor for analysis.

Suggestion 3. Tracking of just one particle does not provide sufficient information. Several particles are needed to probe different regions of phase space. On the other hand, not much additional information is obtained if several particles are started with the same initial betatron oscillation amplitude but with different initial betatron phases. Therefore initial particles should be distributed over the space of betatron amplitudes ( horizontal, vertical, longitudinal amplitude).

Suggestion 4. A single tracking run does not provide sufficient information. There are several important "control parameters" which need to be scanned in order to explore the dynamical situation in the accelerator studied. The most important parameters are the tunes. The dependence of the tracking results on the tunes is a valuable source of information about the nature of the nonlinear impact. Important parameters to be varied are also chromaticities, momentum amplitude, strength of the nonlinearities, linear optics distortions and closed orbit.

Suggestion 5. If closed orbit distortions are introduced into the lattice in order to study the interference with the nonlinearities, one should keep in mind that a corrected closed orbit is qualitatively very different from an uncorrected one. The closed orbit deviations of an uncorrected orbit are strongly correlated whereas the corrected orbit deviations are uncorrelated. To study closed orbit effects, the tracking code used should include an orbit correction algorithm (most codes do not).

Suggestion 6. In order to do systematic tracking, a convenient way of changing the tunes should be provided. Changing the tunes should not disturb the beam optics so that it will be possible to distinguish between the effect of modified tune and modified lattice functions. Dependence of the tunes on the oscillation amplitude is also an important effect which needs to be studied. Since it would be cumbersome to reoptimize the lattice design, optics matching and the arrangement of nonlinear elements which enable these parameters to be changed without disturbing the whole set up, it is convenient to introduce artificial elements. They only act on the tune and they are described by a transfer matrix of the form (for example only horizontal plane)

$$\begin{pmatrix} \cos(2\pi\Delta Q_x) + \alpha_x \sin(2\pi\Delta Q_x) & \beta_x \sin(2\pi\Delta Q_x) \\ -\gamma_x \sin(2\pi\Delta Q_x) & \cos(2\pi\Delta Q_x) - \alpha_x \sin(2\pi\Delta Q_x) \end{pmatrix} \quad (21)$$

$\Delta Q_x$  is the tune change required,  $\alpha$ ,  $\beta$ ,  $\gamma$  are the usual linear lattice functions.  $\Delta Q$  may depend on the oscillation amplitudes, the momentum, on external excitation and on noise effects. Most tracking programs unfortunately do not provide the option of defining tune changes in this way but it would be worthwhile to modify a particular code to include the option.

### **4. INTERPRETATION OF TRACKING DATA**

The concepts of nonlinear dynamics in accelerators which will be used in the following sections to interpret tracking data are:

- Nonlinear Distortions
- Nonlinear Resonances
- Nonlinear Tuneshift with Amplitude (Detuning)
- Chaotic Motion

These concepts or models are approximations which give a good description of the nonlinear dynamics of accelerator particles well in particular situations. The numbers produced by these models have to be interpreted very carefully whenever the models are applied outside the (usually quite small) range of validity. Parameters calculated in this case are then no longer a precise description of the dynamics, but a "measure" of the nonlinear effect which can only be used for comparisons and scaling. Nonetheless, these models provide a good qualitative description of the beam behaviour simulated by particle tracking.

#### **4.1 THE CONCEPT OF NONLINEAR DISTORTIONS**

An accelerator beam which performs a nonlinear motion is difficult to control. Therefore, in order to be able to operate the machine, one has to require that the beam is inside a region where nonlinearities cause only small distortions, so that the motion and the effect of adjustments can be predicted. This leads to the requirement that the motion should be sufficiently linear. The boundary of the region where this is the case is sometimes called the "linear aperture". In order to determine the linear aperture one needs a criterion for "nonlinearity". The model of nonlinear distortions is an attempt to provide such a criterion.

Consider a single particle travelling around an accelerator ring. If one measures the position and the slope of the trajectory at a fixed position in the ring every time the particle goes by, one obtains a phase space plot. Note that the phase space coordinates are not recorded continuously but for constant "time" intervals. The particle positions are the canonical coordinates, the slopes of the trajectory are the canonical momenta. If the motion of the particle is determined by linear forces, the phase space vectors are found on the boundary of an ellipse<sup>1</sup> for the motion in one degree of freedom. If a small nonlinearity is turned on, the ellipses will be distorted but will still enclose the same area. If the nonlinearity is increased (or the initial amplitude is increased), the distortions will become larger and the phase space trajectories will show new features such as chains of so called islands. Beyond this, the phase space coordinates are no longer to be found on a closed curve (contour) but they are scattered irregularly over an area in phase space with no apparent structure. Finally there is a domain, where the particle amplitude is rapidly growing and the particles will soon be outside the physical aperture. The concept of nonlinear distortions is only valid in the first part of this scenario. In the following paragraph, the basic ideas for obtaining a description of the nonlinear distortions are presented. For the sake of clarity, motion in only one degree of freedom is considered whenever the treatment of the more general case is similar. More stringent and comprehensive descriptions are to be found in various issues of the proceedings of the CERN and US Accelerators schools (see for example [15]) as well as in internal laboratory reports [16]).

In general nonlinear problems cannot be treated exactly. However there are cases in which perturbation theory provides an adequate description of the nonlinear motion. One of the concepts is the one of nonlinear distortions. It leads to an interesting insight into the nature of nonlinear motion which helps to recognize the structure hidden inside the results of particle tracking. By nonlinear distortions we understand small deviations from the motion described by linear theory of betatron (and synchrotron) motion. There are several procedures which arrive at an appropriate description. In order to interpret tracking data it turns out that the classical treatment which goes back to Poincaré [17] is quite useful.

The basic idea of classical perturbation theory is to keep the analytical form of the solution of the linear equation of motion and to vary the constants of motion obtained from linear betatron theory which are  $\epsilon = x^2\gamma + 2\alpha xx' + x'^2\beta \equiv 2\epsilon$  and the initial betatron phase  $\phi = \arctan\left(\frac{x_0\alpha + x'_0\beta}{x_0}\right)$  in order to satisfy the nonlinear equation of motion.

$$x(s) = \sqrt{2\epsilon\beta(s)} \cos(\psi(s) + \phi) \quad (22)$$

---

<sup>1</sup>If  $\tilde{x}' = x'\beta + x\alpha$  is plotted versus  $x$ , the phase space becomes circular.  $\beta$  and  $\alpha$  are the usual lattice functions provided by the linear beam optics

( $\beta(s)$ ,  $\alpha(s)$ ,  $\gamma(s)$ , and  $\psi(s)$  are the usual linear lattice functions.) If one inserts the solution of the linear equation of motion into the nonlinear equation which may be derived from a Hamiltonian

$$\mathcal{H} = \frac{1}{2} p_x^2 + \frac{1}{2} k(s) x^2 + H \quad \begin{matrix} \text{linear part} & \text{nonlinear part} \end{matrix} \quad (23)$$

and one allows the constants of motion to vary, one arrives at a differential equation for these constants which is of Hamiltonian form

$$\frac{d\epsilon}{ds} = -\frac{\partial H}{\partial \phi} \quad \frac{d\phi}{ds} = +\frac{\partial H}{\partial \epsilon} \quad (24)$$

This means that we have performed a canonical transformation from the coordinates  $\{x, x'\}$  to  $\{\epsilon, \phi\}$  which corresponds to a transformation to action ( $\epsilon$ ) and angle ( $\phi$ ) variables as described in any textbook on classical mechanics.

The new, nonlinear Hamiltonian  $H$  is the nonlinear part of the original Hamiltonian, which, according to the theory of betatron motion is the vector potential from which the nonlinear fields are derived. It may be represented by a multipole expansion

$$H(x, s) = \sum_{n=0}^{\infty} A_n(s) x^n \quad (25)$$

or

$$H(x, y, s) = \sum_{n,m=0}^{\infty} A_{nm}(s) x^n y^m \quad (26)$$

for two degrees of freedom. This expansion is particularly useful if small field distortions of the main accelerator magnets are considered <sup>2</sup>. This Hamiltonian is expressed by the new coordinates  $\epsilon, \phi$

$$H = \sum_{n \text{ even}} \left( \frac{n}{2} \right) \left( \frac{\beta}{2} \right)^{n/2} A_n \cdot \epsilon^{n/2} + 2 \sum_n \sum_{\nu} \left( \frac{n-\nu}{2} \right) \left( \frac{\beta}{2} \right)^{n/2} A_n \cdot \epsilon^{n/2} \cos(\nu\psi(s) + \nu\phi) \quad (27)$$

(where  $\nu$  runs in steps of 2 from  $n, n-2, \dots$  to 1 or 0). The multipole coefficients  $b_k$ , and the lattice functions  $\beta(s)$ ,  $\psi(s) - 2\pi Q$  ( $Q$  being the linear tune  $2\pi Q = \psi(s+L) - \Psi(s)$ , the betatron phase advance for one turn around the accelerator) are periodic functions. The period length is the machine circumference  $L$ . We can take advantage of this periodicity by expanding  $H$  in Fourier series (index  $q$ )

$$H(s) = \frac{1}{2} \sum_n h_{n00} \cdot \epsilon^{n/2} + \sum_n \sum_{\nu} \sum_q h_{n\nu q} \cdot \epsilon^{n/2} \cos(\nu\phi + (\nu Q + q)2\pi s/L + \phi_{n\nu q}) \quad (28)$$

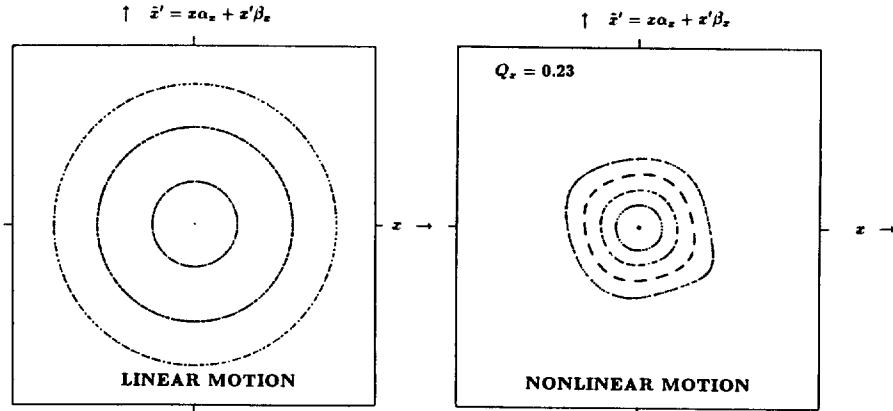
with

$$h_{n\nu q} = \left| \frac{2}{L} \int_s^{s+L} ds \left( \frac{n}{2} \right) \left( \frac{\beta}{2} \right)^{n/2} A_n e^{i(\nu\psi(s) + (\nu Q + q)2\pi s/L)} \right|.$$

The impact of terms in this expansion, which have a rapid oscillating "time" ( $= s$ ) dependence ( $\nu Q + q \geq 1$ ) on the variation of  $\epsilon$  and  $\phi$ , will essentially cancel over many turns, provided that the rate of change in  $\epsilon$  and  $\phi$  is small during a period  $\Delta s = \frac{L}{2\pi(\nu Q + q)}$ . This rate depends on the strength of the coefficients  $h_{n\nu q}$  as well as on the actual value of  $\epsilon$  itself. We may distinguish between three different cases:

- (a)  $\epsilon, \phi$  change only little during a period  $\frac{L}{2\pi(\nu Q + q)}$ , thus  $\Delta\epsilon/\epsilon \ll 1$ . In this case it is most useful to describe the motion by nonlinear distortion functions generated by perturbation theory.
- (b)  $\epsilon, \phi$  change significantly since there is at least one term in the expansion for which the denominator  $\nu Q + q$  is very small. In this case, the nonlinear force has a component which oscillates with the betatron frequency and even a small force can build up a large impact over a sufficient amount of "time". This is the case of a nonlinear resonance. Resonant phenomena play an important role in accelerators. They can be described quite elegantly by a technique called secular perturbation theory as we will see in the next section.

<sup>2</sup>The expansion coefficients  $A_{nm}$  are related to the usual multipole coefficients in the expansion  $B_y + iB_z = B_0 \sum_k (b_k + ia_k)(x + iy)^k / r_0^k$  by  $A_{nm} = (-1)^m b_k k! / n! m!$  for  $m$  even or  $A_{nm} = (-1)^m a_k k! / n! m!$  for  $m$  odd where  $k = n + m - 1$ .



**Figure 1:** Phase space plot for linear and nonlinear motion with one degree of freedom. Plotted is  $\tilde{z}' = z\alpha_z + z'\beta_z$  versus  $z$  which corresponds to a polar coordinate plot of  $\sqrt{\epsilon} = f(\phi)$ . The distortion pattern show four maxima and minima suggesting a prominent contribution from a term with  $\nu = 4$  which could be generated by octupole fields in first order or by sextupole fields in second order.

The full scale of these plots is in the order of a beam pipe radius.

(c)  $\epsilon, \phi$  change significantly while the denominator  $\nu Q + q$  is considerably different from zero. Thus there are terms  $h_{n\nu q}\epsilon^{n/2}$  which are strong enough to have a large impact within  $\Delta s = \frac{L}{2\pi(\nu Q+q)}$ . In this case, the perturbation series converges only poorly if it converges at all. Also the resonance treatment cannot be applied since there is usually more than just one single strong term. Note that the situation (a) will always transform into the situation described here if only the amplitude, thus  $\epsilon$ , is increased sufficiently. Therefore, this behaviour is quite often found close to the stability limit. The difficulty in describing the situation is the reason why it is not yet possible to provide a reliable analytic estimate of the dynamic aperture.

In the model of nonlinear distortions one pretends that it is possible to adjust the tunes far from any possible resonances  $\nu Q + q \approx 0$  and that all the terms in the Fourier expansion are sufficiently small since one is considering trajectories with sufficiently small amplitudes. In this case we are looking for a canonical transformation which transforms the system  $H, \epsilon, \phi$  into a new system  $K, J, \varphi$  where the Hamiltonian only depends on the action variable  $J$ . The solution of such a system is trivial since

$$J' = -\partial K / \partial \varphi = 0 \rightarrow J = \text{const} \quad \text{and} \quad \varphi' = \partial K / \partial J = 0 \rightarrow \varphi = f(J) \cdot s. \quad (29)$$

Following the procedure of classical perturbation theory, we choose a generating function  $S$  which is mixed in old and new canonical variables

$$S(J, \phi, s) = J\phi + \sigma(J, \phi, s) \quad (30)$$

from which one derives the transformation equations

$$\begin{aligned} \epsilon &= J + \partial\sigma(J, \phi, s)/\partial\phi \\ \varphi &= \phi + \partial\sigma(J, \phi, s)/\partial J \\ K &= H + \partial\sigma(J, \phi, s)/\partial s \end{aligned} \quad (31)$$

One can construct a transformation which removes approximately all those terms from  $H$  which depend on the action and on phase and "time":

$$S = J\phi + \sum_n \sum_\nu \sum_q \frac{-h_{n\nu q}}{\nu Q + q} \cdot J^{n/2} \cos(\nu\phi + (\nu Q + q)2\pi s/L + \phi_{n\nu q}) \quad (32)$$

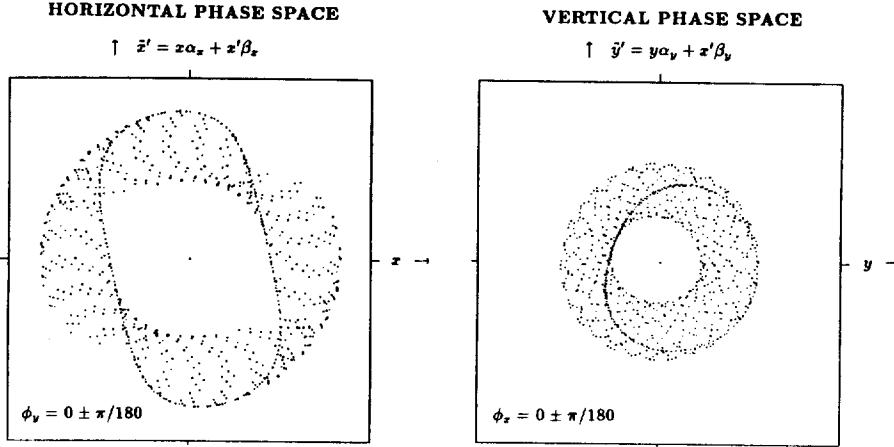


Figure 2: Projection of a four-dimensional nonlinearly distorted phase space onto the  $\tilde{x}' - x$  and the  $\tilde{y}' - y$ -plane respectively. The fat points which lie in a small phase interval in the vertical plane form a closed contour in the horizontal plane. The full scale of the plot scales with the strengths of the nonlinearities.

Applying the corresponding transformation replaces  $\varepsilon^{n/2}$  by  $(J^{n/2} - \varepsilon^{n/2} \simeq \frac{n}{2J} \partial S / \partial \phi + \dots)$  which leaves us with

$$\begin{aligned}
 K &= \frac{1}{2} \sum_n h_{n00} \cdot J^{n/2} \\
 &- \sum_{nn'} \sum_{\nu\nu'} \sum_{qq'} \frac{n\nu'}{4} \frac{h_{n\nu} h_{n'\nu'q'} J^{(n'+n-2)/2} \cos((\nu \pm \nu')\phi + ((\nu \pm \nu')Q + q + q')2\pi s/L + \phi_{nn'\nu\nu'qq'})}{\nu'Q + q'} + \dots \quad (33) \\
 &\equiv \frac{1}{2} \sum_n k_{n00} \cdot J^{n/2} + \sum_n \sum_\nu \sum_q k_{n\nu q} \cdot J^{n/2} \cos(\nu\varphi + (\nu Q + q)2\pi s/L + \varphi_{n\nu q})
 \end{aligned}$$

which is of second and higher order in  $h_{n\nu q}$ , the effective strength of the nonlinearity. We can see that this new Hamiltonian now contains new potentially resonant terms which, for a given maximum number of  $n$ , will drive resonances in the order of up to  $2n$ . These potential resonances are usually referred to as "higher order resonances". The tune parameter  $\tilde{Q}$  in eqn.31 is defined by  $\tilde{Q} = Q + \frac{n}{2} \sum_n h_{n00} J^{(n-2)/2}$ . It replaces the linear (zero amplitude) tune  $Q$  in order to cancel those second-order contributions which arise from expressing the terms  $\sum_n h_{n\nu q} \varepsilon^{n/2}$ , the ones which are retained in the new Hamiltonian, by the new action variable  $J$ . As we will see later,  $\tilde{Q}$  has a physical meaning. It is the tune which includes the amplitude dependence of the tune due to the nonlinearities. Therefore, the retained terms are also called "detuning terms". Thus the occurrence of a resonance does not only depend on what value the "linear tune" has been adjusted to, but also on the actual oscillation amplitude. In order to arrive at the final goal of a "trivial" Hamiltonian, one can introduce a second generating function which removes the second order terms from the new Hamiltonian but introduces third-order terms. Thus we have to consider infinite perturbation series consisting of repeating the first step over and over again. It is apparent from eqn.31 that the series will only converge as long as the denominators  $\nu\tilde{Q} + q$  do not become too small, so that the contributions from higher-order terms have a chance to become smaller on each iteration. Since during the perturbation series more and more potential resonances will appear and since the tune will change due to the additional detuning terms which appear on each iteration, it is not trivial to avoid the higher order resonances by a careful adjustment of the linear tune. This is a big technical problem in carrying out the transformation series. It is also the reason why the series converges very poorly and why it is very complicated to prove convergence. Nonetheless, despite the fact that for every rational number of the tune  $\tilde{Q}$ , there is a potential resonance to be excited, there are dense layers of orbits for which the perturbation series converges (under certain conditions which are stated by the KAM theorem, see for example [18]). Such orbits are called regular orbits. In accelerators one is interested in a large area in phase space around the closed orbit which contains at most regular orbits, since regular orbits are also stable orbits which are bound to a closed surface in phase space. This can be seen from the final form of the generating function

obtained by carrying out the sum over the Fourier series analytically. The form of the generating function which is obtained on the first iteration step is preserved throughout the whole perturbation series. The result is written down here for two degrees of freedom corresponding to transverse motion of accelerator particles with constant momentum (coasting beam)

$$S = J_x \phi_x + J_y \phi_y + \sum_{n,m,\nu\mu} \frac{\sigma_{nm\nu\mu} J_x^{n/2} J_y^{m/2} \sin(\nu\phi_x + \mu\phi_y + \phi_{nm\nu\mu})}{\sin \pi(\nu\bar{Q}_x + \mu\bar{Q}_y)}. \quad (34)$$

Note that the generating function does not contain any explicit "time" dependence apart from a ring periodic  $s$ -dependence, which is suppressed in the above representation since we are considering a fixed point in the lattice. Thus the generating function is a ring periodic function like the linear lattice functions. One can consider it as a "nonlinear lattice function".

What do we learn for the interpretation of tracking data from these considerations? First, let us consider tracking data from a regular orbit for which one can in principle carry out the complete transformation series so that we can assume that  $J_x, J_y$  are constant since  $\partial K / \partial \varphi_{x,y} = 0$ . Therefore, according to eqn. 30, the original action variables  $\varepsilon_{x,y}$  which we can calculate from the actual particle coordinates  $x, x', y, y'$  obtained by the numerical tracking, are functions of the original phase variables  $\phi_{x,y}$  which are also determined by the particle coordinates

$$\varepsilon_{x,y} = \partial S / \partial \phi_{x,y} = J_{x,y} + \frac{\partial}{\partial \phi_{x,y}} \sigma(J_x, J_y, \phi_x, \phi_y). \quad (35)$$

This property suggests a special way of displaying the tracking results. For motion in one degree of freedom, a phase space plot  $\tilde{x}'$  versus  $x$  results in closed in phase space contours  $\sqrt{\varepsilon} = f(\phi)$ . From these curves, we can read which terms are important contributions to the nonlinear distortion just by counting the maxima and minima of the distortion pattern (see Fig.1). The motion in two degrees of freedom, extends over a four-dimensional phase space. If the phase-space trajectories are projected onto the  $\tilde{x}'$ - $x$  and  $\tilde{y}'$ - $y$  plane respectively, which is the most convenient way of plotting the tracking results, the structure of the phase space distortions is invisible (See Fig 2). If one selects however a subset of data points which correspond to the same value of the vertical betatron phase angle (for example  $\phi_y = 0$ ), these points lie on a contour in the horizontal phase space since

$$\varepsilon_x = J_x + \sum_{n,m,\nu\mu} \frac{\nu \sigma_{nm\nu\mu} J_x^{n/2} J_y^{m/2} \cos(\nu\phi_x + \phi_{nm\nu\mu})}{\sin \pi(\nu\bar{Q}_x + \mu\bar{Q}_y)} = f(\phi_x). \quad (36)$$

In other words we are looking at a phase space cut for  $\phi_x = \text{const}$ . The resulting contour contains the same information as in the case with one degree of freedom (see Fig.2). It reveals the strength of different contributions to the perturbation series especially if prominent contributions are present. This procedure of data selection is reminiscent of a "stroboscopic" observation method.

There is another benefit from the perturbation treatment. A central figure in accelerator language is the beam emittance  $\epsilon$  which also characterizes the oscillation amplitude of a single particle. (The beam emittance corresponds to the oscillation amplitude of a particle at the "surface" of the beam.) However, the emittance is only a measure of the betatron amplitude in the linear theory. The concept of nonlinear distortion suggests that a natural generalisation of the linear emittances are the invariants  $J_{x,y}$  which are related to  $\varepsilon_{x,y}, \phi_{x,y}$  by

$$J_{x,y} = \frac{1}{2\pi} \oint d\phi_{x,y} \varepsilon_{x,y}. \quad (37)$$

In order to obtain the generalized emittances  $J_x$  from the tracking data, we again make a phase space cut for  $\phi_y = \text{const}$  and integrate the selected values of  $\varepsilon_x$  over the horizontal phase  $\phi_x$ . The integrals for different choices for the phase angle of the cut should be the same, so that it is allowed to average over the corresponding values for  $J_x$  to make efficient use of all the tracking data. Note that averaging the values of  $\varepsilon$  does not give the same result since particles are found mainly around maxima of the distortion pattern, so that the average value of  $\varepsilon$  is larger than the integral <sup>3</sup>.

<sup>3</sup>The phase space distortions have to preserve the phase space area (a consequence of Liouville's Theorem). Thus where the phase space is stretched in the phase variable it is compressed in the action variable and vice versa. Therefore, the rate by which the distorted phase advances is increased with respect to the linear phase advance near minima of the phase space distortions but decreased near maxima of the distortions. Thus particle coordinates are preferably found near distortion maxima.

### HERA-e VERTICAL PHASE SPACE

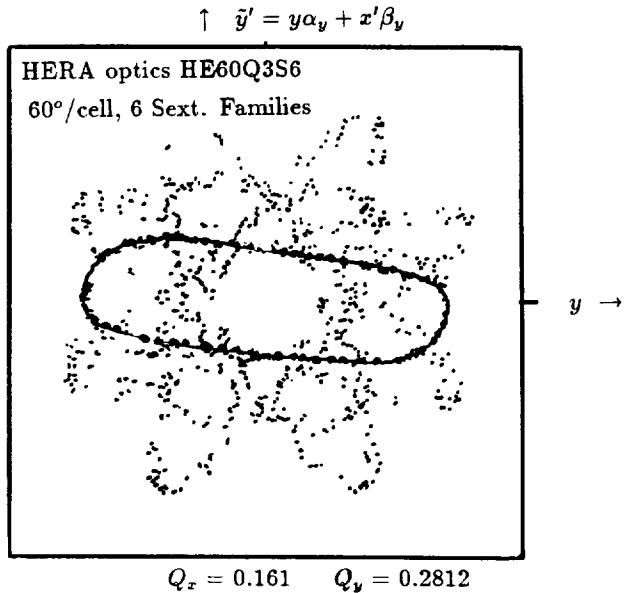


Figure 3: Vertical phase space trajectory close to the stability limit projected onto the  $\tilde{y}'$ – $y$ -plane for the HERA electron ring taking into account chromaticity compensating sextupole fields. The closed curve joins data points corresponding to  $\phi_x \simeq 0$ . The area enclosed by the curve is the effective vertical emittance. The dominant contribution to the phase space distortion are terms with  $\mu = 2, 4$  which are generated in second order by sextupoles. The full scale of the plot is 2.8 mm which is 1.9× the vertical rms beam size for full coupling.

The generalized emittances are adequate and physically motivated observables which are well suited to parameterize the stability limit ("dynamic aperture") of betatron oscillations. The dynamic aperture expressed in generalized emittances is less unambiguous than the values of the coordinates  $x$  and  $y$ . This is particularly helpful to avoid misinterpretation of tracking data in the case of strongly distorted vertical phase space for a flat electron beam (see Fig.3).

Quite often, one will observe that for a particle injected close to the stability limit, the stroboscopic observation will not be successful since no closed curve  $\epsilon_{x,y}(\phi_{x,y})$  will be obtained. This is the first hint that one is examining a nonregular orbit which is potentially unstable. Thus even the failure of the stroboscopic method provides useful information about the tracking data.

One may summarize that the concept of nonlinear distortions provides useful tools to display and to interpret particle tracking data.

#### **4.2 THE CONCEPT OF NONLINEAR RESONANCES**

The force experienced by a particle travelling around the accelerator has a strong "time" dependence imposed by the variation of the nonlinear fields along the closed orbit but also owing to the motion of the particle itself. (Note that this point of view implies a perturbation treatment starting from an unperturbed motion which is used to calculate the perturbations.) If a small nonlinear force has a frequency component close to the betatron frequency, this force stays in phase with the particle for quite a long period so that it can cause the particles to go unstable. This situation is called a nonlinear resonance <sup>4</sup>. Nonlinear resonances play a central role in the nonlinear dynamics of accelerator beams. In some cases, the resonant behaviour of the beam can be described analytically. We can take up the considerations of the previous

<sup>4</sup>One can convince oneself that a force  $F = \frac{m(s)}{2}x^2$  where  $m(s)$  is a periodic function which can be expanded in a Fourier series  $m(s) = \sum_q m_q \cos \frac{2\pi q}{L}s$  and where the unperturbed  $x$  is described by a betatron oscillation with the tune  $Q$ ,  $x = \sqrt{\epsilon\beta} \cos \frac{2\pi Q}{L}s$  has a component close to the tune if  $3Q = \text{integer}$  and  $Q = \text{integer}$

section in order to arrive at such a description. Remember that the generating function  $S$  eqn. 33 can become very large in case of a resonance  $\nu\tilde{Q} + q \simeq q$ , so that convergence of the perturbation series is expected to be very poor. Therefore we have to exclude the terms  $\frac{h_{n\nu q}}{\nu\tilde{Q}+q}$  with small denominators from the transformation, so that these terms will also be retained in the new Hamiltonian  $K$ . Let us consider the case of a single resonance  $\nu\tilde{Q} + q$ .

$$K = \frac{1}{2} \sum_n h_{n00} \cdot J^{n/2} + \sum_n h_{n\nu q} \cdot J^{n/2} \cos(\nu\varphi + (\nu Q + q)2\pi s/L + \varphi_{n\nu q}) + \text{higher order terms} \quad (38)$$

As long as they are non-resonant, the higher-order terms can be treated by the distortion concept described in the previous section. Thus they will be taken into account by the coordinate transformation  $\epsilon, \varphi \rightarrow \epsilon, \phi$  which will not interfere with treatment of the resonant terms. The detuning terms which also appear in the Hamiltonian will be neglected for the moment. The "time" dependence of  $K$  can be removed by introducing the coordinates  $\Psi = \varphi + (Q + q/\nu)\frac{2\pi s}{L} \equiv \varphi + \Delta \cdot s$  and  $I = J$  by a new canonical transformation (generated by a generating function  $F = I \cdot (\varphi + \Delta \cdot s)$ , yielding in the most simple case, ( only a single value of  $n$ ,  $\varphi_{n\nu q} = 0$ )

$$K = \Delta I + h_{n\nu q} I^{n/2} \cos \nu\Psi \quad (39)$$

The value of the Hamiltonian does not change during the motion of the particle, it is an invariant. By plotting  $K$  versus the action  $I$  for the two extreme values of the cosine, we can see under what circumstances one obtains stable and unstable motion (see Fig.4). The lower curve (for  $\cos \nu\Psi = -1$ ,  $\Psi = \pi/\nu$ ) has a maximum at  $I = I_0$ . Every initial value of  $I$  which is larger than  $I_0$  belongs to an unstable orbit. Initial values of  $I \leq I_0$  and of  $K \leq \Delta I_0 - h_{n\nu q} I_0^{n/2}$  belong to stable orbits. A corresponding two dimensional phase space plot for  $\nu = 3$ ,  $n = \text{odd}$  ( $\frac{1}{3}$ -integer resonance) is shown in Fig.5. The coordinates of the fixed point  $(K_0, I_0)$  are obtained by solving

$$\frac{\partial K}{\partial \Psi} = \frac{\partial K}{\partial I} = 0 \quad (40)$$

which yields in the case of a  $\frac{1}{3}$ -integer resonance driven by sextupoles ( $n = 3$ ,  $\nu = 3$ )

$$\begin{aligned} \cos 3\Psi = -1 (\Delta > 0) &\rightarrow \Psi_0 = k\pi/3, \quad k = 1, 3, 5 \\ \Delta - \frac{3}{2}\sqrt{I} h_{33q} = 0 &\rightarrow I_0 = (2\Delta/3h_{33q})^2 \quad K_0 = (4/27)(\Delta^3/h_{n\nu q}^2) \end{aligned} \quad (41)$$

The parameter  $\Delta$  is the distance of the tune from the resonance  $\Delta = Q - q/\nu$ . For a fixed value of the action  $I$ , there is a certain value of  $\Delta$  for which  $I = I_0(\Delta)$ . This value  $\Delta_0$  is called "width of the resonance". The coordinates  $\sqrt{I} = f(\Psi)$  of the separatrix are related by

$$K = \Delta I + h_{n\nu q} I^{n/2} \cos \nu\Psi = K_0 = \Delta I_0 - h_{n\nu q} I_0^{n/2}. \quad (42)$$

Higher dimensional problems are closer to reality in accelerators and they are more relevant for the interpretation of particle tracking results. Whereas the algebra necessary to derive the Hamiltonian and to calculate the resonance widths and separatrices is somewhat more complicated in higher dimensional systems, the basic physical ideas are the same. The procedure to obtain the Hamiltonian is exactly the same. The result for transverse motion (coasting beam) in the vicinity of a resonance  $\nu Q_x + \mu Q_y \simeq \text{integer}$  is

$$K = \Delta_x I_x + \Delta_y I_y + h_{mn\nu\mu q} I_x^{n/2} I_y^{m/2} \cos \nu\Psi_x + \mu\Psi_y \quad (43)$$

( $n + m$  multipole order,  $\nu = n, n-2, \dots$ ,  $\mu = m, m-2, \dots$ ) where the  $\Delta_{x,y}$  are defined by

$$\Delta_x = \frac{\nu}{\nu^2 + \mu^2} (\nu Q_x + \mu Q_y) \frac{2\pi}{L}; \quad \Delta_y = \frac{\mu}{\nu^2 + \mu^2} (\nu Q_x + \mu Q_y) \frac{2\pi}{L}. \quad (44)$$

As in the one dimensional case,  $K$  is invariant since it does not contain "time" explicitly. There is another invariant because

$$\frac{1}{\nu} \frac{\partial K}{\partial \Psi_x} = \frac{1}{\mu} \frac{\partial K}{\partial \Psi_y} \rightarrow \nu I_x - \mu I_y = \text{const} = C \quad (45)$$

We see that there are two classes of resonances in the two dimensional case.

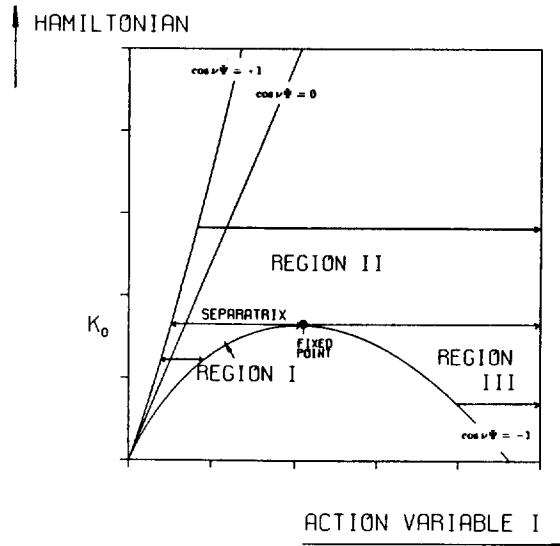


Figure 4: The value of the Hamiltonian  $K$  is plotted versus the action variable  $I$  for different values of  $\cos \nu \Psi$ . Particles move along horizontal lines with  $K = \text{const}$ . The motion is confined in region I thus for  $K < K_0$ ,  $I < I_0$ . In region II and III, the motion is unstable. The point  $K_0, I_0$  is called the fixed point. The orbit through the fixed point is called separatrix.

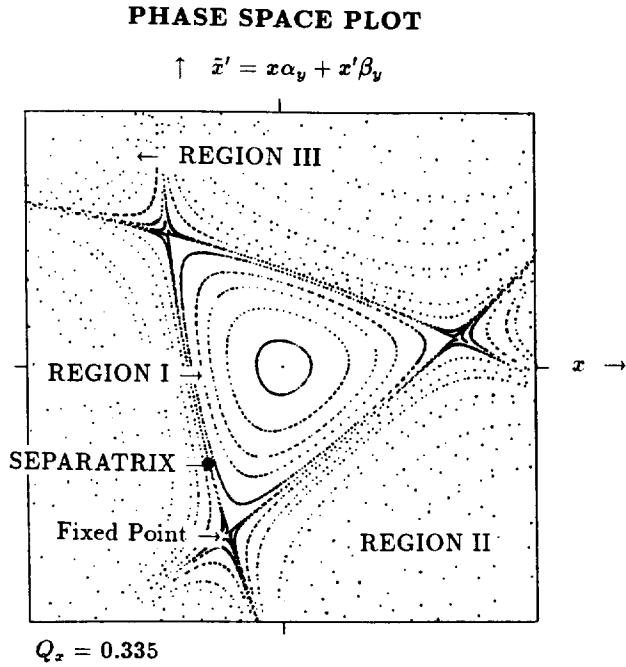


Figure 5: Nonlinear phase space plot  $\tilde{x}' = x\alpha + x'\beta$  versus  $x$  for horizontal betatron motion close to the  $\frac{1}{3}$  integer resonance ( $\nu = 3$ ). Indicated are the regions I,II,III of Fig.4, the separatrix and the fixed points. Note that the radius  $\sqrt{x^2 + (x\alpha + x'\beta)^2} = \sqrt{2}\epsilon\beta$  corresponds to  $\simeq \sqrt{2}I$  if one neglects higher-order distortions.

(a) The sign of  $\nu$  and  $\mu$  are equal. In this case, both  $I_x$  and  $I_y$  can become arbitrarily large without violating  $\mu I_x - \nu I_y = C$ . These resonances are called "sum" resonances. The motion in the vicinity of sum resonances is potentially unstable. For illustration see Fig.6a.

(b) If the signs of  $\nu$  and  $\mu$  are different, the motion is confined between two maximum values of  $I_x$  and  $I_y$  (see Fig.6b). In this case we are talking about a difference resonance. Difference resonances are stable.

The analytical calculation of the fixed points, resonance widths and separatrices can become very difficult because it involves finding the roots of high-order polynomials in two variables. The problem can be reduced to a one-dimensional one by eliminating  $I_x$  or  $I_y$ , respectively by using  $\mu I_x - \nu I_y = C = \mu I_x(s=0) - \nu I_y(s=0)$ . This gives two expressions for the fixed points

$$\frac{\partial K(I_x, C)}{\partial I_x} = 0 \rightarrow I_x^0(C); \quad \frac{\partial K(I_y, C)}{\partial I_y} = 0 \rightarrow I_y^0(C). \quad (46)$$

The constant  $C$  can be eliminated. This results in a separatrix in action space  $I_y^0 = f(I_x^0)$  which is independent from the initial conditions. The result for the resonance  $Q_x + 2Q_y \simeq \text{integer}$  is

$$I_y^0 = \frac{2}{9} \left( \frac{\Delta}{\kappa} \right)^2 \pm \frac{2}{3} \left( \frac{\Delta}{\kappa} \right) \sqrt{I_x^0 - 4I_x^0} \quad (47)$$

( $\kappa$  stands for  $h_{nm\nu\mu q}$ ,  $n = 1, m = 2, \nu = 1, \mu = 2$ ) which is illustrated in Fig.7.

The considerations show so far that resonances may drastically influence the performance of an accelerator. Detection of resonant behaviour is therefore essential for the interpretation of tracking results. For example it is important to recognize that a dynamic aperture reduction is caused by a strong single resonance, since the resonant instability can in principle be cured by cancelling the driving terms  $h_{nm\nu\mu q}$  which may require only two<sup>5</sup> correction elements.

The most convenient way to detect resonant behaviour is to plot the action variables in an "action plot". However, one cannot expect to observe clean resonant behaviour, as shown for the example in fig 6a,b. Since one cannot calculate  $I_x, I_y$  directly from the tracking data, one has to plot  $\varepsilon_x$  and  $\varepsilon_y$ , which are determined by the impact of both resonant and non-resonant, distortion terms. Therefore the motion in action space does not usually extend along a single line but is "smeared" out owing to the nonlinear distortions. Note that there is quite some arbitrariness about which contribution of the nonlinear force is considered to be "resonant" and which is considered to be "distortion". There is a continuous transition from one extreme to the other. Sometimes, the motion is influenced by more than one resonance. The corresponding phase space pattern is then more complicated. Both our models, the resonant and the distortion concept fail to describe this situation. However, in many cases the structures obtained in action space remind one of the motion in the vicinity of a single resonance. See for example Fig.8 where the action plot exhibits two slopes  $\partial I_y / \partial I_x = 2$  and  $\partial I_y / \partial I_x = 0$  corresponding to the two resonances  $Q_x + 2Q_y \simeq \text{integer}$  and  $3Q_x \simeq \text{integer}$  which govern the motion in this example.

Another way of detecting resonant behaviour is the stroboscopic method of displaying phase space projection which has been described in the previous section. By inserting equation 45 into equation 43 one can see that also in a resonant situation where the convergence of the perturbation series is not guaranteed there is a relationship between the action variable in one plane and the phase variables which allows to draw phase space contours in the stroboscopic way. Fig.9 shows an example of the motion near the difference resonance  $Q_x - 2Q_y = \text{integer}$ . If the phase space trajectories are projected onto the  $x - x'$  and  $y - y'$  plane respectively, there is no resonant structure visible. The stroboscopic method results in contours showing the characteristic phase space pattern of this resonance.

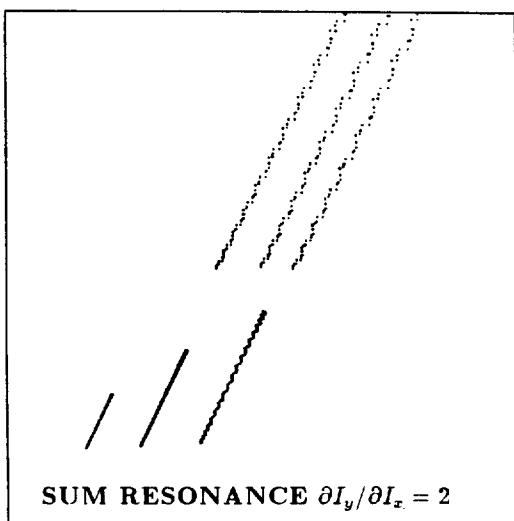
#### 4.3 THE CONCEPT OF DETUNING

In the previous section, we have neglected the detuning terms in the Hamiltonian  $\sum_n k_{n\nu q} J^{n/2}$ , which determine the amplitude dependence of the tunes since

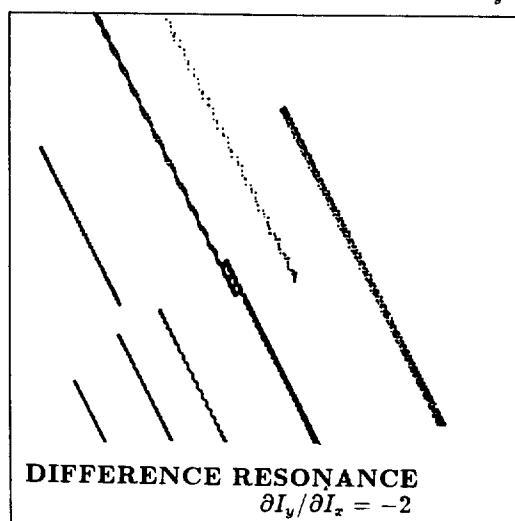
$$J' = -\partial K / \partial \varphi = 0 \rightarrow J = \text{const}; \quad \varphi' = \partial K / \partial J \rightarrow \varphi = f(J) \cdot s \rightarrow \Delta Q = \frac{L}{2\pi} f(J) \quad (48)$$

<sup>5</sup> $h_{nm\nu\mu q}$  is a complex number which can be eliminated by adding two orthogonal correction elements of proper strength at appropriate positions in the lattice.

↑ VERTICAL ACTION VARIABLE  $I_y$



↑ VERTICAL ACTION VARIABLE  $I_y$



HORIZONTAL ACTION VARIABLE  $I_x \rightarrow$

HORIZONTAL ACTION VARIABLE  $I_x \rightarrow$

Figure 6a: Projection of the motion in 4-dimensional phase space into the action space  $I_y$  versus  $I_x$  in the case of a sum resonance.

Figure 6b: Projection of the motion in 4-dimensional phase space into the action space  $I_y$  versus  $I_x$  in the case of a difference resonance.

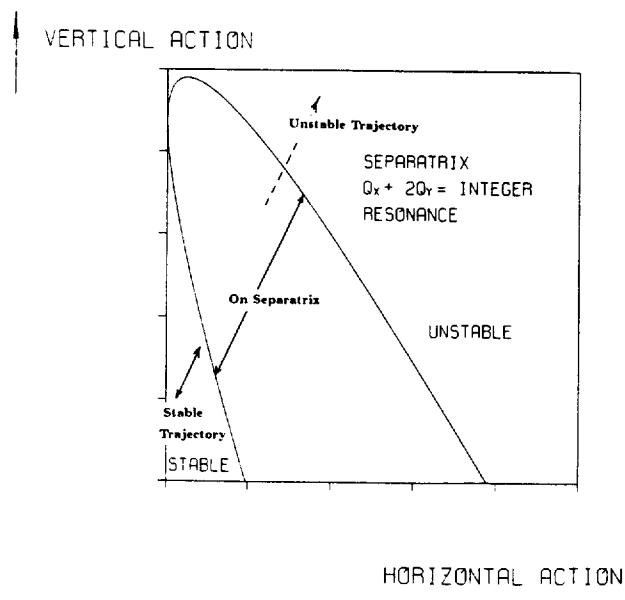


Figure 7: Separatrix for the resonance  $Q_x + 2Q_y \simeq \text{integer}$  projected onto action space  $I_x, I_y$ .

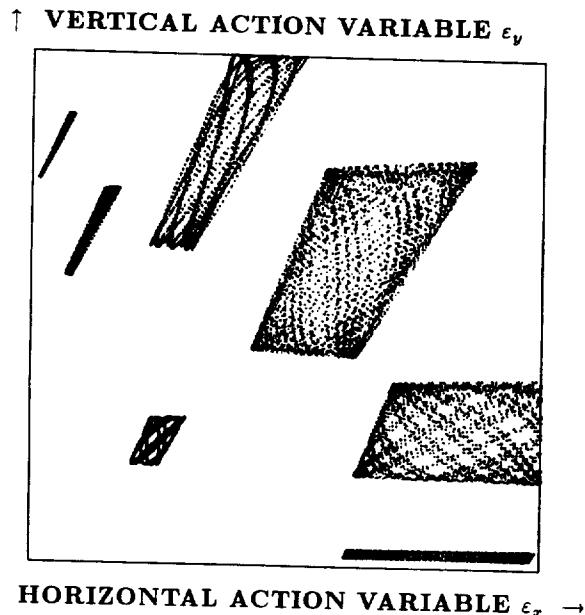


Figure 8: Action plot  $I_y$  versus  $I_x$  for the motion near the resonances  $Q_x + 2Q_y \simeq q$  and  $3Q_x \simeq q$ .

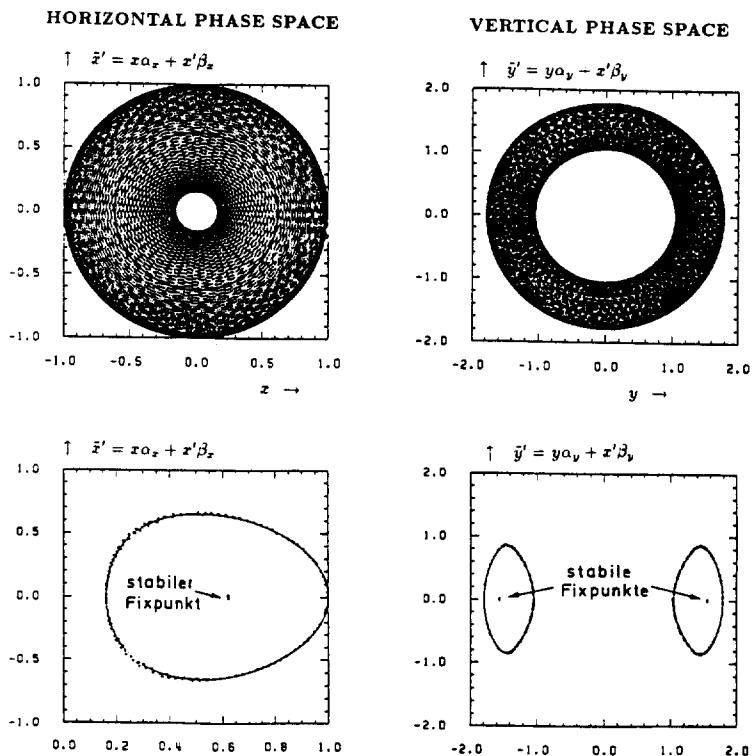


Figure 9: "Stroboscopic" representation of phase space pattern near the difference resonance  $Q_x - 2Q_y \simeq q$  which has been taken from reference[19].

Such terms are produced by nonlinear field components which are asymmetric with respect to the closed orbit, thus by quadrupole, octupole, duodecapole, ... fields. The horizontal tuneshift with amplitude due to an octupole component  $b_3$  of a dipole magnet of strength  $\frac{1}{\rho}$  measured at a radius  $r_0$  is

$$\Delta Q = \frac{1}{16\pi} \int_s^{s+L} ds \frac{b_3 \beta_x^2}{\rho r_0^3} \times J. \quad (49)$$

The detuning terms, if they are strong enough, have a drastic impact on the resonant behaviour. Consider a single resonance Hamiltonian with detuning terms (using  $I = J$  and  $\Psi = \phi + \Delta \cdot s$  as canonical variables again)

$$K = \Delta I + \sum_n h_{n00} I^{n/2} + h_{n\nu q} I^{n/2} \cos \nu \Psi. \quad (50)$$

If  $K$  is plotted versus  $I$  for  $\cos \nu \Psi \pm 1$ , it becomes apparent that the motion of the particles between the two extreme values of the cosine will be bounded in action (amplitude) everywhere if the detuning terms are strong enough (see Fig.10). For the case of a single detuning term and a single resonance term of the same order  $n$  there is the simple rule; the resonance is stabilized if the detuning term  $h_{n00}$  is larger than the resonance driving term  $h_{n\nu q}$ . The phase space plot shows a new feature near stabilized resonances. The separatrix branch which is departing from the unstable (hyperbolic) fixed point bends backwards after a while and approaches the next hyperbolic fixed point. The region enclosed is called a "resonance island". It comprises closed concentric phase space contours with a stable (elliptic) fixed point in the centre (see Fig.11). Similar features exist in higher dimensional phase spaces which become visible if the stroboscopic plotting method of phase space data is used.

The width of the island can be obtained by expanding the Hamiltonian around the separatrix up to second order in  $\Delta I = I - I_0$ . The maximum value for  $\Delta I$  is obtained for  $\cos \nu \Psi = +1$  which gives a resonance width of

$$\Delta I = \sqrt{\frac{4h_{n\nu q} I_0^{n/2}}{\partial Q(I_0)/\partial I}} \quad (51)$$

Note that the resonance island becomes larger if the resonance occurs at larger amplitudes. Since  $I_0$  is in the order of  $h_{n\nu q}^{2/(n-2)}$  (see eqn. 40) one finds that the island width is

$$\Delta I \sim \sqrt{I_0} \quad (52)$$

The amplitude dependent tunes and the detuning terms  $h_{n00}$  can be determined numerically from the tracking data. The tunes are defined as the phase advance  $\Delta\phi$  per revolution in units of  $2\pi$ . We can determine  $\Delta\phi = \left( \arctan \left( \frac{x_N \alpha + x'_N \beta}{x_N} \right) - \arctan \left( \frac{x_{N-1} \alpha + x'_{N-1} \beta}{x_N} \right) \right)$  directly from the tracking data of two successive turns  $N, N-1$ .  $\Delta\phi$ , also contains, besides the detuning effects, phase distortions  $\delta\phi$  described by

$$\delta\phi = \partial\sigma(J, \phi)/\partial J = \sum_{n\nu} \frac{n\sigma_{n\nu} J^{(n-2)/2} \sin(\nu\phi_x + \phi_{n\nu})}{2 \sin \pi\nu \bar{Q}_x} \quad (53)$$

(see eqn. 30). We conclude that we have to integrate over all phases  $\phi$  in order to average over the distortional phase variations  $\delta\phi$ . This can be done approximately by averaging over discrete values of  $\phi$  obtained on successive turns around the accelerator. Near a resonance, the particles are to be found, most of the time, near the hyperbolic fixed point, so that the corresponding phases are overemphasized (for the reasons discussed in section 4.1). It appears as if the particles would "lock on" to the resonance. The amplitude dependent tune can be calculated by

$$Q(J) = Q_0 + \frac{1}{2\pi N} \sum_{M=0}^{N-1} (\phi_{M+1} - \phi_M) \quad (54)$$

What remains to be determined are the corresponding action variables. These are the effective emittances discussed in section 4.1 obtained by calculating the emittance  $\epsilon$  from the particle coordinates and integrating it over the phase in order to average over the nonlinear distortions

$$J = \frac{1}{2\pi} \oint d\phi \epsilon(\phi). \quad (55)$$

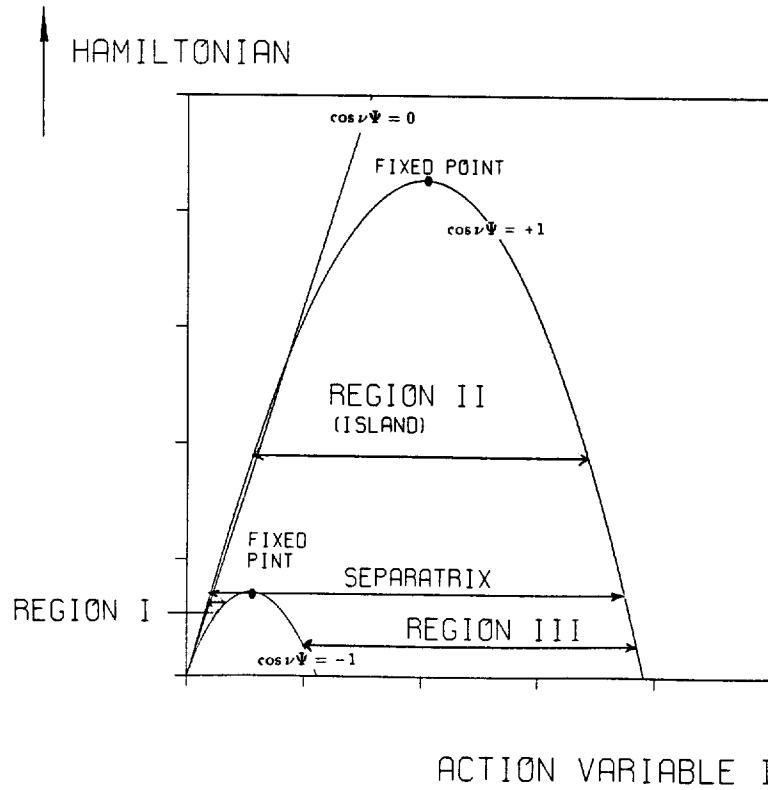


Figure 10: Stabilization of a nonlinear resonance by detuning terms. The value of the Hamiltonian  $K$  is plotted versus the action value  $I$  for three values of the cosine function  $\cos \nu \Psi \in [1, 0, -1]$ . For strong enough detuning, the motion is bounded everywhere.

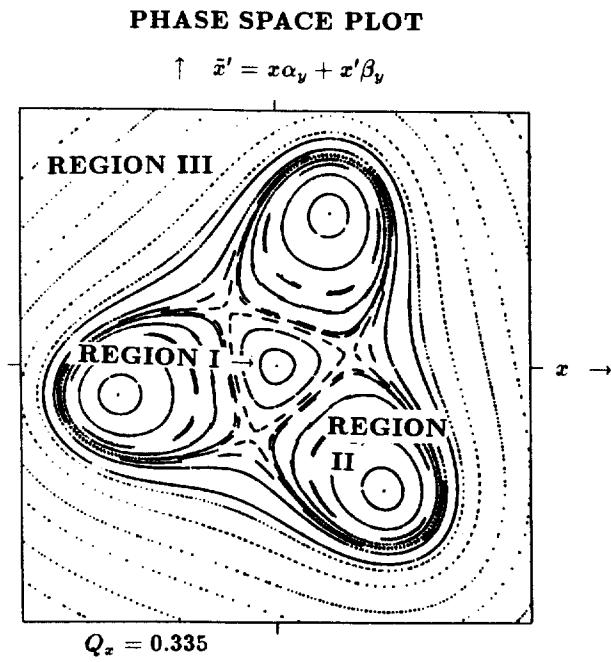


Figure 11: Phase space plot  $x\alpha + x'\beta$  versus  $x$  for the motion in the vicinity of the stabilized nonlinear resonance  $3Q = \text{integer}$ .

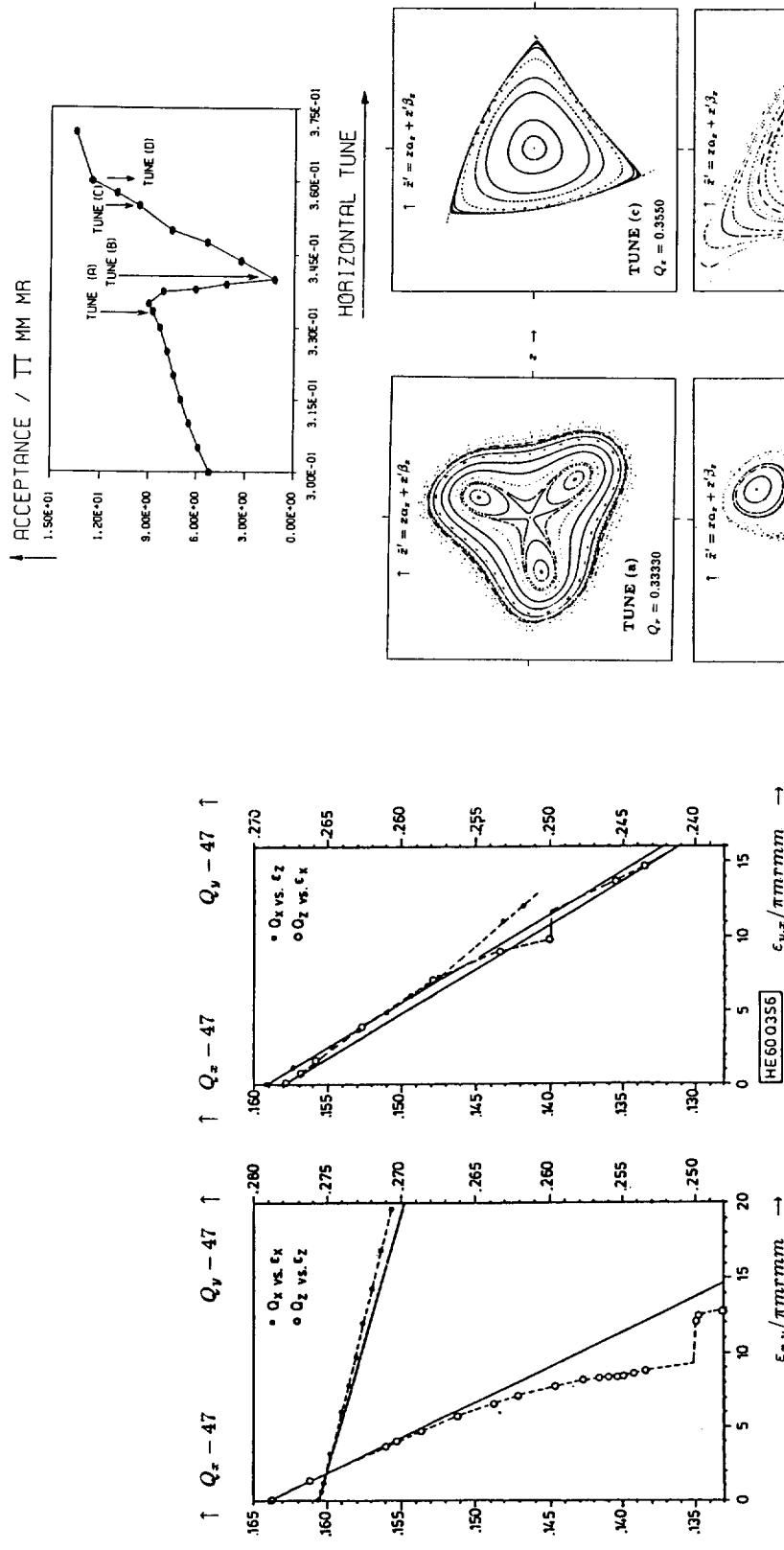


Figure 12: Amplitude dependent tunes in the HERA electron ring calculated analytically (solid lines), and from particle tracking (dots). The open dots are the vertical tunes, the full dots the vertical tunes. In the left figure  $Q_{z,y} = f(J_{z,y})$  is plotted, and in the right figure  $Q_{x,y} = f(J_{x,y})$  is plotted (Taken from ref [20]).

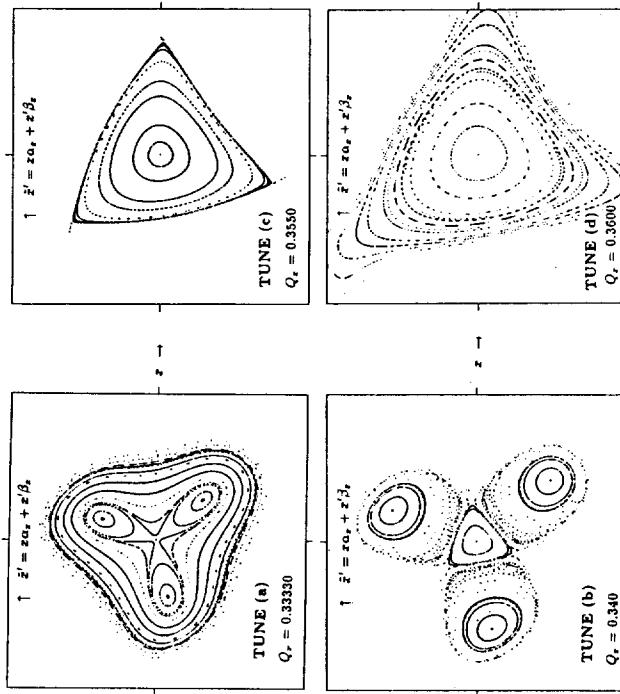


Figure 13: Mechanism of dynamic aperture reduction due to nonlinear resonances in the presence of a tuneshift with amplitude. The top figure is the dynamic aperture plotted versus the tune. The figures below are phase space plots  $\vec{x}' = x\alpha + z\beta$  versus  $x$  for motion with one degree of freedom (horizontal betatron motion) at selected tune values (see text).

In the case with higher degrees of freedom one has to use the stroboscopic method to obtain the effective emittance as discussed in section 4.1.

$$J_z = \frac{1}{2\pi} \oint d\phi_z \epsilon(\phi_x, \phi_y) = \text{const} \quad (56)$$

Figure 12 (taken from ref.[20]) shows that this procedure has been very successful in explaining tracking results for the HERA electron ring with 400 sextupoles for chromaticity correction. Strong detuning is generated as a second-order effect of sextupole fields which can be calculated analytically (solid lines Fig.12) (see for example [21] for the explicit formulae) and which agrees well with the analysis of the tracking data (dots in Fig.12). This detuning turned out to be one of the factors which determined the dynamic aperture in the HERA electron ring since the detuning pushes the tunes towards the integer resonance where the motion becomes unstable.

A tune scan of the dynamic aperture very often shows reductions of the maximum stable amplitude around certain linear tunes as sketched in Fig.13. Usually, the linear tunes at the acceptance minima are not close to a resonance. Detuning is the key to the understanding of this behaviour. Consider Fig.13. There are two dips of the dynamic aperture over the plotted tune range. Furthermore, there are schematic phase space plots for four different tunes around the dip.

- (a) The linear tunes are adjusted close to the resonance. The stabilized resonance island occurs at small amplitudes. At larger amplitudes the phase space is almost circular again. For large enough amplitudes, the nonlinearities are strong enough to cause the particles to go unstable.
- (b) The tunes are adjusted farther away from the resonance. The resonance now occurs at larger amplitudes — and the resonance islands become larger. If the separatrix touches the dynamic aperture, one finds that most of the trajectories inside the island become unstable.
- (c) The hyperbolic fixed points and the separatrix occur near the aperture limit. This marks the tune where one starts to see the influence of the resonance.
- (d) The dynamic aperture limit is restored. The resonant tune occurs at amplitudes which are larger than the dynamic aperture.

This interpretation of the results of a tune scan is of course only quantitative. No attempt is made at this point to explain the mechanism which determines the stability limit. Global dynamic aperture limitations cannot in general be explained by single resonance effects. Nevertheless one should exploit the information contained in such a tune scan. The width of the resonant island is in the order of the reduction of the dynamic aperture. The detuning is the (inverse) slope of the dynamic aperture as a function of the tune inside the dip. This allows a rough estimate of the strength of the resonance, since resonance strength  $h_{n\nu q}$ , island width  $\Delta I$ , and detuning  $\partial Q/\partial I$  are related by

$$h_{n\nu q} = \frac{(\Delta I)^2 \partial Q / \partial I}{4I_0^{n/2}}. \quad (57)$$

So far we have seen that detuning stabilizes resonances, thus preventing instabilities from occurring. But on the other hand it may push the particle tunes with increasing amplitudes towards distant strong resonances which can cause instability. There is another aspect of detuning. The distance of "island chains" measured in the values of the actions at which the resonances occur is inversely proportional to the detuning  $\partial Q/\partial I$ , the size of the island is inversely proportional to the square root of the detuning (eqn. 51). Thus increasing the strength of the detuning will cause separated island chains to touch. Chirikov has discovered that overlapping island chains coincid with the phenomenon of chaotic motion [22] which is found in many nonlinear systems. Chaotic regions in phase space are of great importance for the stability of accelerator beams which is the subject of the following section.

#### 4.4 THE CONCEPT OF CHAOTIC MOTION

The phase space of accelerator particles may be considered as composed of several domains. For oscillations with small amplitudes around the closed orbit the motion is almost linear. Phase space trajectories

are determined by the linear optics and nonlinear perturbation theory. With increasing amplitudes, the distortions become larger and chains of resonant islands appear. For large amplitudes, the motion is unstable, the value of the action or amplitude is growing quickly as a function of "time". Inbetween, near the stability limit, there is a region where phase space trajectories show an irregular complicated structure which fills a band in phase space. The motion in this domain is called chaotic, it is characteristic for nonlinear systems. A definition for a chaotic motion is that one can never collect enough information about the motion to make a prediction about its future. In other words, the rate of information obtained from performing measurements on the system never saturates. For accelerators, or to be more precise, for our nonlinear model of an accelerator, we know from experience that a particle which has been injected into the chaotic domain will very likely enter the domain of unbounded motion. Sometimes however, this may take quite a long time. For hadron beams, where there is no damping, the lifetime of particles in the chaotic domain will be very low. In the SPS for example, the lifetime of "chaotic particles" has been determined to be in the order of seconds. This is the conclusion drawn from both particle tracking and machine experiments [23]. Therefore, the boundary between regular and chaotic motion has been proposed as the dynamic aperture limit for hadron accelerators [24,19]. The scenario just described is shown in Fig.14.

It is of course only a coarse grain model. If one is more precise, chaotic motion is found everywhere in nonlinear phase space. Around the microscopic island chains (for every rational number of the tunes, there is a nonlinear, usually stabilized resonance) tiny chaotic layers exist. Due to the complexity of high dimensional ( $> 2$ ) phase space, these chaotic domains may be interconnected and connected with the unstable domain. Therefore, even a particle which is near the closed orbit may find its way out after a finite time and may become unstable. This is called "Arnold Diffusion". The occurrence of Arnold Diffusion has not yet been verified in simulations of particle motion in accelerators.

Since the main task of particle tracking is to find the dynamic aperture limitation, detecting chaotic behaviour is essential. There are several methods to detect chaos or, what is more complicated, to determine that the motion is non-chaotic or regular.

- (a) The straightforward approach is to track particles for a long enough time. This will not prove that the trajectories are non-chaotic and stable but it will exclude chaotic behaviour with a large probability. The procedure is limited by the impact of round off errors and consumption of computing time.
- (b) Strong chaos can be recognized just by inspection of the tracking data displayed in a phase space plot. Regular motion extends, in contrast to chaotic motion, over a region in phase space which has sharp borders (which may take many data points to show). This is still observable if the phase space data are projected onto subspaces  $(\tilde{x}', x)$ ,  $(y', y)$ . For chaotic motion, the points of a phase space plot appear to be irregularly scattered. (See Fig.15.)
- (c) It has already been mentioned in section 4.1 that the method of stroboscopic display of phase space data is only guaranteed to work if the perturbation series converges, thus only if a generating function  $S$  (eqn. 32) exists. The failure of the stroboscopic method is therefore a strong hint for a nonregular, chaotic trajectory.
- (d) Chirikov's criterion of overlapping island chains can, in principle, be calculated analytically. It turns out that one has to involve very high order contributions to driving and detuning terms in order to make a reliable prediction. Because of the size and complexity of accelerator lattices, this appears to be very impractical for the time being.
- (e) Another criterion for chaotic motion is the Lyapunov exponent [25]. It parameterizes the exponential growth rate of the distance of two, initially very close, phase space points. This has turned out to be a useful tool to detect the weak chaos which is found very close to the border of stability[19]. This method will be discussed in more detail below.
- (f) The Lyapunov exponent is the average growth rate of the distance of two phase space vectors  $d_N = |\tilde{z}_1(t_N), -\tilde{z}_2(t_N)|$ . It is defined by

$$\sigma = \lim_{d_0 \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{N} \ln \frac{d_N}{d_0} \quad (58)$$

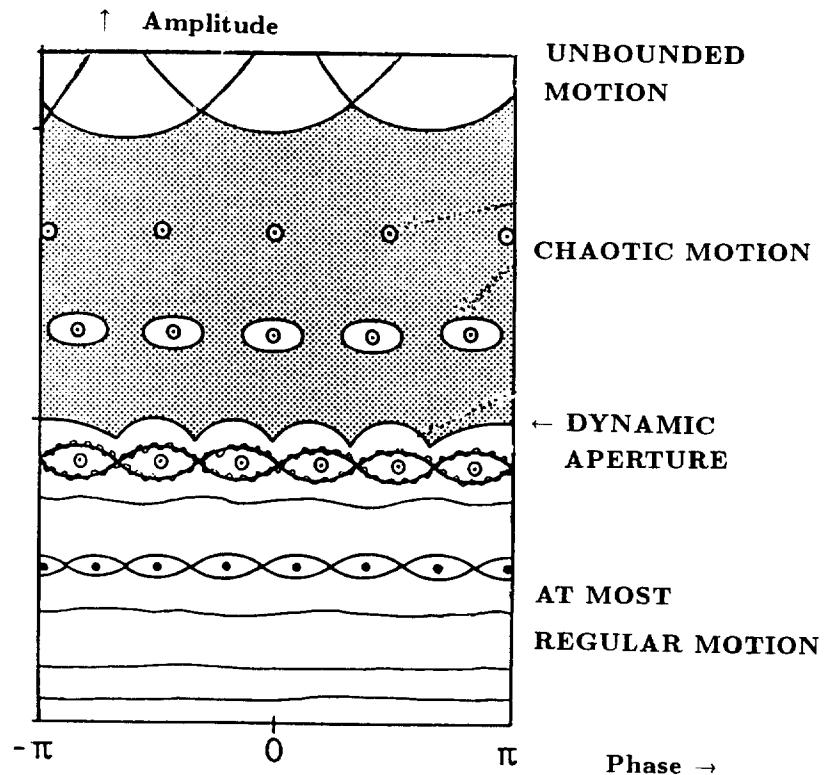


Figure 14: Dynamic aperture concept for hadron accelerators. Oscillation amplitude versus phase is plotted (schematically). Stable and unstable domains in phase space are separated by quasi-stable chaotic layers. The onset of chaos defines the dynamic aperture limit.

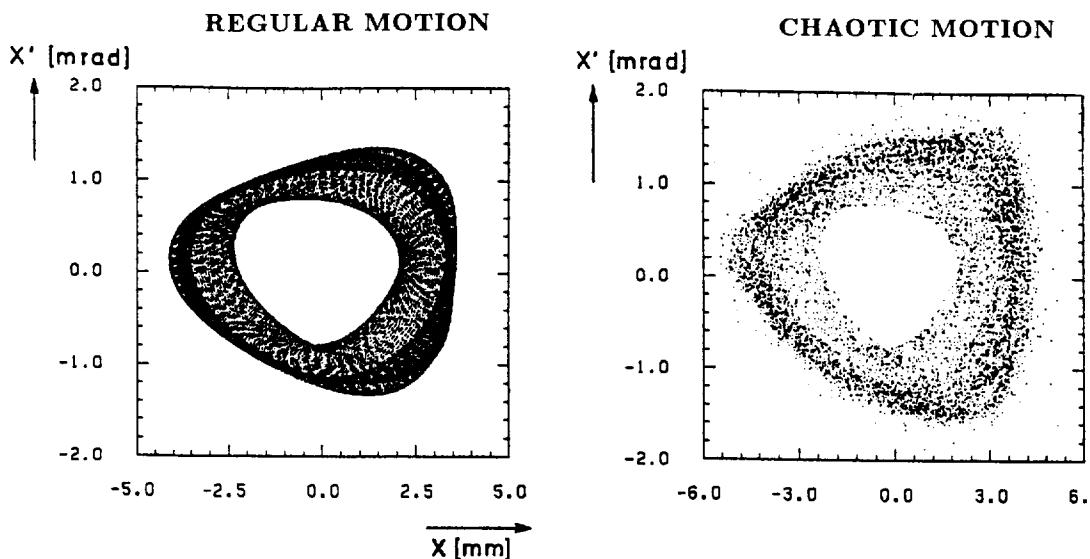


Figure 15: Strong Chaos is recognized by irregularly scattered points in a phase space plot.

Regular trajectories are known to produce zero Lyapunov exponents. If the Lyapunov criterion is different from zero, motion is chaotic. A pragmatic method has been developed to determine the dynamic aperture in the HERA proton ring [19]. The criterion is tested by injecting two particles as close as possible in phase space. The minimum distance is limited by round-off errors and is well above the resolution of the digital numbers in the computer. The distance in phase space is recorded turn by turn and plotted versus number of turns in a double logarithmic plot. Any non-exponential evolution of  $d_N$  will result in a linear divergence in this plot, whereas exponential growth still appears exponential. Problems of the method are large fluctuations of the distance on a short time scale which make it difficult to distinguish between non-exponential and exponential growth. This can be overcome by appropriate averaging and binning. The method turned out to be useful for practical investigation of the dynamic aperture limitation in the superconducting HERA proton ring. Fig.16 shows examples for the evolution of phase space distances for regular and chaotic motion.

Another interpretation of the Lyapunov exponent is that it measures the density of linearly unstable regions in the vicinity of the trajectory considered. This is motivated by the following consideration. The phase space distance  $d_N$  for  $d_0 \rightarrow 0$  can be rewritten as

$$\lim_{d_0 \rightarrow 0} d_N = \lim_{d_0 \rightarrow 0} |\Delta \vec{z}_N| = |\underline{J}_{(0,N)} \Delta \vec{z}_0|. \quad (59)$$

$\underline{J}_{(0,N)}$  is the Jacobian of the mapping  $N$ -times around the accelerator. Since  $\underline{J}_{(0,N)}$  is a symplectic map it is the product of the single turn Jacobians ( $\underline{J}_{(K-1,K)}$ ,  $K \in [1, N]$ ) the  $N$ -turn Jacobian can be calculated successively

$$\underline{J}_{(0,N)} = \prod_{K=1}^N \underline{J}_{(K-1,K)} \quad (60)$$

from turn to turn. Suppose that the initial vector  $\Delta \vec{z}_0$  is expressed in the base generated by the normalized eigenvectors  $\vec{e}_N^i$  ( $\vec{e}_N^{i+} \underline{S} \vec{e}_N^j = 2i\delta_{ij}$ ,  $\underline{S}$  = symplectic form) of  $\underline{J}_{(0,N)}$

$$\Delta \vec{z}_0 = \sum_i \alpha_{iN} \vec{e}_N^i. \quad (61)$$

Then we obtain

$$\underline{J}_{(0,N)} \Delta \vec{z}_0 = \sum_i \alpha_{iN} \lambda_N^i \vec{e}_N^i. \quad (62)$$

( $\lambda_N^i$  is an eigenvalue of  $\underline{J}_{(0,N)}$ ,  $\underline{J}_{(0,N)} \vec{e}_N^i = \lambda_N^i \vec{e}_N^i$ ) It can be shown [26] that the Lyapunov exponent is independent from a particular definition of a norm. A convenient choice for the norm of phase space vectors is  $|\Delta \vec{z}| = \sqrt{\Delta \vec{z}^+ S \Delta \vec{z}}$ . Eventually the Lyapunov exponent is expressed by

$$\sigma = \lim_{N \rightarrow \infty} \sigma_N = \lim_{N \rightarrow \infty} \frac{1}{N} \ln \sqrt{\frac{\sum_i \alpha_{iN}^* \alpha_{iN} \lambda_N^{i*} \lambda_N^i}{\alpha_{iN}^* \alpha_{iN}}}. \quad (63)$$

If the motion around the trajectory under consideration is linearly unstable, which is the case at a hyperbolic fixed point <sup>6</sup>, the absolute value of at least one pair of eigenvalues  $\lambda_N^i$  will become greater than one. A frequent occurrence of unstable fixed points during the motion leads to a non-vanishing Lyapunov exponent as can be seen from eqn.63.

The Lyapunov exponent can be successively calculated during the tracking run. Besides tracking the particle coordinates, the code must provide a one turn Jacobian matrix [27]. Note that a code like RACETRACK already contains this capability in form of a routine which is able to calculate the linear optics expanded around the closed orbit (and any other trajectory). Fig.17 shows regular and chaotic trajectories and the corresponding evolution of the Lyapunov coefficient. If the motion is regular, the coefficient  $\sigma_N$  ( $\sigma = \lim_{N \rightarrow \infty} \sigma_N$ ) decreases linearly in a double logarithmic plot. If the motion is chaotic,  $\sigma_N$  does not decrease linearly but evolves in a non-monotonous manner (note the somewhat different scales in Fig.18).

<sup>6</sup>If the nonlinear motion is expanded about an hyperbolic (unstable) fixed point, the corresponding linear motion is unstable, thus the linear transfer matrix associated with the expansion has eigenvalues with an absolute value greater than one.

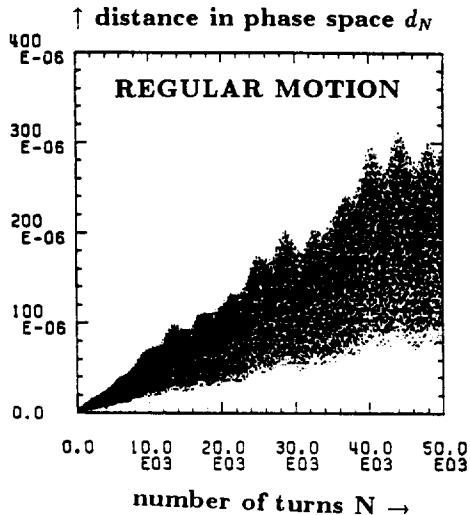


Figure 16a: Distance  $d_N$  of two initially close trajectories in phase space as a function of completed turns around the accelerator  $N$ .  $\ln d_N$  is plotted versus  $\ln N$ . For a regular trajectory one obtains a linear growth.

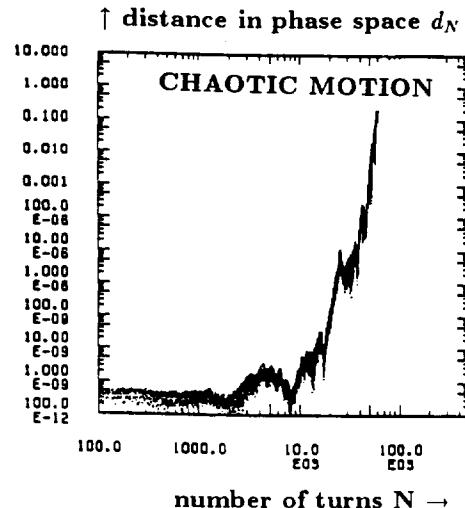


Figure 16b: Distance  $d_N$  of two initially close trajectories in phase space as a function of completed turns around the accelerator  $N$ .  $\ln d_N$  is plotted versus  $\ln N$ . For a chaotic trajectory one obtains an exponential growth.

## **5. CLOSING REMARKS**

This lecture about interpretation of particle tracking data is by far incomplete. The aim of the author was to show by examples how the nonlinear theory of particle motion in accelerators (which is also a separate lecture in this course) can be used for practical work on accelerator design and performance.

In order to minimize the formal overhead, the examples in this lecture have been chosen from motion in one and two degrees of freedom. This corresponds to transverse motion of a coasting beam around the accelerator. All the concepts, are at least in principle, also valid for the motion which also includes momentum oscillations. (For a nonlinear theory of synchro-betatron motion of linearly coupled systems see for example references[28,29]). Some of the methods such as the stroboscopic method may become impractical since too many data points are required.

The field of numerical simulation of particle motion in accelerators is evolving quickly. There are interesting new techniques using Lie algebra, differential algebra and normal forms which have a large potential for improving the understanding of the nonlinear dynamics in accelerators. Comprehensive discussion of these subject is outside the scope of this lecture.

## **ACKNOWLEDGEMENT**

The author wishes to thank Leo Michelotti and Rodney Gerig from Fermilab, Frank Schmidt from CERN and Gerhard Ripken, Helmut Mais and Reinhard Brinkmann from DESY for many valuable and interesting discussions. He is especially grateful to Susan Wipf for a careful revision of the manuscript.

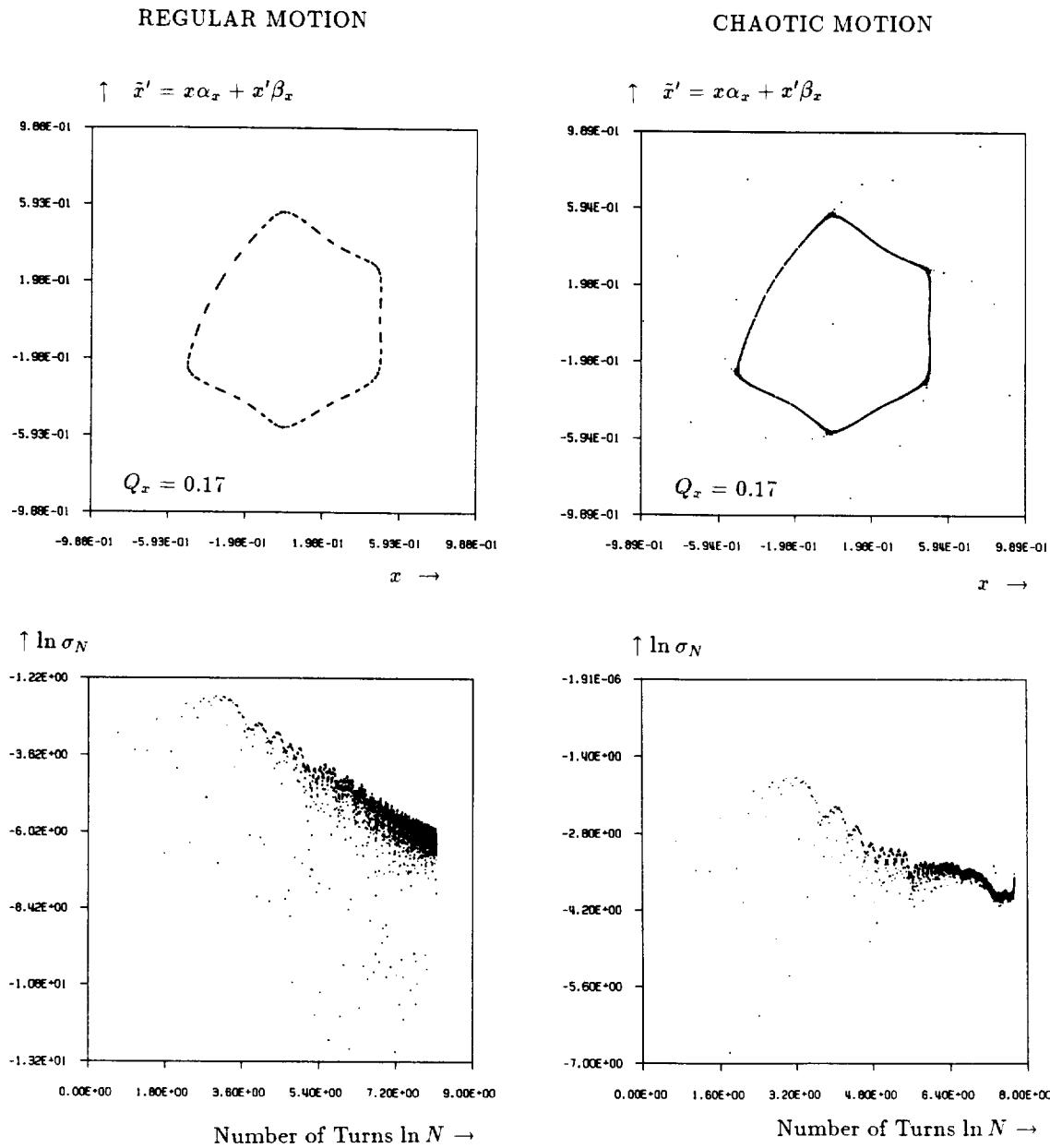


Figure 17: Evolution of  $\sigma_N$  over many turns for a Regular case (left part) and a chaotic case (right part).

## References

- [1] G. Ripken, F. Willeke: "Methods of Beam Optics", DESY 88-114 (1988)
- [2] A. Wrulich: "Particle Tracking for HERA", Accelerator Physics Issues for a Superconducting Super Collider (M. Tigner ed), Ann Arbor(1983)
- [3] A. Wrulich: "RACETRACK, a Computer Code for the Simulation of Nonlinear Particle Motion in Accelerators", DESY 84-026 (1984)
- [4] PATRICIA was originally written by H. Wiedemann. More recent descriptions are to be found in:  
S. Keifets: "Tracking Studies in PEP and Description of the Computer Code PATRICIA", SLAC-PUB-2922, (1982)  
G.F. Dell: "Studies of the Chromatic Properties and Dynamic Aperture of the BNL Colliding Beam Accelerator", IEEE Trans NS-30 (1985) 2469
- [5] G. Ripken: "Non-linear canonical equations of coupled synchrotron motion and their solution within the framework of a non-linear 6-dimensional (symplectic) tracking program for ultra relativistic protons"; DESY M-85-84 (1985)  
G. Ripken, D. P. Barber, F. Schmidt: "A Non-Linear Canonical Formalism for the Coupled Synchro-Betatron Motion of Protons with Arbitrary Energy"; DESY 87-036 (1987)
- [6] L.Schachinger, R. Talman: "Teapot: A Thin-Element Accelerator Program for Optics and Tracking", Particle Accelerators, Vol 22 pp 35-56 (1987)
- [7] F. Schmidt: "Tracking Results with SIXTRACK for the 1988 Dynamical Aperture Experiment"; CERN SPS/88-49 (AMS) (1988)
- [8] M. Berz: "A Survey of Differential Algebra and its Use for the Extraction of Maps to Arbitrary Order", SSC-166 (1988)
- [9] E. Forest, M. Berz and J. Irwin: "Normal Form Methods for Complete Solution Using Differential Algebra and Lie Operators"; SSC-166 (1988)
- [10] L. Michelotti: "Differential Algebra without Differentials"; Proceedings of the US Particle Accelerator Conference, Chicago (1989)
- [11] for a review see:  
A. J. Dragt, F. Neri, G. Rangarajan, D.R. Douglas, L.M. Healy, R. D. Ryne: "Lie Algebraic Treatment of Linear and Nonlinear Beam Dynamics", Ann. Rev. Nucl. Part. Sci. 38, 455-496 (1988)
- [12] A. J. Dragt: "Lectures in Nonlinear Orbit Dynamics"; Fermilab Summer School on High Energy Particle Accelerators, AIP Conference Proceedings, vol 87 (1982)
- [13] F. Neri: "Notes on Symplectification through Fifth Order", Dept.Phys. Tech. Rep., University of Maryland (1986)
- [14] A. J. Dragt, L.M. Healy, F. Neri, R. D. Ryne D.R. Douglas, E. Forest: "MARILIE 3.0, A Program for Nonlinear Analysis of Accelerator And Beamline Lattices" IEEE Trans Nucl. Sci, Vol NS-32 No 5, (1985)
- [15] see for example E. Wilson, this course  
There is a joint US-European Accelerator School which was held in Sardinia in 1985. The proceedings contain a comprehensive introduction to nonlinear accelerator theory: M. Month and S. Turner (ed): "Non-Linear Dynamics Aspects of Particle Accelerators", Springer Lecture Note in Physics, Sardinia (1985)  
see also the various issues of the proceedings of the CERN Accelerator School and the US High Energy Particle Accelerator Schools, AIP Conference Proceedings

- [16] There are a large number of internal laboratory reports, see for example:  
A. Schoch: "Theory of Linear and Non-Linear Perturbations of Betatron Oscillations in Alternating Gradient Synchrotrons", CERN 57-21 (1958)  
R. Hagedorn: "Stability and Amplitude Range of Two-Dimensional Non-linear Oscillations with Periodic Hamiltonian Applied to Non-Linear Betatron Oscillations of Circular Particle Accelerators", CERN 57-1 (1957)  
G. Guignard: "A General Treatment of Resonances in Accelerators", CERN MA-78-11 (1978)
- [17] H. Poincaré: "Les Méthodes Nouvelles de la Méchanique Céleste", Gauthier-Villars, Paris (1893)
- [18] V. I. Arnold: "Small Denominators and Problems of Stability in Classical and Celestial Mechanics I, II, III", Part I: Akad. Nauk. SSR Ser Mat 25 (1961) 21-86; Part II: Russ Math Surv 18/5 (1963); Part III: Russ Math Surv 18/6 (1963)
- [19] F. Schmidt, Thesis, University of Hamburg, DESY HERA 88-02 (1988)
- [20] R. Brinkmann, F. Willeke: "Chromatic Corrections and Dynamic Dynamic Aperture in the HERA Electron Ring I", DESY 86-079 (1986)
- [21] F. Willeke: "Analytical Study of the TEVATRON Nonlinear Dynamics", FERMILAB FN-422 (1985)
- [22] B.V. Chirikov: "A Universal Instability of Many-Dimensional Oscillator Systems", Physics Reports 52, No 5 (1979)
- [23] J. Gareyte, A. Hilaire, F. Schmidt: "Dynamic Aperture and Long-Term Particle Stability in the Presence of Strong Sextupoles in the CERN SPS", CERN SPS/89-2 (AMS)
- [24] H. Mais, G. Ripken, A. Wrulich, F. Schmidt: "Particle Tracking", Lecture given in the CERN Accelerator School, Oxford 198, CERN 87-03 (1987)
- [25] A.M. Lyapunov: "Stability of Motion". Academic Press New York (1966)
- [26] See for example  
A.J. Lichtenberg, M.A. Lieberman: "Regular and Stochastic Motion", Springer Verlag (1983)
- [27] This procedure has been implemented by L. Michelotti and R. Gerig from Fermilab in a tracking code named "TEVLAT". The procedure has been used successfully for FERMILAB main ring and SSC simulations.
- [28] D. Barber, H. Mais, G. Ripken, F. Willeke: "Nonlinear Theory of Coupled Synchro-Betatron Motion", DESY 86-147 (1986)
- [29] G. Ripken, F. Willeke: "The Impact of Linear Coupling on Nonlinear Dynamics", DESY 90-001 (1990)



## INTERNAL TARGETS

C. Ekström

The Svedberg Laboratory, Uppsala, Sweden

### ABSTRACT

Storage-ring operation with the combination of thin internal targets and phase-space cooling of the stored beams allows high-precision experiments to be made under very clean conditions. Here we will discuss the advantages and drawbacks of the internal-target systems currently in use or under development.

### 1. INTRODUCTION

The operation of internal targets in multi-pass experiments dates back to the seventies when gas-jet targets were used in the high-energy accelerators at Serpukhov and Fermilab [1-3]. These targets, later also used at Brookhaven [4], were pulsed and produced strong pressure bumps close to the target. In the late seventies, the continuous operation of thin internal targets, compatible with the vacuum requirements in storage rings, was investigated at the CERN accelerators LEAR, ISR and SPS. The proposals at ISR [5] and SPS [6] in 1980 led to the construction and installation of so-called cluster-jet targets for high-energy physics experiments. The early plans for an internal target at LEAR and a thorough discussion of the merits of internal-target operation in conjunction with phase-space cooling of the stored beams may be found in the report by Kilian et al. [7]. By this time, Garçon et al. [8] had installed a cluster-jet target at the Saturne accelerator at Saclay. A light-ion storage ring with internal targets and electron cooling, intended mainly for intermediate-energy physics experiments, was proposed in 1980 by Pollock et al. [9] at IUCF, Indiana. This Indiana Cooler ring has been followed by a large number of cooler storage-ring projects for internal-target experiments with light and heavy ions, i.e. at Aarhus, Darmstadt, Heidelberg, Jülich, Osaka, Stockholm, Tokyo and Uppsala. There is now a rapidly growing interest in using the technique of internal targets also at electron accelerators and stretcher rings (see e.g. Ref. [10]).

The use of internal targets in storage rings has been discussed at several conferences and workshops [10-14], to which here frequent references will be given. An extensive review on the possibilities for internal targets at the Indiana Cooler ring has been presented by Meyer [15].

Following some general physics-experiment considerations we will here discuss the requirements on internal targets in storage-ring environment, and in some detail the techniques and advantages of gas-jet targets, polarized targets and solid targets.

### 2. PHYSICS EXPERIMENTS

When building accelerators and targets it is important to remember that the main goal is to optimize the conditions for the *physics experiments* to be made. The figure of merit for physics experiments is usually the so-called luminosity, i.e. the target thickness (number of target atoms per  $\text{cm}^2$ ) times the particle-beam intensity (number of projectiles passing through the target per second). The total reaction rate in the target is then given by

the product of the luminosity and the total cross section for all reactions. In a given experiment, one is normally interested in only one specific channel. The detected count rate for this channel is obtained as the product between luminosity, differential cross section, detector efficiency and the solid angle covered by the detector. The quality of the experiment is, however, not only given by the count rate, but is very much dependent on other factors like background from other sources, energy resolution and angular resolution. These factors become of increasing importance in high-sensitivity experiments like on rare reactions and rare decays, and in high-precision experiments. So, in designing an experiment, one has to consider seriously these factors in addition to the general figure of merit, the luminosity.

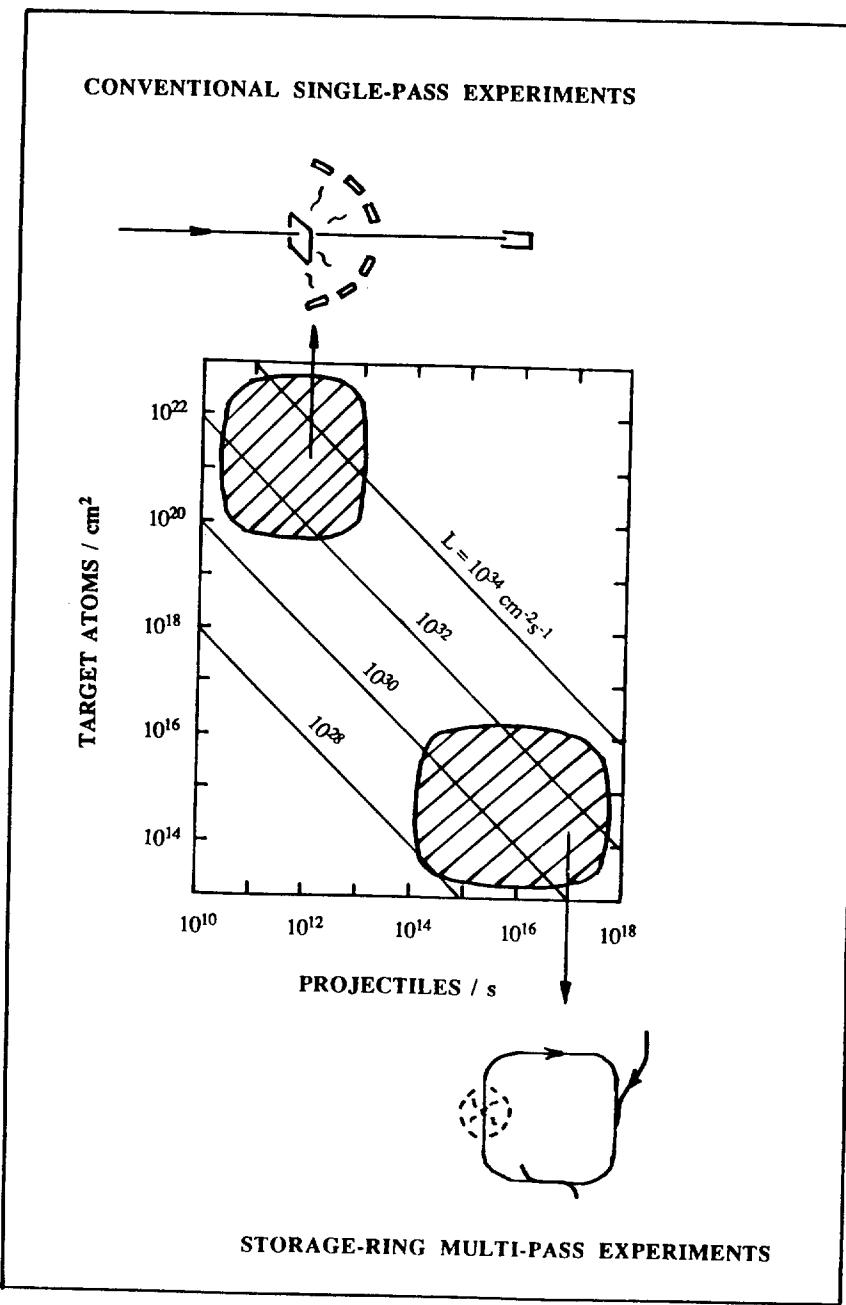


Fig. 1 Target thickness - beam intensity diagram showing the regions for conventional single-pass experiments and storage-ring multi-pass experiments. Lines of equal luminosity are indicated.

### 3. COMPARISON BETWEEN CONVENTIONAL AND INTERNAL-TARGET EXPERIMENTS

It is of interest to compare conventional single-pass experiments with storage-ring multi-pass experiments. In the diagram shown in Fig. 1 is given the target thickness as a function of particle-beam intensity. Lines of equal luminosity are indicated. The conventional single-pass experiments are made with extracted beams of  $10^{10}$  -  $10^{13}$  projectiles/s, i.e. currents of 1.6 nA - 1.6  $\mu$ A singly-charged particles, from the accelerator passing through a rather thick target of  $10^{20}$  -  $10^{23}$  atoms/cm<sup>2</sup>, i.e. 0.16 mg/cm<sup>2</sup> - 0.16 g/cm<sup>2</sup> single-nucleon target thickness. We are thus in the upper left corner of the diagram with luminosities in the range  $10^{30}$  -  $10^{35}$  cm<sup>-2</sup>s<sup>-1</sup>. Comparable or somewhat lower luminosities may be obtained in storage-ring multi-pass experiments with a stored beam interacting with an internal target. The intensity of the stored beam is given by the number of projectiles times the revolution frequency which is typically 1 MHz in small rings. Phase-space cooling is generally included to balance the beam heating and to improve on the beam properties. In the lower left corner the combination of beam intensity and target thickness gives too low luminosity for a realistic experiment, whereas in the upper right corner, these high intensities are not available for conventional experiments, and the high target thickness would directly kill the circulating beam in a storage-ring experiment.

The comparison shows that the two types of experiments are working at about the same luminosity but at quite separate regions of the target-thickness beam-intensity diagram. The storage-ring operation is, however, considered to be superior to the conventional experiments in several respects.

#### *Efficient use of the particle beams from the accelerator.*

Storage-ring experiments allow parallel operation since only a fraction of the beam from the accelerator is used, injected periodically into the ring. The particles make typically 1 million revolutions per second in small rings, and since the target is extremely thin, particles which do not interact strongly in the target or are scattered out of the ring acceptance, are recirculated and used again, maybe after cooling to retain their beam properties.

#### *High energy and angular resolutions.*

The high energy resolution of the cooled particle beams may be fully utilized in the experiments thanks to the small target thickness. The angular resolution profits from the small, well defined intersection point between particle beam and target.

#### *Undisturbed reaction products.*

The reaction products may be detected more or less without energy loss in the thin target. The detectors are placed either inside the scattering chamber or outside behind thin foil windows.

#### *No background from target container or beam dump.*

This is evident, since neither target container nor beam dump are required when using internal targets in a storage ring.

#### *Possibility for use of special targets.*

Here can be mentioned atomic-beam targets of almost 100% polarization.

#### *High luminosities and count rates.*

As shown above, the luminosities in storage-ring experiments are comparable or somewhat lower than those in conventional experiments. For a given energy resolution they are of course superior.

#### 4. REQUIREMENTS ON INTERNAL TARGETS

The internal targets to be used in storage-ring experiments have to fulfil a number of requirements. They should be thick enough to give sufficient count rate, and at the same time thin enough not to degrade the circulating beam too rapidly. They should also be thin enough to allow heavy recoil products to escape and be detected. From the experimental point of view, a wide choice of target material is desirable. To reduce the risk of background from non-target material, like narrow collimators, the region around the target should be wide enough. The disturbance to the ring vacuum should of course be kept at a minimum. For the detector arrangements, a large solid angle should be available around the target. Finally, most experiments require a well defined reaction vertex.

All these requirements are of course impossible to meet at the same time, and in building different target systems, a trade-off between different requirements is common, in order to meet the needs of a specific experiment.

#### 5. CHOICE OF INTERNAL TARGETS

There is a wide range of potential internal targets to be used in storage-ring experiments. One should, however, remember that it is only the gas-jet targets at the high-energy machines at Serpukhov, Fermilab, CERN, Saclay and Brookhaven that have been used in regular data-taking experiments. The target systems at the smaller storage rings around the world are at the same stage as the rings themselves, either under construction or in the running-in phase.

The internal targets may be divided schematically into gas targets and solid targets. The main advantage with gas targets is that the target beams are well matched to the ion beams, both in dimensions and thickness, whereas the disadvantage is the pressure bump around the target beam. The solid targets are locally very thick, but they occupy on the other hand only a small fraction of the area of the stored beam. Limited beam lifetimes and low duty cycles are seen as the main drawbacks in using these targets.

There are three types of molecular-beam targets. If we assume a distance of 25 cm between the nozzle and intersection point between target beam and circulating beam, to allow for the differential pumping, we may reach a target thickness of the order  $10^{10}$  atoms/cm<sup>2</sup> in the case of molecular effusion through a nozzle. At higher input pressures we get a supersonic gas jet which at room temperature reaches thicknesses of  $10^{13}$  atoms/cm<sup>2</sup>. Another factor of 10 may be obtained in the so-called cluster jet, obtained by cooling the nozzle.

Polarized atomic-beam targets may be obtained for hydrogen and deuterium using multipole magnets and rf-transitions, and for <sup>3</sup>He and the alkali metals using laser light. In optimized systems, target thicknesses of  $10^{13}$  atoms/cm<sup>2</sup> are foreseen. To increase further the target thickness a factor of 10, storage cells for the gas are under development. Special attention is given to the wall material, the temperature and the openings of the storage cells.

The solid targets to be used in storage-ring experiments may be of different materials and forms. We have thin fibers or micro-ribbons mounted on some fork-structure and introduced into the circulating beam. Microparticles may be enclosed in a cell, like in the case of gas storage-cell, or prepared in the form of a beam, crossing the stored beam and collected in a beam dump. Frozen droplets, so-called pellets, may be seen as a special case of the microparticle-beam target. The solid targets are, as mentioned above, locally very thick, but the

*effective target thickness*, taking into account the area of the stored beam occupied by the target, may be adjusted to optimum values of the order  $10^{16}$  atoms/cm<sup>2</sup>. In fact, a main problem is to come down to this effective value without reducing the duty cycle by keeping the target outside the beam part of the time.

The techniques and advantages of the different target systems will be discussed briefly below, starting with the gas-jet targets.

## 6. GAS-JET TARGETS

A review on the use of gas-jet targets in storage rings was given by Macri [16] at the CERN Accelerator School at Geneva in 1983. Various source configurations for beam production were discussed as well as the implementation of gas-jet targets in storage rings and the influence on stored beams. As examples, the cluster-jet targets at ISR [5,17] and SPS [6,18] at CERN were described in some detail. Here, we will present briefly some developments made and systems put into operation since 1983.

### 6.1 Supersonic gas-jet targets

Supersonic gas-jet targets are being used for experiments at extracted beams at a number of low-energy accelerators. At a target workshop at Jülich in 1987 [14], several groups (eg. Gaul et al [19] at Münster) presented their systems and the potential use in storage rings. In the present set-ups, different gases may be pressed through a tube and nozzle, passing a skimmer and being collected in a catcher. There is a very narrow space for the ion beam to pass between the skimmer and catcher, and the detectors occupy only a small solid angle in the horizontal plane. The target thicknesses are very large,  $10^{17}$  atoms/cm<sup>2</sup>, thanks to the closeness to the nozzle exit. The pressure in the target region is on the other hand very large, of the order  $10^{-4}$  mbar, which of course is not compatible with storage-ring operation. To cure this situation and reach acceptable ring pressures, a number of differentially pumped stages on each side of the target region is proposed to be added. The differential pumping along the circulating beam, using narrow tubings, of course increases the problem of background count rates, and reduces the flexibility of mounting external detectors. Also, the broad gas distribution gives a less defined reaction vertex and contributes to the background.

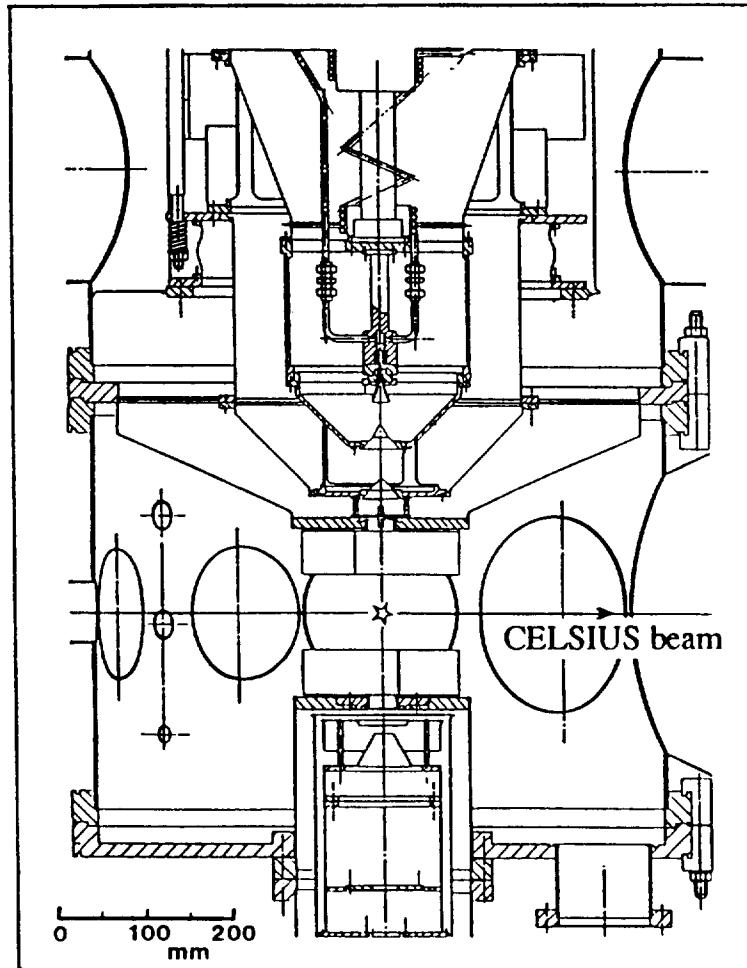
A supersonic gas-jet target has been used by Sperisen et al. [20] at the Indiana Cooler ring at IUCF, Bloomington. The circulating beam is crossing the target beam in the 1 cm space between the nozzle and catcher. Narrow tubing along the circulating beam provides the differential pumping. Hydrogen and nitrogen target beams have been tested with this system. At a target thickness of  $1.2 \times 10^{15}$  atoms/cm<sup>2</sup> N<sub>2</sub> the pressure in the different sections of the target ranges from  $6 \times 10^{-4}$  mbar to  $2 \times 10^{-8}$  mbar. The detectors cover the narrow forward cone between 2° and 12°. Because of background problems due to the tails of the transverse beam distribution interacting with the narrow apertures of the differential pumping system, the gas-jet target is currently removed for modifications.

At the Experimental Storage Ring (ESR) at GSI, Darmstadt, a gas-jet target is under construction by Gruber et al. [21]. The target is designed to deliver a molecular density of  $10^{11} - 10^{14}$  atoms/cm<sup>3</sup> within an interaction length of 3 - 5 mm. The strong vacuum requirements in the ring and the scattering chamber will be satisfied by a differential pumping in four stages along the target beam, both at the injection side and the beam-dump side. By adjusting the gas-input parameters, cluster formation in the target beam will be avoided. This is important since

the high power deposition in the clusters from the high-Z projectiles traversing the target beam will cause evaporation of the clusters. A recirculation and cleaning system will be used to handle more economically the large throughput of gas. The differential pumping along the target beam allows detectors to cover almost the whole horizontal plane and vertical angles between  $\pm 10^\circ$ .

## 6.2 Cluster-jet targets

The cluster-beams are formed by pressing a gas through a trumpet-shaped nozzle with temperature and pressure conditions close to the phase transition to liquid. The central part of the target beam is forming large molecules, so-called clusters, with typically  $10^5$  atoms. The technique was invented by Becker et al. [22] at Karlsruhe, and the dependence of the cluster formation on different input parameters has been studied in detail by Hagen and Obert [23]. The technique has been used to produce target beams of hydrogen for high-energy physics experiments at CERN [17,18] and Saclay [8]. A cluster-jet target, similar to the one used at the UA6 experiment at the SPS at CERN [18], has been installed at the CELSIUS storage ring at Uppsala [24,25]. It is being used also for heavier gases like nitrogen and argon.



*Fig. 2 Vertical cross section of the central part of the cluster-jet target at the CELSIUS ring, showing, in the upper part, the differentially pumped beam source with the cooled nozzle, skimmer and collimators, the central scattering chamber, and, in the lower part, the cryogenic target-beam dump.*

Figure 2 shows the central part of the cluster-jet target at Uppsala, with the differentially pumped beam source, the interaction region and the target-beam dump. The differential pumping is made along the target beam, giving an open construction with possibilities for valves on both sides of the scattering chamber and wide angles for detection of the reaction products. For the target-beam production a gas is pressed by a few atmospheres through the nozzle, which may be cooled down to 20K by a two-stage cryogenerator. A cluster beam is formed which is made to pass a skimmer and a set of collimators, giving an intense beam with a well bounded intensity profile. The cross section of the target beam at the intersection with the circulating CELSIUS beam, 250 mm from the nozzle, is oval-shaped, defined by the oval skimmer, with a length of 8 mm along the circulating beam and 5 mm across. At the bottom, the target beam is collected in a cryogenic beam dump which has to be regenerated typically once per day. Funds have been granted to add a gas recirculation and cleaning system.

The target-beam profiles, cross-sections and thicknesses have been measured by scanning a narrow tube connected to a vacuum gauge across the target beam at the intersection point. The beam profiles are well bounded as shown in Fig. 3 by the curves taken along and perpendicular to the circulating beam. The corresponding curves have also been obtained for the heavier gases nitrogen and argon by varying the input pressure and the nozzle temperature. The pressure values have also been converted to give the target thicknesses of the different gases.

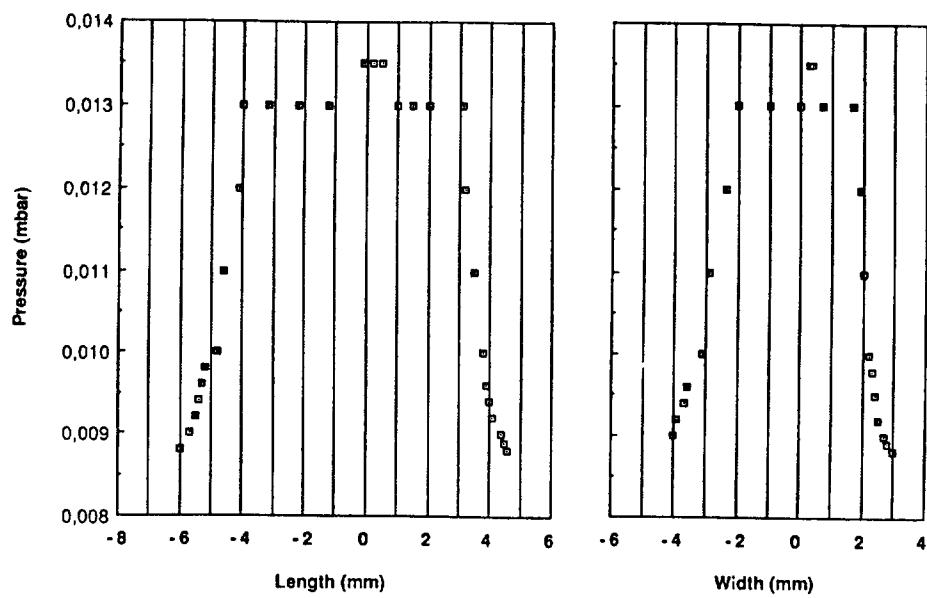


Fig. 3 Intensity profiles of a hydrogen cluster beam, along and perpendicular to the circulating CELSIUS beam, recorded at the intersection point, 250 mm from the nozzle aperture. The hydrogen input pressure was 1.5 bar.

In Table I is given a compilation of running conditions and target thicknesses for the gases hydrogen, nitrogen and argon, taken when the pressure in the scattering chamber reached  $2 \times 10^{-7}$  mbar. The target thicknesses are all of the order  $1 \text{ ng/cm}^2$  or ranging from  $3 \times 10^{14}$  to  $2.4 \times 10^{13}$  atoms/ $\text{cm}^2$  between hydrogen and argon.

Table I. Input parameters and target thicknesses for different gases in the cluster-jet target.

Gas	Z	A	Input pressure [bar]	Nozzle temp. [K]	Gas flow [mbar l/s]	Target thickness [atoms/cm <sup>2</sup> ]	Target thickness [g/cm <sup>2</sup> ]
Hydrogen	1	1	1.5	34	10.0	$3.0 \times 10^{14}$	$0.5 \times 10^{-9}$
Nitrogen	7	14	7.0	160	9.5	$5.0 \times 10^{13}$	$1.2 \times 10^{-9}$
Argon	18	40	2.0	130	2.5	$2.4 \times 10^{13}$	$1.6 \times 10^{-9}$

Life-times of the proton beam at CELSIUS have been measured at different energies with cluster-jets of hydrogen, nitrogen and argon. The target-related partial half-lives are collected in Table II and compared with calculated half-lives due to multiple Coulomb scattering [26] using an acceptance of the CELSIUS ring of  $100\pi$  mm mrad and, in the case of hydrogen, the target thickness of Table I. For nitrogen and argon, the target thicknesses used were determined from triple-coincidence measurements of scattered protons in a monitor telescope. There is a good agreement between the experimental and calculated half-lives, as well as between the target thicknesses of nitrogen and argon measured by the two methods of Table I and II.

Table II. Comparison between experimental and theoretical partial half-lives of proton beams of different energies traversing cluster-jet targets of hydrogen, nitrogen and argon.

Gas	Target thickness [atoms/cm <sup>2</sup> ]	Partial half-life of proton beam					
		48 MeV		72 MeV		600 MeV	
		Exp.	Theor.	Exp.	Theor.	Exp.	Theor.
Hydrogen	$3.0 \times 10^{14}$			2 min	1 min	21 min	20 min
Nitrogen	$3.6 \times 10^{13}$		22 s	27 s		20 min	22 min
Argon	$2.0 \times 10^{13}$		21 s	9 s			

At the Low-Energy Antiproton Ring (LEAR) at CERN, the JETSET collaboration [27,28] is building a cluster-jet target for hydrogen. The antiprotons in LEAR will interact with the protons in the hydrogen target beam and the reaction products will be detected in a close to  $4\pi$  detector system surrounded by a superconducting solenoid. To allow for the installation of the detector system, the distance between the beam production and beam dump is as long as 2 m resulting in a decrease in target thickness as compared to that of previous cluster-jet targets. The design value of the target thickness is  $5 \times 10^{13}$  atoms/cm<sup>2</sup>.

The charmonium spectroscopy experiments at the CERN-ISR [17], using the antiproton beam and the hydrogen cluster-jet target, will be continued at the Fermilab antiproton Accumulator Ring [28,29] at a factor of 10 higher luminosity and a factor of 5 larger acceptance. The hydrogen cluster-jet target, having a distance of 36 cm between the nozzle and the intersection point, produces a target thickness of  $10^{14}$  atoms/cm<sup>2</sup>. The vacuum requirements of  $10^{-10}$  mbar in the Accumulator Ring is met by strong pumping of the target and the interaction region.

In addition to the cluster-jet targets at ISR, LEAR and FNAL, the productive Genova group has constructed a cluster target for heavy gases [30] to produce tagged photon beams at the ADONE electron accelerator at Frascati. Tests with cluster beams of argon and krypton have shown good agreement with the law of corresponding jets [23].

## 7. POLARIZED TARGETS

Important additional information on the nuclear interactions may be obtained in experiments using polarized targets and polarized beams. A general problem, however, is to reach the luminosities available in non-polarized experiments. The polarized atomic-beam targets may be obtained either by state selection in multipole magnets or by optical pumping induced by laser light. To increase the target thickness, the polarized target atoms may be stored in a gas cell. Reviews on free and stored atomic beams as internal polarized targets have been given by eg. Gruebler [31] and Haeberli [32].

### 7.1 Atomic-beam targets

The nuclear polarization is generally obtained by electron-spin polarization which through the hyperfine interaction affects the nuclear spin projection. The hyperfine structure (hfs) diagrams of hydrogen and deuterium are shown in Fig. 4.

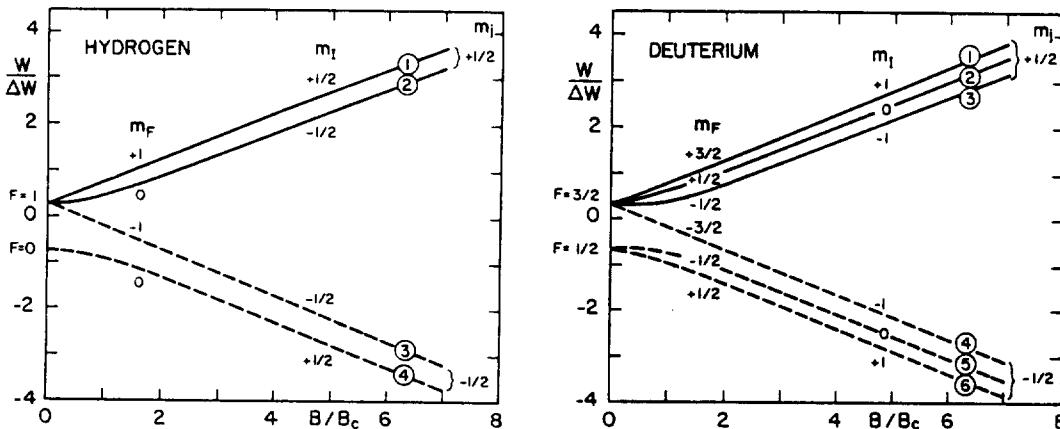
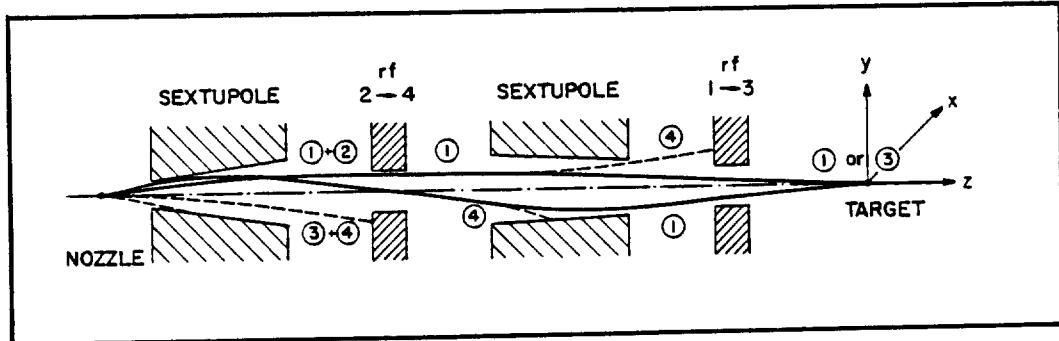


Fig. 4 Energy level diagrams of hydrogen and deuterium in an external magnetic field. The energies are given in units of  $\Delta W = 1420.4$  MHz and  $327.4$  MHz, respectively, and the magnetic fields in units of  $B_c = 507$  G and  $117$  G. From Ref. 31.

In the case of hydrogen we have a simple atomic system with both nuclear and electronic spin equal to  $1/2$ . The resulting hyperfine levels with  $F = 1$  and  $0$  are split up into magnetic substates by an external magnetic field. In the strong-field limit there are two groups characterized by the electron spin projections  $m_J = \pm 1/2$ . The proton spin projections, which are the ones of interest in the experiments, are here  $m_I = \pm 1/2$  in the two groups. The proton polarization is given by  $P_z = N_+ - N_-$ , the difference in number of protons in  $m_I = +1/2$  and  $-1/2$ , with the

total number of particles normalized to 1. The quantization axis is normally chosen along the magnetic field, resulting in vanishing components  $P_x$  and  $P_y$  of the vector polarization. The states 1 and 3 are pure in  $m_I = +1/2$  and  $-1/2$ , with a polarization  $P_z = +1$  and  $-1$ , respectively, independent of the external magnetic field. In the mixed states 2 and 4, the polarization depends on the strength of the field,  $P_z = -1$  and  $+1$ , respectively, in strong fields, going down to 0 in weak fields. In the case several states are occupied, the polarization is given by the weighted mean value.



*Fig. 5 Schematic diagram of a polarized atomic-beam target system with multipole magnets and rf-transitions to produce single substates. The trajectories of hydrogen atoms in different magnetic substates, cf. Fig. 4, are indicated. From Ref. 31.*

A polarized beam of atomic hydrogen may be obtained by state selection in inhomogeneous magnetic fields and rf-transitions between the hyperfine states (cf. Fig. 5). Molecular hydrogen is dissociated by an rf-discharge to give a beam of atomic hydrogen. The first sextupole focus atoms in the states 1 and 2 with  $m_I = +1/2$ , while atoms in 3 and 4 are defocused and lost. In the central field, an rf-transition is induced between the states 2 and 4, so, at the entrance of the second sextupole we have atoms in the states 1 and 4. In this magnet, atoms in the state 4 are again defocused, and we have a remaining hydrogen beam of state 1 atoms focused to the intersection with the circulating beam. The proton polarization is here  $P_z = +1$ , since all atoms are in the state 1 with  $m_I = +1/2$ . An rf-transition between the states 1 and 3 directly after the second magnet provides a rapid switching between the polarizations  $P_z = +1$  and  $-1$ .

Deuterium, with a nuclear spin  $I = 1$ , has a hfs diagram similar to hydrogen (cf. Fig. 4). The  $m_I$  components are here 1, 0 and  $-1$ , and the polarization of the deuteron may be described by a vector polarization,  $P_z = N_+ - N_-$ , the difference in number of deuterons in  $m_I = +1$  and  $-1$ , and a tensor polarization  $P_{zz} = 1 - 3N_0$ , where  $N_0$  is the number of deuterons in states with  $m_I = 0$ . The only pure states are 1 and 4 with  $P_z = \pm 1$  and  $P_{zz} = 1$ . For the other mixed states, the vector and tensor polarizations depend on the strength of the external field. By state selection and rf-transitions one may obtain deuterium beams with different combinations of the deuteron vector and tensor polarizations.

An atomic-beam target consists essentially of the atomic-beam part of a polarized ion source, for which over the years large efforts have been put into increasing the intensity. By cooling the atoms to low temperature (35K), Singy et al. [33] at Zürich have obtained recently an intensity of  $10^{17}$  atoms/s and a density of  $2 \times 10^{12}$  atoms/cm<sup>3</sup> at the exit of the atomic-beam stage.

At CERN an optimized system with cooled beams and superconducting magnets has been studied by Dick and Kubischa [34]. The calculations give target thicknesses of up to  $5.7 \times 10^{12}$  polarized hydrogen atoms per  $\text{cm}^2$  with a target length of 25 mm and a total atomic-beam intensity of close to  $10^{18}$  atoms/s.

Atomic-beam sources with multipole magnets and rf-transitions are also constructed for use as injectors of polarized atoms into storage cells. This way of increasing the target thickness will be discussed separately below.

A Michigan-MIT Collaboration [35] is developing an ultra-cold polarized atomic hydrogen jet target for experiments at the UNK proton accelerator at Serpukhov. Microwave-induced extraction of polarized hydrogen atoms stored at 0.3K and 5T in a dilution refrigerator will give a target beam with a thickness of  $10^{14}$  atoms/ $\text{cm}^2$ .

## 7.2 Laser manipulation

To increase the number of available targets with polarized atoms, a promising development is being made by Ertmer et al. [36] at Bonn. It concerns target beams of laser-polarized alkali atoms. Starting with an effusive beam of  $^{23}\text{Na}$  from some kind of oven with a 1 mm diameter aperture, and a distance of 25 cm to a 1 cm diameter target area, the target thickness is  $6 \times 10^9$  atoms/ $\text{cm}^2$ . This is too small for storage-ring experiments. A large increase in target thickness may be obtained by collimation of the beam through transverse momentum transfer by irradiating the target beam perpendicularly by laser light using a cone-shaped mirror. The gain factor may be as large as 400. Using a spherical mirror to focus the laser light onto the atomic-beam axis leads to a contraction of the phase space of the target beam, and an increase in target thickness by a factor of 3 to 10. Finally, a counterpropagating laser beam results in a deceleration or cooling of the target beam. A deceleration by a factor of 10 gives a corresponding increase in target thickness. So, in total, if all systems can be made to work simultaneously, target thicknesses of  $2 \times 10^{14}$  atoms/ $\text{cm}^2$  may be obtained. The elements suitable for this kind of laser manipulation are the alkali metals lithium, sodium, potassium, rubidium and cesium.

## 7.3 Storage-cell targets

A way to increase the target thicknesses, particularly in the case of polarized atomic-beam targets, is to feed a so-called storage cell, through which the circulating beam will pass. A schematic drawing of a storage cell is given in Fig. 6.

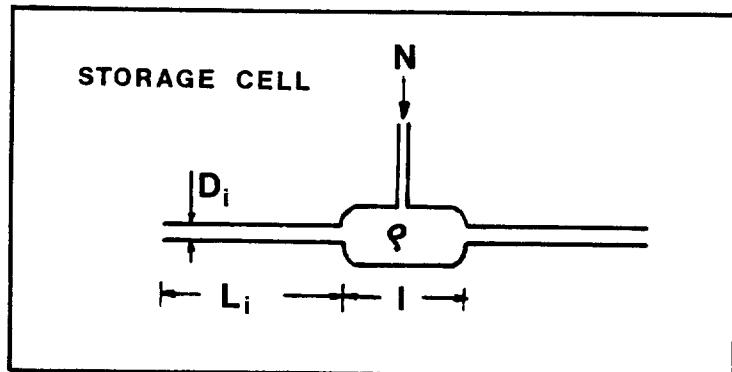


Fig. 6 Schematic diagram of a storage-cell target with openings for the gas injection and for the circulating beam.

The average density in the storage cell is the feed rate divided by the total conductance of the constrictions.

$$\rho = N/C$$

- $\rho$  average density in the storage cell [atoms/cm<sup>3</sup>]
- $N$  feed rate [atoms/s]
- $C$  total conductance of the constrictions [cm<sup>3</sup>/s]

The target thickness is as usual given by the density times the length of the target along the circulating beam.

$$t = \rho l$$

- $t$  average target thickness [atoms/cm<sup>2</sup>]
- $l$  length of the storage cell [cm]

The conductance depends on the square root of the temperature of the system divided by the molecular weight of the target gas, and the sum of the constrictions with diameter in cube divided by the length.

$$C = 3.81 \times 10^3 \sqrt{\frac{T}{M}} \sum_i \frac{D_i^3}{L_i + 1.33D_i}$$

- $T$  absolute temperature [K]
- $M$  molecular weight of the target gas
- $D_i$  diameter of a constriction [cm]
- $L_i$  length of a constriction [cm]

Combining these equations we get an expression for the target thickness in the storage cell. From the simplified relation below it is evident that high target thicknesses require a large feed rate, a long storage cell, a heavy gas at low temperature, and few, long and narrow constrictions.

$$t \propto N l \sqrt{\frac{M}{T}} \sum_i \frac{L_i}{D_i^3} \quad L \gg D$$

In a typical case we have feed rates of  $10^{17}$  atoms/s, a cell of 10 cm length, three constrictions of 10 cm length and 1 cm diameter. For atomic hydrogen at room temperature the target thickness will be  $6 \times 10^{13}$  atoms/cm<sup>2</sup> and at 20K  $2 \times 10^{14}$  atoms/cm<sup>2</sup>. This can be compared with the value  $10^{12}$  atoms/cm<sup>2</sup> for a free atomic beam of  $10^{17}$  atoms/cm<sup>2</sup> s and the velocity  $10^5$  cm/s. The gain, a factor of 100, has to be paid by the narrow constrictions which may cause background problems, the target length which results in an undefined reaction vertex, and the cell walls which may disturb the reaction products. Special attention has also to be paid to the wall coating and the cooling to keep recombination and depolarization at low values.

A storage-cell target for polarized hydrogen is being prepared by a collaboration from Heidelberg, Marburg, Munich and Madison [37]. It is based on an atomic-beam source with multipole magnets and rf-transitions of the

kind described above giving  $10^{17}$  polarized hydrogen atoms per second. This is expected to give a target thickness of  $10^{14}$  atoms/cm<sup>2</sup> in the storage cell. The system will be used to produce polarized beams in the so-called FILTEX experiments at the test storage ring (TSR) at Heidelberg for protons and at LEAR, CERN for antiprotons. A similar system is proposed for the study of reactions between polarized electrons in the HERA-ring at Hamburg and the polarized target protons.

An Argonne-Novosibirsk collaboration [38] is preparing a polarized deuterium storage-cell target for the 2 GeV electron storage ring VEPP-3 at Novosibirsk. In the first two phases, deuterium will be polarized conventionally by a sextupole and rf-transitions, and injected into a passive and active storage cell, respectively, the latter with a clam-shell mechanism to allow for narrower openings following the electron-beam injection. This arrangement will increase the target thickness from  $6 \times 10^{11}$  to  $1 \times 10^{13}$  atoms/cm<sup>2</sup>. In a third phase, the target thickness will be further raised to  $4 \times 10^{14}$  atoms/cm<sup>2</sup> by feeding the storage cell with deuterium polarized through spin exchange with optically pumped alkali atoms.

A storage-cell target for polarized  $^3\text{He}$  has been proposed by Milner et al. [39] at Pasadena. The system consists of two stages; a cell fed by helium gas and polarized by laser light and the storage-cell target fed by the polarized helium through a capillary tube. In the first cell, the  $^3\text{S}_1$  metastable state is populated by a discharge and polarized through optical pumping by polarized laser light to the  $^3\text{P}_0$  state. The  $^1\text{S}_0$  ground state atoms are then polarized through exchange collisions. A flow of  $10^{17}$  polarized  $^3\text{He}$  atoms per second are expected to enter the storage cell through the capillary tube. The target thickness is of course dependent on the geometry of the storage cell, but assuming the same dimensions as discussed above, it will reach  $10^{14}$  atoms/cm<sup>2</sup>.

## 8. SOLID TARGETS

We will now turn to the solid targets and start by looking at some characteristics, Table III. Fibers of the refractory elements carbon, molybdenum and tungsten, with diameters 7, 13 and 4  $\mu\text{m}$ , microparticles in the form of 1  $\mu\text{m}$  nickel-dust and 25  $\mu\text{m}$  frozen hydrogen pellets, all have target thicknesses of the order  $10^{19}$  atoms/cm<sup>2</sup>. They are locally very thick, but taking into account the ratio between the target area and stored-beam area, the effective target thicknesses come into the range acceptable for storage-ring operation. A reduction of the effective target thickness may also be obtained by a periodic sweeping of the beam over the target.

Table III. Characteristics of some solid targets.

	Fibers			Microparticles	
	$^{12}\text{C}$	$^{96}\text{Mo}$	$^{184}\text{W}$	"dust"	"pellet"
Fiber diameter [ $\mu\text{m}$ ]	7	13	4		
Microparticle diameter [ $\mu\text{m}$ ]				1	25
Mass density [ $\text{g}/\text{cm}^3$ ]	2.27	10.2	19.3	8.9	0.142
Target thickness [ $\text{mg}/\text{cm}^2$ ]	1.3	10.5	6.1	0.55	0.22
- " - [ $10^{19}$ atoms/cm <sup>2</sup> ]	6.3	6.5	2.0	0.6	6.6

### 8.1 Fiber targets

The fibers given in Table III have been tested by Holmqvist et al. [24] with 3 MeV protons to evaluate the luminosities under storage ring conditions, assuming a proton-beam energy of 200 MeV, a beam diameter of 3 mm at the target position and the same power deposition per unit length of the fibers. The luminosities  $2.9 \times 10^{32}$  and  $3.5 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$  for molybdenum and tungsten, respectively, are upper limits where the fibers were burned off within less than 1 h of irradiation, whereas in the case of carbon the fibers were still intact after 10 hours of irradiation at a power deposition corresponding to a luminosity of  $1.5 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ . Carbon fibers have also been impregnated by Hoff [24] with different elements like cobalt, lanthanum, thorium and uranium to form low vapour pressure carbides, thus increasing the number of available target materials.

At the CELSIUS storage ring at Uppsala carbon fibers of 7  $\mu\text{m}$  diameter have been used in initial test experiments and as a diagnostics tool in combination with pick-up electrodes in the ring and a spectrum analyzer to measure position, profile and life-time of the stored proton beam. The measurements have been made by introducing vertical as well as horizontal fibers into the uncooled beam at the injection energies 48 and 72 MeV, and, following acceleration, at 500 and 600 MeV. By scanning the fibers across the stored beam and measuring the life-time of the proton-beam intensity as a function of fiber position a quite different behaviour has been found for vertical and horizontal fibers. In the case of vertical fibers scanned horizontally over the beam, the life-time showed a maximum with the fibers at the center of the beam. At the tail of the beam distribution a long beam life-time was observed following the short initial decay. Typical values are 20 ms half-life with three twisted carbon fibers of 7  $\mu\text{m}$  diameter at the center of a 72 MeV uncooled proton beam, going down to 10 ms off center and staying at 5 s after a 10 ms initial decay at the tail. Horizontal fibers scanned vertically over the beam showed, on the other hand, clear minima with the fibers at the center of the beam, e.g. 20 ms and 175 ms half-lives at 72 MeV and 600 MeV proton energies, respectively, using one carbon fiber. The difference in behaviour between introducing vertical and horizontal fibers has been explained by Reistad [40] as due to the horizontal dispersion at the target position.

The target group at the Indiana Cooler ring [41] has also used 7  $\mu\text{m}$  carbon fibers in test experiments. In addition they have manufactured so-called micro-ribbons of carbon with a linear density of  $1 \times 10^{15}$  atoms/cm as compared to  $4.3 \times 10^{16}$  atoms/cm for the 7  $\mu\text{m}$  carbon fibers. The effective target thickness of  $1.3 \times 10^{16}$  atoms/cm<sup>2</sup> of the micro-ribbons, taking into account the width of the beam, has been reduced to  $6 \times 10^{14}$  atoms/cm<sup>2</sup> by a periodical sweeping of the ribbons across the proton beam. The life-time of a 108 MeV cooled proton beam was 520 ms at a 5% duty cycle, 1 ms overlap time out of the 20 ms sweep time. To further reduce the overlap time, and thus the peak luminosity, higher wiggling frequencies have been used by sweeping the proton beam over stationary micro-ribbon targets.

Ribbon targets of  $40 \mu\text{g}/\text{cm}^2$  carbon and  $200 \mu\text{g}/\text{cm}^2$  gold, suspended from a support at the top, have been tested at different accelerators by Koch et al. [42] at Jülich. An average luminosity of  $3 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$  was obtained with  $10^{11}$  protons of 150 MeV in the SATURNE ring at Saclay and the carbon ribbon target. The proton beam, injected every 1.3 s, had a half-life of 15 ms with the lower end of the ribbon at the center of the beam. To reduce the peak luminosity, a periodical sweeping of the proton beam over the target has been suggested. In the case of gold ribbons, the average luminosities were  $2 \times 10^{30} \text{ cm}^{-2}\text{s}^{-1}$  with  $5 \times 10^{10}$  protons of 150 MeV and  $1 \times 10^{31} \text{ cm}^{-2}\text{s}^{-1}$  with  $10^{11}$  protons of 300 MeV. Because of the strong dependence of the Coulomb scattering on the charge of the target nuclei, the gold target gave much shorter proton-beam half-lives than the carbon target; 0.6 ms as compared to 15 ms at 150 MeV. At 300 MeV the proton-beam half-life was 1.2 ms using the gold target.

## 8.2 Microparticle targets

There are different ways of introducing microparticles into the circulating beam. These so-called dust targets may be used in a storage cell, much the same as the storage cells discussed in connection with the polarized gas targets. The problems with the charging and the uncontrollable diffusion of dust out of the cell have turned the attention to producing microparticle beams instead. So-called dust guns have been developed as well as systems for beams of gas-particle mixtures where the gas is differentially pumped away to give a beam of microparticles. The advantages with the microparticle target are that the material is commercially available for many elements, there is a very small target heating, the target thickness can be matched to optimum storage-ring operation, and that there are no special vacuum problems, except of course for the differentially pumped system.

A dust gun for microparticle beams has been developed by Tanabe et al. [43] at Tokyo. The technique is based on contact-charging dust particles in electric fields on metal surfaces, followed by extraction and acceleration to prepare a dust beam. They used  $1 \mu\text{m}$  Ni-particles collected in a cup, measuring the current. A permanent magnet prevented the dust to fall back into the system. The target thickness of the dust beam was shown to be  $10^{14} - 10^{16}$  atoms/cm<sup>2</sup> by bombarding with a 65 MeV alpha beam and detecting the elastically scattered alpha particles. A similar system was developed at IUCF [15], but later abandoned because of problems to get sufficient target thickness in a reliable and reproducible way. The third group working on the problems with a dust gun is Prasuhn et al. [44] at Bonn. Starting with a simple arrangement similar to the previous ones, they are now improving the system by shaking the dust on a foil with an electromagnet, and optimizing the geometry of the extraction and focusing elements.

Recently, the Indiana group [41] has developed a microparticle-beam target, letting a gas-microparticle mixture enter a vacuum chamber through a capillary tube and be collected in a catcher. The microparticles are evenly mixed with the carrier gas through an electromagnetic agitator, and the properties of the particle beam have been investigated by scanning it over a laser beam. Different combinations of particle materials and sizes, carrier gases, and capillary-tube lengths and diameters have been tested with respect to microparticle-beam profiles, velocity distributions and stability in time. The particle velocities, being in the range 100 to 300 m/s, have been well reproduced by calculations. The dependence on the different input parameters is thus to a large extent understood.

To use the microparticle target in a storage-ring experiment, the carrier gas has to be removed, leaving a pure dust beam. A scheme for differential pumping on the beam source side is now being applied to the Indiana microparticle target in preparations for experiments at the Cooler ring [41]. Microparticle beams of carbon, nickel and copper have been produced in target thicknesses of the order  $10^{15}$  atoms/cm<sup>2</sup>. A similar system has been described by Popov [45], Novosibirsk.

## 8.3 Pellet targets

The so-called hydrogen pellet target may be seen as a special case of the microparticle targets. It has been proposed to be used in experiments at the CELSIUS ring to detect rare decays of neutral pions produced in proton-proton reactions [46]. It seems to fulfil a number of stringent requirements, like luminosities above  $10^{32}$  cm<sup>-2</sup>s<sup>-1</sup>, access to close to  $4\pi$  detection, a good spatial vertex definition, minimum Coulomb scattering of leptons, tagging possibilities, acceptable perturbation to the stored beam and acceptable pressure load on the ring. Using  $10^{10}$  protons of 550 MeV within a diameter of 3 mm in the ring and a pellet of  $25 \mu\text{m}$  in the beam at every

moment, the luminosity will be  $10^{32} \text{ cm}^{-2}\text{s}^{-1}$ , giving 400  $e^+e^-$ -decays per day at a 50% acceptance.

The production of pellets has been studied in fusion research, e.g. by Foster et al. [47] at Urbana including the droplet production by pressing liquid hydrogen through a capillary nozzle and applying an acoustic excitation to break up the liquid stream into a droplet train, the charging of the droplets, and the cooling of the droplets to frozen pellets by evaporation at a pressure slightly below the triple point pressure of 72 mbar. Trostell et al. [48] at Studsvik and Uppsala have obtained funds for building a prototype pellet generator to reproduce these points and to further study the continuous injection into vacuum, the particle optics for steering the pellets, the diagnostics and the pellet dumping. The pressure load in the scattering chamber due to heating of the pellets by the walls and the circulating beam has been calculated by Trostell [49] to be of the order  $5 \times 10^{-3}$  mbar l/s. Heavy pumping around the target will thus be needed to meet the vacuum requirements of storage rings.

## 9. CONCLUSION

There is currently a strong development in the field of internal targets to be used in storage-ring experiments. It is however important to be aware of the advantages and drawbacks of the different target systems. There is obviously no general-purpose system available, but one has in each case to choose and optimize the target in order to meet the specific needs of the experiment to be done.

\* \* \*

## REFERENCES

- [1] V. Bartenev et al., *Cryopumped, Condensed Hydrogen Jet Target for the National Accelerator Laboratory Main Accelerator*, Adv. Cryog. Eng. **18**(1973)460.
- [2] P. Mantsch and F. Turkot, *Design of a Gas Jet Target Operating at Ambient Temperature in the Main Ring*, Fermilab Report TM586 (1975).
- [3] A.C. Melissinos and S.L. Olsen, *Physics and Technique of Gas Jet Experiments*, Physics Reports **17**(1975)77.
- [4] B.C. Stringfellow et al., *Accelerator Internal Target Experiments Using a Supersonic Gas Jet*, Nucl. Instr. and Meth. in Phys. Res. **A251**(1986)242.
- [5] C. Baglin et al., *Charmonium Spectroscopy at the ISR Using an Antiproton Beam and a Hydrogen Jet Target*, CERN/ISRC/80-14, 1980.
- [6] J. Antille et al., *Proposal for the Study of  $e^+e^-$ ,  $\gamma$ ,  $\pi^0$  and Hyperon Production in  $\bar{p}p$  Reactions at  $\sqrt{s} = 22.5$  GeV Using an Internal Jet Target at the SPS*, CERN/SPSC/80-63, 1980.
- [7] K. Kilian, D. Möhl, J. Gspann and H. Poth, *Internal Targets for LEAR*, Workshop on Physics at LEAR with Low-Energy Cooled Antiprotons, Erice, Sicily, 1982 (Plenum, New York, 1984).

- [8] R. Burgei, M. Garçon, M. Grand, B. Gonel, R. Maillard, A. Malthiery and J. Martin, *A Clustered Hydrogen Gas Jet Target for the Saturne Accelerator*, Nucl. Instr. and Meth. **204**(1982)53.
- [9] R. Pollock, D. Miller and P. Singh, *The IUCF Cooler-Tripler*, Proposal for an Advanced Light-Ion Physics Facility, Indiana University, 1980, and R. Pollock, D. Miller and P. Singh, *The IUCF Cooler*, Proposal for an Advanced Light-Ion Facility, Indiana University, 1982.
- [10] Topical Conference on Electronuclear Physics with Internal Targets, SLAC, Stanford, California, 1989.  
To be published by World Scientific Publishing Company.
- [11] Workshop on Physics at LEAR with Low-Energy Cooled Antiprotons, Erice, (Plenum, New York, 1984).
- [12] Workshop on Polarized Targets in Storage Rings, ANL, Argonne, Illinois, 1984. ANL-84-50.
- [13] Workshop on Nuclear Physics with Stored Cooled Beams, McCormick's Creek State Park, Spencer, Indiana, 1984. AIP Conference Proceedings, No. 128.
- [14] Workshop on Internal Targets for COSY, KFA, Jülich, 1987. JüL-Spez-409.
- [15] H.O. Meyer, *Internal Targets for the Indiana Cooler: Possibilities*, CEBAF Workshop, Newport News, U.S.A., 1985.
- [16] M. Macri, *Gas Jet Internal Targets*, CERN Accelerator School - Antiprotons for Colliding Beam Facilities, Geneva, Switzerland, 1983. Report CERN 84-15, 1984, p. 469.
- [17] C. Baglin et al., *J/Ψ Resonant Formation and Mass Measurements in Antiproton-Proton Annihilations*, Nucl. Phys. **B286**(1987)592.
- [18] L. Dick and W. Kubischa, *Physics with Jet Targets at the SPS  $\bar{p}p$  Collider*, Hadronic Physics at Intermediate Energy, T. Bressani and R.A. Ricci (editors), Elsevier Science Publishers B.V., 1986, p. 209.
- [19] G. Gaul, W. Bickel, D. Grzonka, M. Heupel and G. Höcker, *Development of a Gas-Jet Target for Experiments in Storage Rings*, in Ref. 14.
- [20] F. Sperisen, A. Berdoz, J. Doskow, P. Li, W. Lozowski, H.O. Meyer, P. Pancella, T. Rinckel, A. Ross, R. Barbieri and R.A. Bonham, *Gas Jet Target*, IUCF Scientific and Technical Report 1987/88, p. 194.
- [21] A. Gruber, W. Bourgeois, B. Franzke, A. Kritzer and C. Treffert, *Internal Gas-Jet Target for the ESR at GSI*, Nucl. Instr. and Meth. in Phys. Res. **A282**(1989)87.
- [22] E.W. Becker, K. Bier and W. Henkes, *Strahlen aus Kondensierten Atomen und Molekülen im Hochvakuum*, Z. Phys. **146**(1956)333.
- [23] O.F. Hagena and W. Obert, *Cluster Formation in Expanding Supersonic Jets: Effect of Pressure, Temperature, Nozzle Size and Test Gas*, J. Chem. Phys. **56**(1972)1793.

- [24] C. Ekström et al., *The CELSIUS Project*, Physica Scripta **T22**(1988)256.
- [25] C. Ekström, *A Cluster-Jet Target for Heavy Nuclei and the Physics Programme at CELSIUS*, in Ref. 10.
- [26] D. Prasuhn, private communication.
- [27] G. Bassompierre et al., *JETSET: Physics at LEAR with an Internal Gas Jet Target and an Advanced General-Purpose Detector*, CERN/PSCC/86-23, 1986.
- [28] M. Macri, *Study of Resonant States Formed in  $\bar{p}p$  Annihilations Using a  $H_2$  Jet Target at LEAR and Fermilab Antiproton Accumulator*, Folgaria School of Hadronic Physics, Folgaria, Italy, 1989. To be published by World Scientific Publishing Company.
- [29] K.K. Seth, *High Resolution Charmonium Spectroscopy with an Internal Gas Jet Target*, in Ref. 10.
- [30] M. Taiuti, *Argon Gas Jet Target for a Tagged Photon Beam at ADONE*, in Ref. 10.
- [31] W. Grüebler, *Polarized Atomic Beams for Targets*, in Ref. 12.
- [32] W. Haeberli, *Free and Stored Atomic Beams as Internal Polarized Targets*, in Ref. 13.
- [33] D. Singy, P.A. Schmelzbach, W. Grüebler and W.Z. Zhang, *Production of Intense Polarized Hydrogen Atomic Beams by Cooling the Atoms to Low Temperature*, Nucl. Instr. and Meth. in Phys. Res. **A278**(1989)349.
- [34] L. Dick and W. Kubischa, *The CERN Polarized Atomic-Beam Target Project*, in Ref. 10.
- [35] T. Roser et al., *An Ultra-Cold Polarized Atomic Hydrogen Jet for Internal Target Applications*, in Ref. 10.
- [36] W. Ertmer, *Atomic Beam Targets*, in Ref. 14.
- [37] E. Steffens, *Target for the FILTEX Experiment*, in Ref. 14.
- [38] L. Young et al., *Internal Polarized H and D Targets*, in Ref. 10.
- [39] R.G. Milner, R.D. McKeown and C.E. Woodward, *A Polarized  $^3He$  Target for Nuclear Physics*, Nucl. Instr. and Meth. in Phys. Res. **A274**(1989)56.
- [40] D. Reistad, *Status of the CELSIUS Cooler-Storage Ring Project*, 12th Int. Conf. on Cyclotrons and their Applications, Berlin, FRG, 1989.
- [41] A. Berdoz, J. Doskow, W. Lozowski, H.O. Meyer, P. Pancella, T. Rinckel, H. Rohdjess, A. Ross and F. Sperisen, *Target Development*, IUCF Scientific and Technical Report 1988/89, p. 132.
- [42] H.R. Koch, G. Riepe, A. Hamacher, J. Pfeiffer, O.W.B. Schult, A. Nakach, M. Olivier, A. Dewald and

P. von Brentano, *Thin Ribbons as Internal Targets in Storage Rings*, Nucl. Instr. and Meth. in Phys. Res. **A271**(1988)375.

- [43] T. Tanabe, I. Sugai and M. Yasue, *Study of a Microparticle Internal Target for a Cooler Ring*, Nucl. Instr. and Meth. in Phys. Res. **A256**(1987)439.
- [44] D. Prasuhn, *First Ideas and Tests of Dust Targets*, in Ref. 14.
- [45] S. Popov, *Nuclear Physics Experiments with Internal Targets in Electron Beams at Novosibirsk*, in Ref. 10.
- [46] L. Bergström et al., *WASA: Wide Angle Shower Apparatus for Particle Physics at CELSIUS*, Proposal for CELSIUS experiment, University of Uppsala, 1987.
- [47] C.A. Foster, K. Kim, R.J. Turnbull and C.D. Hendricks, *Apparatus for Producing Uniform Solid Spheres of Hydrogen*, Rev. Sci. Instr. **48**(1977)625.
- [48] B. Trostell, private communication.
- [49] B. Trostell, *The Thermodynamics of Hydrogen Micro-Spheres as Internal Targets in Ion Storage Rings*, Proc. 1st European Particle Accelerator Conference, Rome, 1988. (World Scientific, Singapore, 1989).

## LIVING WITH RADIATION

*S. Kullander*

Department of Radiation Sciences, Uppsala University, Uppsala, Sweden

### ABSTRACT

Visible radiation, indispensable for plants to grow and animals to see, is absorbed like all electromagnetic radiation in the form of photons. Ultraviolet and other ionizing radiations from the outer space and heat emitted by the earth are absorbed and reflected by the atmosphere so that the climate is pleasant. Synchrotron radiation from remote radiogalaxies helps to unravel the secrets of the cosmos and the Northern Light gives valuable information on charges circulating in the ionsphere near the pole. Nowadays, radiation is produced artificially in accelerators and reactors and is essential for contemporary natural science, medicine and technology. A beautiful example is the diffraction of synchrotron radiation by large biomolecules revealing their atomic structure and which can be used in the design of drugs.

### 1. INTRODUCTION

The 1989 CERN Accelerator School was held in Uppsala which is situated 70 km north of Stockholm. Uppsala was a political and religious centre already in the year 500 AD. During the following four centuries it was the centre of the Swedish vikings. The university was founded in 1477 and is the oldest one in Scandinavia. Like the great French and Italian universities of the time, Uppsala had faculties of Theology, Jurisprudence, Medicine and Philosophy.

One effect of the Reformation was that the church ceased to assume any responsibility for Uppsala University. Thus the development was slow and not until the 17th century was the basis for a university of international standards created. The foremost natural scientist of the university during the 17th century, Olof Rudbeck, famous in particular for his discovery of the lymphatic vessels, was a professor of medicine.

The natural sciences at the university had a time of prosperity during some decades about the middle of the 18th century. The astronomer Anders Celsius, the mathematician and physicist Samuel Klingenstierna, the chemist Johan Gottschalk Wallerius and the botanist and zoologist Carl von Linné, the most famous scientist among them all, were at the time professors at the university.

In the 19th century, Anders Jonas Ångström measured the solar radiation, this resulting in a spectrum atlas. His last name is best known as a length unit defining the atomic scale. In the 20th century, a mark of continued excellence is that six Nobel Prizes have been awarded to Uppsala scientists. One of the laureates, the chemist The Svedberg, initiated and was responsible for the first high energy accelerator in Scandinavia. In the beginning of the 1950s, protons from the Gustaf Werner synchrocyclotron were accelerated to 185 MeV which for some years was the highest energy in Western Europe. The magnet of the first CERN accelerator, the 600 MeV synchrocyclotron was based on the Uppsala magnet design. Nowadays, the rebuilt Gustaf Werner synchrocyclotron and the cooler storage ring, CELSIUS, are important radiation sources in Uppsala.



Fig. 1. Students and professors relax and philosophize in Uppsala at the turn of the 19th century

In the present exposé, some examples show that radiation is an important ingredient of everyday life and is not restricted to the use of workers in accelerator laboratories. The stepwise absorption of radiation will be illustrated by using photography as an example. The quantum description of radiation is suitable for describing not only absorption but also production and stimulated emission of radiation. The latter process explains the functioning of the laser and a novel application of the cyclotron maser that will be described briefly. The changing composition of our atmosphere will be discussed since this problem is one of the major global problems at present and the effects on the radiation environment could be dramatic. The importance of radiation in medicine is well known and the increased use of accelerators in diagnosis and therapy will be mentioned. Finally, it will be pointed out that not only light from stars provides us with information from the cosmos but also radiation by molecules, ions and electrons in the sky is a useful source of knowledge.

## 2. RADIATION AND PHOTOGRAPHY

Water waves are absorbed efficiently by the barriers surrounding a harbour but how are waves of electromagnetic radiation extinguished? We shall illustrate the mechanisms of absorption of radiation by describing in some detail how light is absorbed in a photographic film. The film's emulsion consists of fine light-sensitive crystals of silver salts evenly dispersed in gelatine. Silver salts of bromine, iodine and chlorine are most commonly used. On exposure to light, small invisible clusters of precipitated silver form an image, usually called the latent image. In the subsequent development more silver from the developing fluid accumulates around the primary clusters of precipitated silver. Normally during development, the amount of precipitated silver in the emulsion increases a million times or more and a picture becomes visible.

We shall limit our discussion to the primary photographic process in which absorption of light leads to a latent image of silver. The emulsion is assumed to consist of silver bromide crystals slightly contaminated with silver sulphide. The positive silver ions and negative bromide ions in the crystal are arranged regularly and well separated. To be precipitated as a free silver atom, the positive silver ion must pick up a free electron. Such an electron not belonging to any special ion is usually called a conduction electron since it is able to move around freely and transport electricity in the crystal.

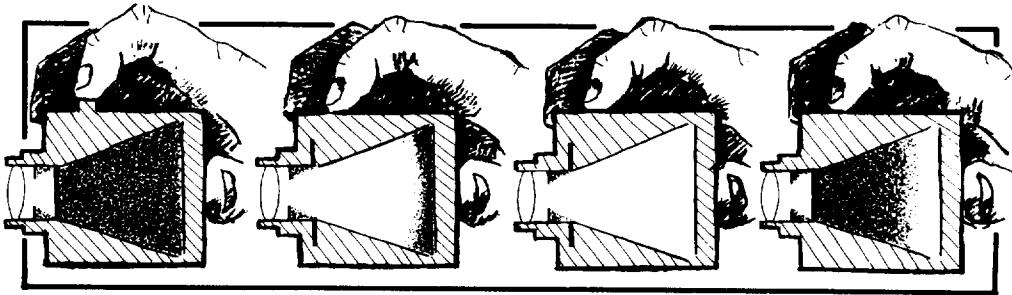


Fig. 2. When photographing, the exposure must be carefully matched to the sensitivity of the film. The light pulse is limited in space to the interior of the camera and in time to the selected exposure time.

The first step is the liberation of an electron from its binding in a bromide ion. An energy of 1 eV, corresponding to  $1.6 \times 10^{-19}$  J, must be absorbed from the light wave by the bromide ion for the most weakly bound electron to convert to a conduction electron.

We know that a film with a sensitivity of 100 ASA units must be exposed for about 15 lux-seconds to have a picture with good contrast. This exposure corresponds to an absorbed energy of  $2.5 \times 10^{-2}$  J m<sup>-2</sup>. Assuming that the energy is absorbed uniformly over the surface of the film, a bromide ion with an exposed surface of  $12 \times 10^{-20}$  m<sup>2</sup> would receive a maximum energy of  $3 \times 10^{-21}$  J. Hence, according to this reasoning, no silver would be precipitated since we know that  $1.6 \times 10^{-19}$  J is necessary to break the bond between the bromide ion and the electron.

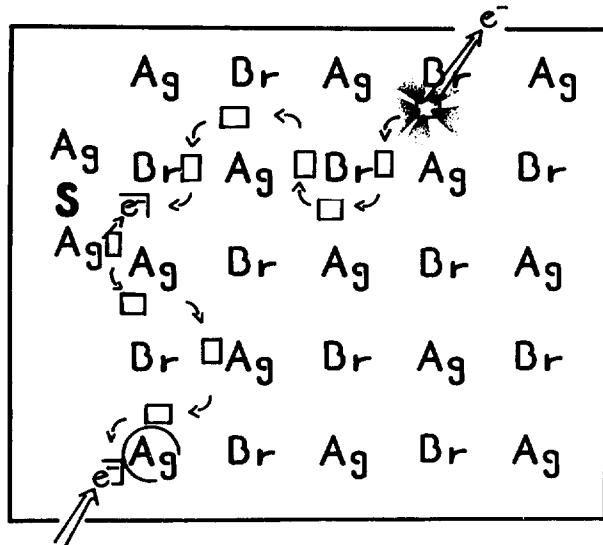


Fig. 3. After photon absorption, the electron originally bound to a bromide ion starts to circulate freely in the crystal. The residual bromine atom attracts an electron from a neighbouring bromide ion which in turn attracts an electron from another ion. In this way positive charge is transported towards the surface of the crystal where a silver ion of an impurity of silver sulphide takes over the transport of positive charge. Near the edge of the crystal, the energy conditions favour the capture of a conduction electron by a silver ion which is precipitated and, as free silver, contributes to the blackening of the film.

What is wrong? The amount of light required for a good picture is known by experience, and the energy necessary to break the bond of the loosest bound electron in a bromide ion is determined in laboratory

experiments. That the energy comes from light absorbed on one bromide ion only seems reasonable since we know that ions in a crystal lattice do not interact strongly with one another. But inspite of these considerations, the film is indeed blackened. We are in the same dilemma as the physicists were at the end of the nineteenth century when they tried to explain why electrons bound in a metal surface are ejected by light. Similar considerations, as in our example with the photographic film, showed that the intensity of the incident light should have been insufficient to liberate the electrons from the metal. It was observed that the energy of the ejected electron was dependent on the frequency of the light and below a certain frequency, ejection of electrons did not occur at all. Einstein in 1905 explained this photoelectric effect by postulating that the energy of light is carried in small localized quanta and that a quantum is transferred to an electron in a point-like fashion.

Hence, the energy of light is not uniformly dispersed in time and space but concentrated in quanta, now called photons. The energy of a photon is related to the frequency,  $f$ , of the light wave according to the formula,

$$E = hf \quad (1)$$

The constant of proportionality,  $h$ , is named after the German physicist Max Planck, and is  $4.135 \times 10^{-15} \text{ eVs}^{-1}$ . By this formula, we find that red light with a frequency of  $4 \times 10^{14} \text{ Hz}$  corresponds to a photon energy of  $1.6 \text{ eV}$ . The photons of violet light carry almost twice the energy of those of red light and the average energy of the photons of sunlight is  $2.5 \text{ eV}$  equal to  $4 \times 10^{-19} \text{ J}$ . Photons of visible light are apparently sufficiently energetic to liberate electrons from bromide ions.

We recall that the absorbed energy density should be around  $2.5 \times 10^{-2} \text{ Jm}^{-2}$  for a visible photographic picture. Hence the density of photons incident on the film can be estimated from the knowledge of the average photon energy of light. The result is that an area on the film equivalent to 100 closely packed bromide ions absorbs on average one photon.

A photon may pass several layers of atoms before being absorbed in the film. Absorption occurs only if a photon meets in one and the same point, an electron in a bromide ion. If this condition is not satisfied the photon passes through the ion and penetrates deeper into one or more neighbouring ions.

The first step in the formation of a silver atom is a reaction between a bromide ion and a photon indicated by the formula,



After the reaction, the electron liberated from the ion moves freely in the lattice of the crystal. The neutral bromine atom attracts an electron relatively easily from a neighbouring negative bromide ion which becomes neutral. The so-formed bromine atom in turn attracts an electron from an adjacent bromide ion and so on. Thus, charges are transported within the lattice and this transport proceeds to the surface of the crystal.

In the early days of experimental photography, it was known that a certain contamination of silver sulphide was needed to get a good picture. The silver sulphide gives off an electron to a neighbouring neutral bromine atom more readily than a bromide ion. When a bromine atom is formed near a molecule of silver sulphide, the transport of charges mediated by bromine ceases. The charge transport continues via the silver ions of the lattice.

At an edge of the crystal, a conduction electron floating freely in the lattice can be captured so that a silver atom is precipitated. The initial bromine atom that absorbed the photon and the finally precipitated silver atom lie near each other because the transport of charges by bromine and silver is limited to distances of fractions of a millimetre within one of the many crystals on a photographic film.

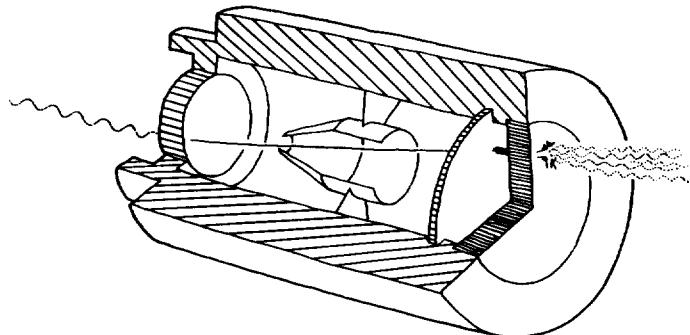


Fig. 4. A photon is absorbed in the photocathode of an image intensifier. The electron liberated is accelerated towards a channel plate at a high positive potential with respect to the photocathode. In one of the many tubes of the channel plate, up to a million electrons are produced and subsequently stopped in a fluorescent screen where photons are produced.

Other electromagnetic radiations of lower frequency than light, such as heat and radio waves are absorbed also in a step-wise fashion. The low energy of the photons of these radiations is however not able to separate electrons from their binding in atoms or molecules of a non-conducting medium. The molecules having absorbed these photons acquire additional motion and the resulting heat distributes itself after some time among all the molecules of the medium. On the absorption of photons by a conductor, excess energy is acquired by the conduction electrons which move about freely in the metal and spread their energy easily. This makes metals conduct heat very efficiently.

In a semiconductor, the difference in binding energy between electrons bound in atoms, the so-called valence electrons, and electrons moving around freely in the lattice, the conduction electrons, is small compared to the situation in an isolator. Normally, in a semiconductor very few of the electrons are free to move as conduction electrons. When irradiated by light, valence electrons may easily acquire sufficient energy to break the bond with their host atoms and become conduction electrons. If an electric tension is applied across a semiconductor, these liberated electrons give rise to a current. Photodiodes and other light-sensitive semiconductor components function according to this principle.

The above examples show how the energy of light is absorbed in a stepwise fashion and converted to kinetic energy of electrons, atoms and molecules. Photosynthesis and vision, most important processes in plants and animals, absorb light similarly.

### 3. RADIATION AND THE ATMOSPHERE

The earth receives visible radiation from the sun and itself radiates heat. The intensity of solar energy reaching the top of our atmosphere is  $1.35 \text{ kW m}^{-2}$  and the wavelength distribution is well described by the radiation spectrum of a black body at a temperature of 5800 K. During its continued transit through the atmosphere, sunlight is absorbed and reflected back into space by atmospheric gases. If there had been no

atmosphere and the earth could have been considered as a blackbody, the average ground temperature would have been around 0 °C. In the presence of nitrogen and oxygen but without so-called greenhouse gases, the average temperature would have been around -20 °C since the atmosphere absorbs and reflects the incoming radiation.

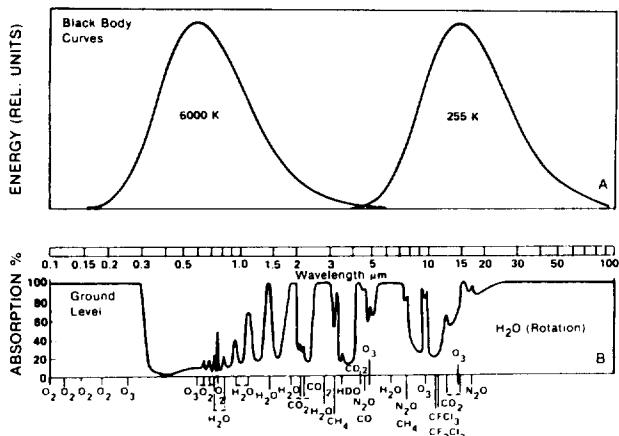


Fig. 5. (A) Black-body curves for 6000 K and 255 K, which approximate the mean emitting temperatures of the sun and the earth. (B) Atmospheric gaseous absorption for radiation passing from the top of the atmosphere to the ground level. From Ref [1].

Half of the radiation energy is absorbed and reflected by the atmosphere so that at the ground level half the energy remains. The ultraviolet portion is absorbed by high-altitude ozone and the absorption elsewhere in the wavelength spectrum is due mainly to water vapour and oxygen.

All the energy incident on the earth's surface must be radiated back into space in the form of heat. Water vapour, carbon dioxide and ozone help to absorb part of the escaping heat radiation. Therefore the average temperature on earth is +15 °C and about 35 degrees warmer than if these absorbing greenhouse gases had not been present. At present, the composition of these and other greenhouse gases in the atmosphere is changing and these changes could have dramatic consequences for the future radiation environment and climate on earth. Because of deforestation and fossil fuel usage, the relative concentration of CO<sub>2</sub> has increased from the fraction, 0.028 percent during pre-industrial times to 0.0347 percent today. In 1958 the measured value was 0.0315 which gives an idea of the rate of increase. Presently five billion tons of carbon, 1000 kg for each person on earth, in the form of CO<sub>2</sub> is produced annually from fossil fuel burning. Half of this remains in the atmosphere and the other half is reabsorbed in vegetation and oceans.

The change in average global temperature is estimated by the use of sophisticated computer codes based on models which are necessarily very approximate. Calculated temperature increases are in the range 1.5 to 5.5 °C for a doubling of the concentration of CO<sub>2</sub> [2]. The increase in temperature is predicted to be higher above land areas than above the sea where the interaction between atmosphere and water moderates the temperature changes. The assumptions about the clouds are very critical since different kinds of clouds at different altitudes affect the radiation balance differently. Some models even predict a cooling due to the changing composition of the atmosphere.

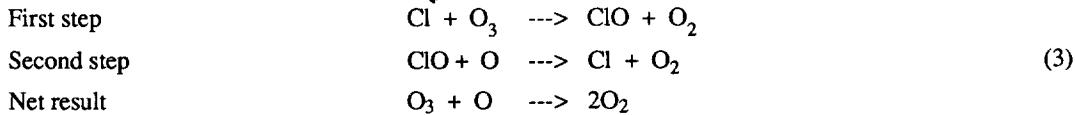
Large scale production of other substances such as freon, methane and nitrous oxide also affect the greenhouse effect. These gases are believed to contribute as much to the temperature increase as CO<sub>2</sub>. In addition to their contribution to the greenhouse effect, the freon molecules most probably destroy the high-altitude ozone layer which protects us from solar UV radiation. There are thus many reasons for reducing drastically the use of freon which will remain for prolonged times in the atmosphere because it does not combine readily with other substances.

A substantial portion of the solar radiation consists of ultraviolet light sufficiently energetic to ionize and act on the hereditary material of the living cell. Fortunately the energetic photons of ultraviolet light are absorbed to a great extent by atmospheric ozone concentrated at altitudes between 20 and 30 km.

Freon, otherwise known as a chlorofluorocarbon, CFC, is a man-made gas and one of the most inert gases in the atmosphere. It consists of fluorine (F), chlorine (Cl) and carbon (C) bound together by strong covalent bonds. The strength of the CF bond is 440 kJ/mole and that of the CCl bond 330. CFCs do not react easily with other gases and are non-toxic. They are hence ideal as propellants in spray cans, as cleaning solvents for microelectronic devices, for making plastic foam, and as cooling media in refrigerators and air conditioners. Since they are very inert, CFCs accumulate in the atmosphere. Only when reaching the upper part of the stratosphere do they disintegrate as a result of the energetic ultraviolet radiation which at lower altitudes has been absorbed efficiently by the layer of ozone. This layer is concentrated at altitudes between 25 and 30 km and protects life by shielding the earth from energetic ultraviolet radiation.

CFCs which initially were considered to be harmless are now thought to be among the most dangerous compounds in our environment since they cause large quantities of both ozone and atomic oxygen to be removed from the stratosphere. Atomic oxygen which can exist at reduced pressure is useful since it fuses with molecular oxygen to make up for natural loss of ozone. The rate of loss of ozone due to CFCs and other gases such as nitrous oxide N<sub>2</sub>O and methane CH<sub>4</sub> assumes enormous proportions, as evidenced by the increasing hole in the ozone layer above the antarctic.

Ironically, the CFCs are not very kind to their protector, the stratospheric ozone. Chlorine atoms liberated from CFCs by ultraviolet radiation in the upper stratosphere are very reactive and destroy ozone without being themselves removed. A chlorine atom first "steals" an oxygen atom from ozone forming chlorine monoxide (ClO) and an oxygen molecule. Then chlorine monoxide reacts with a free oxygen atom so that chlorine and molecular oxygen are formed. The net result is that one ozone molecule and one ozone atom merge into two oxygen molecules.



The same chlorine atom may react hundreds of times thereby consuming large quantities of ozone. However other processes interfere with the catalytic cycle shown in the formulae. For example, chlorine and methane form hydrochloric acid (HCl) while chlorine monoxide (ClO) and nitrogen dioxide (NO<sub>2</sub>) form chlorine nitrate (ClONO<sub>2</sub>). Both HCl and ClONO<sub>2</sub> represent relatively inert compounds which do not attack O<sub>3</sub> and serve as a temporary chlorine reservoir. Hence these interfering reactions reduce the efficiency of ozone destruction by the chlorine cycle particularly in the lower stratosphere.

It seems that polar stratospheric clouds set the scene for a chlorine catalyzed ozone destruction by stimulating the conversion of the major chlorine-reservoir species (HCl and ClONO<sub>2</sub>) to active chlorine (Cl and ClO). These clouds are formed in the polar stratosphere during winter at very low temperatures (around 185 - 195 K) through condensation of nitric acid and water vapours. The HNO<sub>3</sub> stratosphere above the South Pole is

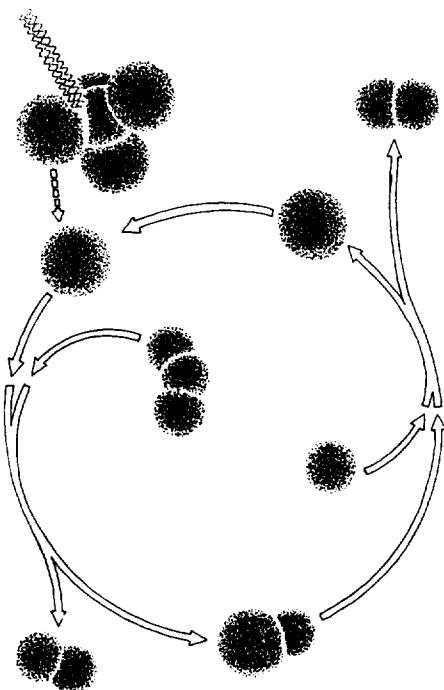


Fig. 6. One possible structure of CFC is  $\text{CF}_2\text{Cl}_2$ . The covalent bond breaks on absorption of a photon of ultraviolet light. The released chlorine atom is oxidized by an ozone molecule. The chlorine monoxide reacts in a second step with atomic oxygen so that molecular oxygen and a free chlorine atom is produced. The same chlorine atom repeats this cycle with the result that hundreds of molecules of ozone may be destroyed.

slightly colder than above the North Pole and therefore the ozone reduction started probably at the South Pole. A critical question is whether or not the ozone depletion will remain restricted to the South Pole or whether the antarctic ozone hole is a first sign of a global erosion of the stratospheric ozone layer. The scenario of a chlorine-catalysed ozone destruction may be influenced also by other atmospheric gases like  $\text{CO}_2$ ,  $\text{CH}_4$  and  $\text{N}_2\text{O}$  which result from combustion and agricultural processes. These gases presently accumulate in the air and cool the stratosphere by emitting infrared radiation and hence they may lower stratospheric temperatures. In fact, as seems probable at present, if the reduction of ozone is favoured by low temperatures, a positive feedback may occur due to the mentioned atmospheric gases.

Initially regarded as harmless, CFCs are now considered to be among the most dangerous compounds in our environment since they erode the ozone layer of the stratosphere. A first manifestation of a reduction of the stratospheric ozone occurred in the late seventies, when at the South Pole, a hole in the ozone layer started to develop each spring. Since then this Antarctic "ozone hole" has returned every spring and has increased in size. In 1987 its area was comparable to that of the Antarctic continent or the United States. There is increasing evidence that the Antarctic ozone hole is caused primarily by the chlorine cycle. Although many questions concerning the formation of the ozone hole are still open to debate, the potential risk of a global erosion of the stratospheric ozone layer has to be taken very seriously. This is particularly true since the residence times of the most important CFCs are long. In other words, even if the release of these substances were stopped today, they could still develop their destructive potential for decades to come. Obviously it would be dangerous to ask for an ultimate proof, namely a significant global reduction of stratospheric ozone before taking serious measures against the release of CFCs.

#### 4. RADIATION SERVING MEDICINE

Radiation becomes more and more important in medicine for diagnosis and for therapy [3]. Tomographic imaging based on X-rays, Nuclear Magnetic Resonance, NMR, and Positron Emission Tomography, PET, give complementary information. In PET, a short-lived positron emitting nuclide such as carbon-11 is first produced from an accelerator and by chemical methods subsequently attached to suitable molecules. These labelled molecules are then injected into the body and positrons emitted continuously for a few hours when the molecules spread and interact with the organs. After having passed a few millimetres in tissue, the positron annihilates with an electron and the two resulting product photons move in opposite directions and are recorded in detectors external to the body. Space lines are reconstructed from detector pairs hit by the photons giving an image of molecular activity. For example images of the metabolism of glucose can be followed as a function of time. Many other different substances can be labelled and traced and it is quite appropriate to say that PET is an effective method for mapping the biochemistry of the body.

Mapping of the atomic structure of large biomolecules is done by means of diffraction techniques using X-rays and neutrons as radiation particles. An exciting scattering experiment was made using a virus as the target for synchrotron radiation from the Cornell synchrotron [4]. A human common cold virus, rhinovirus-14, was used and assembled into a crystallized structure used as the target. The analysis of the very complex pattern of the scattering radiation was done on a supercomputer. The 3D atomic structure, the first ever measured of an animal virus, reveals narrow canyons in the spherical shell of the virus. The canyon is about 2.5 nm deep and may be the site for cell-receptor binding. An antibody molecule whose fragment would have a size of the order of 3.5 nm would not be able to reach the canyon floor. Thus the structure in the deeper part of the canyon would not be prone to immune selection and could remain constant, permitting the virus to retain its ability to seek the same cell receptor.

Research is presently in progress to find inhibitor molecules (virus antibiotics) that are small enough to reach the canyon floor. These inhibitors have either to block the receptor site or prevent ions from entering a channel into the virus thereby preventing it from disassembling and pouring its RNA content into the cell.

It is expected that other important animal viruses of similar structure, e.g. polio and AIDS, could also be known at the atomic level. The polio viruses exist in three different forms (sero types) and are combatted by vaccines. Against the 89 different forms of cold viruses it is not possible to find an effective vaccine and some sort of antibiotics adopted to their atomic structure may evolve as a result of scattering experiments. It will be fascinating to follow how insight at the atomic level of such a large structure as a virus may give clues to some of our common diseases.

Nearly half of all cancer cases are treated by radiation (X-rays and  $\gamma$ -rays) which is as frequent as conventional surgery and more frequent than chemotherapy.

Early work in Berkeley and Uppsala [5] showed the advantage of using protons for medical treatment. A medium-energy proton of 200 MeV has a range of 25 cm in tissue and can reach all points in the human body. The differential ionization and hence the cell damage is largest just prior to the protons coming to rest. In this respect, protons are different from X-rays and neutrons which are attenuated as they penetrate deeper and deeper. Proton beam surgery, for the affection of small regions of the brain, initiated 30 years ago in Uppsala, and

irradiation of small tumours in the eye carried out routinely in several laboratories are two examples of precise radiotherapy in which medium-energy protons are preferred to other radiation particles. Small accelerators for electrons and protons are now in frequent use at hospitals. At many accelerator centres, part of the program is allocated to medical applications. In Harvard, Fermilab, KEK Tsukuba, TSL Uppsala and in three Russian laboratories, INP Gatchina, ITEP Moscow and JINR Dubna, protons of medium energy are used for surgery and therapy. The first medium energy proton accelerator fully dedicated to medicine, a proton synchrotron, is being built presently at the Loma Linda University hospital in California.

##### 5. RADIATION EMITTED BY PARTICLES ACCELERATED IN THE SKY

The Northern Light, or the Auroral Light as it is also called, is a spectacular phenomenon seen normally at northern latitudes near the pole. Its origin is to be found in the sun's "wind" which consists of slow electrons and protons emitted from its corona. At an altitude of some 500 km at the top of the atmosphere there are electrically charged ionized layers with potential differences of 10 000 V. At the North and South poles, the electrons can travel parallel to the magnetic field and penetrate a long distance into the atmosphere. They are accelerated to an energy of some 10 000 eV since they are not deflected by the vertically directed magnetic field. As the density increases, the electrons collide more and more frequently with atoms and molecules of the atmosphere. In these collisions, energy is transferred to the atoms and molecules which are excited and carry excess energy for a short



Fig. 7 Auroral Light over northern Scandinavia seen from 800 km altitude by an American weather satellite. The energy in the Auroral Light corresponds to the energy consumption of Europe.  
Photo: Air Force Geophysical Lab, USA.

while. However, after some billionth of a second, they decay to their stationary states by emitting light. At an altitude of 100 km, the density of molecules and atoms is sufficiently great to cause visible effects on the earth's surface. The blue-green colour, characteristic of the Northern Light is emitted primarily when excited oxygen returns to the normal ground state.

High energy electrons moving in the magnetic fields of galaxies give rise to radiation that may be measured on earth. This form of radiation is named synchrotron radiation since it was seen first in synchrotrons. Electrons

orbiting in a synchrotron lose energy continually in the form of photons emitted tangentially from the orbit. The energy lost every turn by an electron can be computed from the expression:

$$\Delta E = \frac{6 \times 10^{-15}}{R} \left[ \frac{E}{m_0 c^2} \right]^4$$

where  $\Delta E$  is the energy loss per turn,  $m_0 c^2$  the rest energy of the particle,  $E$  its total energy and  $R$  the radius of curvature. The units used are MeV and metre. Since the mass of the electron is 2000 times less than that of the proton, an electron of given energy and radius of curvature loses much more energy than a corresponding proton. Radiogalaxies are gigantic very remote objects in the universe and they emit synchrotron radiation which is detected using large radiotelescopes. Since radio waves easily pass through galactic and intergalactic matter, the intensity of the radiation from radiogalaxies is not absorbed as strongly as visible light.

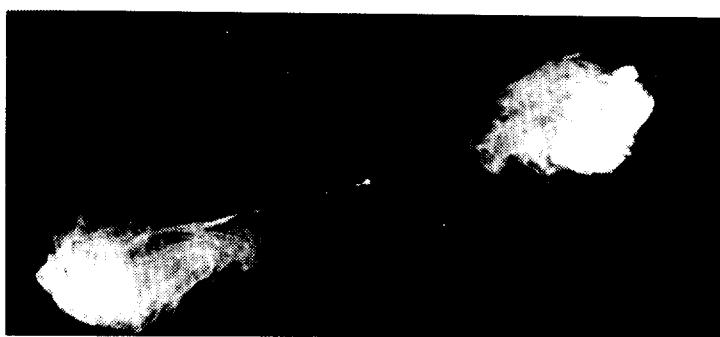


Fig. 8. Radiography of the radiogalaxy Cygnus A. The picture has been made by means of the very large array, VLA. The radiation emitted from the galaxy in the centre interacts with the intergalactic gas and two clouds (lobes) are seen to the left and right of the mother galaxy seen as a small dot in the centre of the picture. Photo: R A Perley, J W Dreher and J J Cowan, The National Radio Astronomy Observatory, Ani, USA.

Microwaves may also be emitted from excited molecules in the remote cosmos and a consequence of changes in the rotational states of excited molecules. The radiation emitted when excited molecules lose their energy appears as lines in a spectrum of wavelengths. Any given species of a molecule can be identified by a number of lines which form a characteristic "finger print" of the molecule.

In 1963 the first molecule, OH<sup>-</sup> was discovered in the interstellar space. Later discoveries include carbon monoxide which is quite abundant and is present in our own galaxy, in a concentration of one part in 10 000 of the hydrogen content. The structure of other galaxies is often studied using carbon monoxide and atomic and molecular hydrogen as radiation sources. Observations of radio waves from space molecules are of increasing importance in astrophysics and, in certain cases, new information about the molecules themselves may be inferred from the radiation observed. One example is the molecular ion NH<sub>2</sub><sup>+</sup> which is so reactive that it cannot be isolated on earth. Among the numerous known space molecules, ethyl alcohol was discovered in 1974. The molecular cloud Sagittarius, B2 could contain as much alcohol as is present in 10<sup>28</sup> bottles of whisky !

#### 6. STIMULATED CYCLOTRON RADIATION

Accelerator science requires knowledge of both classical and quantum physics which is illustrated by a new scheme for fast cooling of particle beams presented recently by Ikegami [6]. Cooling times of the order of a

microsecond are predicted. In the proposed scheme, electrons or ions are cooled very rapidly by making use of the principle behind the cyclotron maser [7]. Here, electrons of a few keV circulate in a strong magnetic field under the influence of an electric rf field. If the frequencies of the cyclotron motion and of the electric-field change are nearly identical the electrons may be stimulated to emit large amounts of radiation. The emitted radiation can become much stronger than the radiation absorbed by the stimulating rf field and this condition is used when designing high-power cyclotron maser amplifiers.

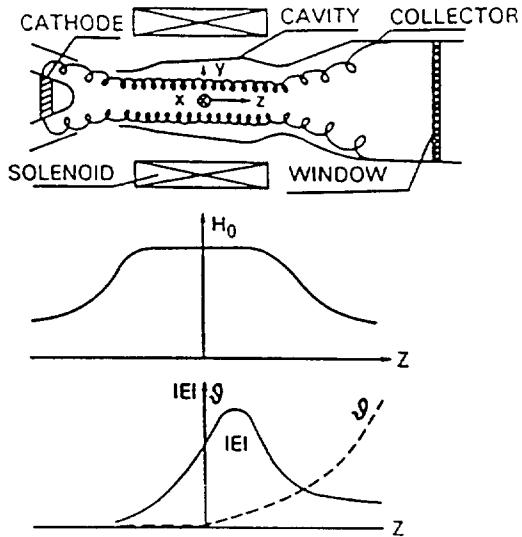


Fig. 9. Arrangement of the gyrotron. Distribution of static  $H_0$  and alternating  $E = |E| [e^{i\omega t - iv}]$  fields.  
From Ref [7].

The idea of stimulated cyclotron radiation was developed originally by Twiss [8] in a paper that described "negative" absorption at radio wavelengths where the medium behaves like an amplifier. Twiss suggested that cyclotron radiation is much more probable at radio than at optical frequencies and may be a common phenomenon in the universe. The mechanism can be understood quantum mechanically as transitions between Landau states that are states associated with the circular motion and separated by the energy  $E = hf$ . The frequency,  $f$ , of the orbiting particle is inversely proportional to its total energy. An external rf field may stimulate emission of photons by a circulating particle or give away photons to the particle. The level spacing is proportional to the particle circulation frequency which is dependent on its total energy. A quantum emitted by the particle in a given level therefore has somewhat higher energy than a quantum being absorbed by the particle in that level. This fact explains why under certain conditions the energy lost by the particle due to stimulated emission may exceed the energy received due to absorption. A quantitative formula for the magnitude of the effect was developed by Schneider [9].

Ikegami suggested that the stimulated emission of radiation from particles in cyclotron orbit can be utilized for the cooling of high-energy particle beams. It is necessary to have a strong axial magnetic field along the main direction of particle motion. In this way, the particles move in helical orbits. The transverse motion is affected by an rf field that acts transversely to the main direction of particle motion. The loss of energy due to the radiation gives rise to a friction term in the equation for the transverse circular motion of the particle.

The idea can be tried at a storage ring such as CELSIUS. A portion of a straight section of a ring must have a strong solenoidal magnetic field in the same direction as the particle beam. In the solenoid, the particles move in a cork-screw orbit and the frequency of the electric rf field applied along a transverse direction has to be the same as the frequency of particle circulation in the transverse plane.

## 7. CONCLUSION

Accelerator science has developed into its own discipline on the border line between classical and modern physics and closely related to plasma-, magnetosphere- and ionosphere-physics. The deep experience gained in accelerator laboratories on how particles can be accelerated, made to radiate and interact with matter is a very important basis for exploration of the microcosmos and for the technology of today and tomorrow.

The examples discussed in this paper show also that knowledge of radiation is much more important in everyday life than is generally believed. Such knowledge will guide us in devising future nuclear and solar energy sources, recyclable material such as wood, ceramics, glass and food produced in harmony with the environment.

## ACKNOWLEDGEMENTS

I wish to thank F. Arnold, H. Ikegami, B. Larsson and W. Mair for many useful discussions on topics related to the present paper. I am grateful also to I. Ericson for assistance with the manuscript.

\* \* \*

## REFERENCES

- [1] F.M. Luther and R.G. Ellingson, Carbon Dioxide and the Radiation Budget, US Dep. of Energy, Report DOE/ER-0237, December 1985, p. 27
- [2] R.T. Watson, M.A. Geller, R.S. Stolarski and R.F. Hampson, Present State of Knowledge of the Upper Atmosphere: An assessment Report , Processes That Control Ozone and Other Climatically Important Trace Gases, NASA-Report RP-1162, May 1986
- [3] S. Kullander, Science, Society and Microcosmos, Proceedings of CERN Accelerator School, Aarhus, CERN 87-10, July 1987
- [4] M. G. Rossman et al., Structure of a Human Cold Virus and Functional Relationship to Other Picornaviruses, Nature 317 (1985) 145
- [5] B. Larsson, L. Leksell, B. Rexed, P. Sourander, W. Mair and B. Andersson, The High-Energy Proton Beam as a Neurosurgical Tool, Nature 182 (1958) 1222
- [6] H. Ikegami, Cyclotron Maser Cooling of Electron and Ion Beams, Proceedings of the Twelfth International Conference on Cyclotrons and Their Applications, Berlin (West), ed. by K. Ziegler (World Scientific, Singapore 1989)
- [7] V.A. Flyagin, A.V. Gaponov, M.I. Petelin and V.K. Yulpatov, The Gyrotron, IEEE Trans. on Microwave Theory and Techniques, MTT-25, 6, June 1977

- [8] R.Q. Twiss, Radiation Transfer and the Possibility of Negative Absorption in Radio Astronomy, *Aust. J. Phys.* 11 (1958) 564
- [9] J. Schneider, Stimulated Emission of Radiation by Relativistic Electrons in a Magnetic Field, *Phys. Rev. Lett.* 2 (1959) 504

**BIBLIOGRAPHY**

S. Kullander and B. Larsson, *Radiation and Matter*, Cambridge University Press, to be published

B. Bohlin, B.R. Döös, J. Jäger and R.A. Warrick, *The Greenhouse effect, Climatic change and Ecosystems*, J. Wiley & Sons, 1986

### LIST OF PARTICIPANTS

ALVES-PIRES, R.	CERN, Geneva, Switzerland
ANDLER, G.	Manne Siegbahn Inst. , Stockholm, Sweden
ANGERT, N.	GSI, Darmstadt, Fed. Rep. of Germany
AUDY, P.	CEN-Saclay, Gif-sur-Yvette, France
BARRATT, P.	RAL, Didcot, United Kingdom
BERGMARK, T.	The Svedberg Lab., Uppsala, Sweden
BIELER, M.	DESY, Hamburg, Fed. Rep. of Germany
BLIND, B.	LANL, Los Alamos, USA
BONGARDT, K.	KFA, Jülich, Fed. Rep. of Germany
BRAUN, H.-H.	PSI-West, Villigen, Switzerland
CASPERS, F.	CERN, Geneva, Switzerland
DECKING, W.	DESY, Hamburg, Fed. Rep. of Germany
DUVAL, P.	DESY, Hamburg, Fed. Rep. of Germany
FINLAYSON, K.	GSI, Darmstadt, Fed. Rep. of Germany
FITZE, H.	PSI, Villigen, Switzerland
FITZGERALD, D.	LANL, Los Alamos, USA
FLOETTMANN, K.	DESY, Hamburg, Fed. Rep. of Germany
GAJEWSKI, K.	The Svedberg Lab., Uppsala, Sweden
GYR, M.	CERN, Geneva, Switzerland
HAMAD, H.A.	SRC, Baghdad, Iraq
HANCOCK, S.	CERN, Geneva, Switzerland
HEINEMANN, K.	DESY, Hamburg, Fed. Rep. of Germany
HERMANSSON, L.	The Svedberg Lab., Uppsala, Sweden
HERR, W.	CERN, Geneva, Switzerland
HERRLANDER, C.J.	Manne Siegbahn Inst., Stockholm, Sweden
HERZOG, M.	Manne Siegbahn Inst., Stockholm, Sweden
HUSSIAN, M.	SRC, Baghdad, Iraq
JAHNEL, L.	LNLS, Campinas, Brazil
KALLBERG, A.	Manne Siegbahn Inst., Stockholm, Sweden
KELLER, R.	LBL, Berkeley, USA
KLEIN, E.	CEN-Saclay, Gif-sur-Yvette, France
KO, I.S.	Pohang Light Source, Pohang, Rep. of Korea
KRAUS, R.	LANL, Los Alamos, USA
LAGNIEL, J.-M.	CEN-Saclay, Gif-sur-Yvette, France
LEMAIRE, J.-L.	CEN-Saclay, Gif-sur-Yvette, France
LILJEBY, L.	Manne Siegbahn Inst., Stockholm, Sweden
LINDHOLM, M.	Scanditronix, Uppsala, Sweden
LOFNES, T.	The Svedberg Lab., Uppsala, Sweden
MAAS, R.	NIKHEF-K, Amsterdam, Netherlands
MALANDAIN, H.	CERN, Geneva, Switzerland

MANGLUNKI, D.	CERN, Geneva, Switzerland
MARTI, F.	Michigan State Univ., East Lansing, USA
MAYER, R.	Dornier, Friedrichshafen, Fed. Rep. of Germany
MEUTH, H.	KFA, Jülich, Fed. Rep. of Germany
MUELLER, R.	BESSY, Berlin, Fed. Rep. of Germany
NORMAN, G.	The Svedberg Lab., Uppsala, Sweden
OLSEN, D.	Oak Ridge Nat. Lab., Oak Ridge, USA
PACE, A.	CERN, Geneva, Switzerland
PAYET, J.	CEN-Saclay, Gif-sur-Yvette, France
PERETTI, A.	INFN - Univ. of Milan, Milan, Italy
PLESKO, M.	Sincrotrone Trieste, Trieste, Italy
PLUM, M.	LANL, Los Alamos, USA
PRUSS, S.	Fermilab, Batavia, USA
RENSFELT, K.-G.	Manne Siegbahn Inst., Stockholm, Sweden
RIFUGGIATO, D.	LNS-INFN, Catania, Italy
RINTA-NIKKOLA, M.	Univ. of Jyvaskyla, Jyvaskyla, Finland
RIPKEN, G.	DESY, Hamburg, Fed. Rep. of Germany
ROJSEL, P.	Max-lab, Lund, Sweden
ROSENGARD, U.	Manne Siegbahn Inst., Stockholm, Sweden
SCHMIDT, M.	Fraunhofer-Inst., Berlin, Fed. Rep. of Germany
SCHUETTE, W.	DESY, Hamburg, Fed. Rep. of Germany
SERLIN, V.	Naval Research Lab., Washington, USA
SPAEDTKE, P.	GSI, Darmstadt, Fed. Rep. of Germany
STECK, M.	MPI, Heidelberg, Fed. Rep. of Germany
STEINBACH, C.	CERN, Geneva, Switzerland
STEINHOF, A.	GSI, Darmstadt, Fed. Rep. of Germany
STOCKHORST, H.	KFA, Jülich, Fed. Rep. of Germany
STREUN, A.	Inst. f. Kernphysik, Mainz, Fed. Rep. of Germany
STRUCKMEIER, J.	GSI, Darmstadt, Fed. Rep. of Germany
UYTHOVEN, J.	Oxford University, Oxford, United Kingdom
VANDEPLASSCHE, D.	CERN, Geneva, Switzerland
VEUILLEN, J.-M.	ESRF, Grenoble, France
VIEIRA, W.M.	LNLS, Campinas, Brazil
von HARTROTT, M.	BESSY, Berlin, Fed. Rep. of Germany
WANZENBERG, R.	THS, Darmstadt, Fed. Rep. of Germany
WESTERBERG, L.	The Svedberg Lab., Uppsala, Sweden
WU, H.	DESY, Hamburg, Fed. Rep. of Germany
WU, Y.	NIKHEF-K, Amsterdam, Netherlands
YANAGIDA, K.	Japan Atomic Energy Res. Inst., Ibaraki, Japan
YOSHII, M.	Nat. Lab. for HEP, KEK, Ibaraki, Japan

### LIST OF PARTICIPANTS

ALVES-PIRES, R.	CERN, Geneva, Switzerland
ANDLER, G.	Manne Siegbahn Inst. , Stockholm, Sweden
ANGERT, N.	GSI, Darmstadt, Fed. Rep. of Germany
AUDY, P.	CEN-Saclay, Gif-sur-Yvette, France
BARRATT, P.	RAL, Didcot, United Kingdom
BERGMARK, T.	The Svedberg Lab., Uppsala, Sweden
BIELER, M.	DESY, Hamburg, Fed. Rep. of Germany
BLIND, B.	LANL, Los Alamos, USA
BONGARDT, K.	KFA, Jülich, Fed. Rep. of Germany
BRAUN, H.-H.	PSI-West, Villigen, Switzerland
CASPERS, F.	CERN, Geneva, Switzerland
DECKING, W.	DESY, Hamburg, Fed. Rep. of Germany
DUVAL, P.	DESY, Hamburg, Fed. Rep. of Germany
FINLAYSON, K.	GSI, Darmstadt, Fed. Rep. of Germany
FITZE, H.	PSI, Villigen, Switzerland
FITZGERALD, D.	LANL, Los Alamos, USA
FLOETTMANN, K.	DESY, Hamburg, Fed. Rep. of Germany
GAJEWSKI, K.	The Svedberg Lab., Uppsala, Sweden
GYR, M.	CERN, Geneva, Switzerland
HAMAD, H.A.	SRC, Baghdad, Iraq
HANCOCK, S.	CERN, Geneva, Switzerland
HEINEMANN, K.	DESY, Hamburg, Fed. Rep. of Germany
HERMANSSON, L.	The Svedberg Lab., Uppsala, Sweden
HERR, W.	CERN, Geneva, Switzerland
HERRLANDER, C.J.	Manne Siegbahn Inst., Stockholm, Sweden
HERZOG, M.	Manne Siegbahn Inst., Stockholm, Sweden
HUSSIAN, M.	SRC, Baghdad, Iraq
JAHNEL, L.	LNLS, Campinas, Brazil
KALLBERG, A.	Manne Siegbahn Inst., Stockholm, Sweden
KELLER, R.	LBL, Berkeley, USA
KLEIN, E.	CEN-Saclay, Gif-sur-Yvette, France
KO, I.S.	Pohang Light Source, Pohang, Rep. of Korea
KRAUS, R.	LANL, Los Alamos, USA
LAGNIEL, J.-M.	CEN-Saclay, Gif-sur-Yvette, France
LEMAIRE, J.-L.	CEN-Saclay, Gif-sur-Yvette, France
LILJEBY, L.	Manne Siegbahn Inst., Stockholm, Sweden
LINDHOLM, M.	Scanditronix, Uppsala, Sweden
LOFNES, T.	The Svedberg Lab., Uppsala, Sweden
MAAS, R.	NIKHEF-K, Amsterdam, Netherlands
MALANDAIN, H.	CERN, Geneva, Switzerland

MANGLUNKI, D.	CERN, Geneva, Switzerland
MARTI, F.	Michigan State Univ., East Lansing, USA
MAYER, R.	Dornier, Friedrichshafen, Fed. Rep. of Germany
MEUTH, H.	KFA, Jülich, Fed. Rep. of Germany
MUELLER, R.	BESSY, Berlin, Fed. Rep. of Germany
NORMAN, G.	The Svedberg Lab., Uppsala, Sweden
OLSEN, D.	Oak Ridge Nat. Lab., Oak Ridge, USA
PACE, A.	CERN, Geneva, Switzerland
PAYET, J.	CEN-Saclay, Gif-sur-Yvette, France
PERETTI, A.	INFN - Univ. of Milan, Milan, Italy
PLESKO, M.	Sincrotrone Trieste, Trieste, Italy
PLUM, M.	LANL, Los Alamos, USA
PRUSS, S.	Fermilab, Batavia, USA
RENSFELT, K.-G.	Manne Siegbahn Inst., Stockholm, Sweden
RIFUGGIATO, D.	LNS-INFN, Catania, Italy
RINTA-NIKKOLA, M.	Univ. of Jyvaskyla, Jyvaskyla, Finland
RIPKEN, G.	DESY, Hamburg, Fed. Rep. of Germany
ROJSEL, P.	Max-lab, Lund, Sweden
ROSENGARD, U.	Manne Siegbahn Inst., Stockholm, Sweden
SCHMIDT, M.	Fraunhofer-Inst., Berlin, Fed. Rep. of Germany
SCHUETTE, W.	DESY, Hamburg, Fed. Rep. of Germany
SERLIN, V.	Naval Research Lab., Washington, USA
SPAEDTKE, P.	GSI, Darmstadt, Fed. Rep. of Germany
STECK, M.	MPI, Heidelberg, Fed. Rep. of Germany
STEINBACH, C.	CERN, Geneva, Switzerland
STEINHOF, A.	GSI, Darmstadt, Fed. Rep. of Germany
STOCKHORST, H.	KFA, Jülich, Fed. Rep. of Germany
STREUN, A.	Inst. f. Kernphysik, Mainz, Fed. Rep. of Germany
STRUCKMEIER, J.	GSI, Darmstadt, Fed. Rep. of Germany
UYTHOVEN, J.	Oxford University, Oxford, United Kingdom
VANDEPLASSCHE, D.	CERN, Geneva, Switzerland
VEUILLEN, J.-M.	ESRF, Grenoble, France
VIEIRA, W.M.	LNLS, Campinas, Brazil
von HARTROTT, M.	BESSY, Berlin, Fed. Rep. of Germany
WANZENBERG, R.	THS, Darmstadt, Fed. Rep. of Germany
WESTERBERG, L.	The Svedberg Lab., Uppsala, Sweden
WU, H.	DESY, Hamburg, Fed. Rep. of Germany
WU, Y.	NIKHEF-K, Amsterdam, Netherlands
YANAGIDA, K.	Japan Atomic Energy Res. Inst., Ibaraki, Japan
YOSHII, M.	Nat. Lab. for HEP, KEK, Ibaraki, Japan