CERN Accelerator School
Intermediate accelerator physics

DESY, Zeuthen, Germany
15 – 26 September 2003

Proceedings
Editor: D. Brandt
Abstract

The fifth CERN Accelerator School (CAS) Intermediate course on Accelerator Physics was held at the DESY-Zeuthen Laboratory, Zeuthen, Germany from 15 to 26 September 2003. Its syllabus was based on the decision to definitively associate the Intermediate level course with the concept of afternoon courses. Consequently, the programme was composed of lectures in the mornings (dedicated to core topics) and of three ‘specialization’ courses in the afternoons, whereby each participant selected one of the proposed subjects and followed this course every afternoon for a detailed tuition on this specific topic. This year, the three proposed afternoon courses were ‘Optics Design, Measurement and Correction’, ‘Instrumentation, Beam Diagnostics and Feedback Systems’, and ‘Linac Structures’. Keeping in mind that the afternoon courses are not meant to be lectures but rather a mixture of theory and experimental applications, it follows that it is not possible to incorporate these courses in the proceedings. Consequently, the proceedings present the lectures given during the morning sessions.

The lectures can be classified in three distinct categories: first, the core topics already introduced at the Introductory level but treated at a more advanced level, such as transverse and longitudinal dynamics, linear imperfections, vacuum, magnet design, particle sources, and beam instabilities. Secondly, the topics requiring some preliminary knowledge, and thus only introduced at the Intermediate level, such as non-linear effects, sources of emittance growth, Landau damping, dynamics with damping, insertion devices, luminosity and beam–beam effects. The third category is composed of plenary lectures directly related to the proposed afternoon courses, i.e., lattice cells, insertions, beam instrumentation and feedback, beam diagnostics, linac structures, and superconducting RF cavities. These 32 lectures were completed by three seminars, six tutorials with six hours of guided study, a 10-hour afternoon course, and a Poster session.
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. Until 1993, this was carried out principally by means of a biennial programme of introductory and advanced two-week courses on general accelerator physics aiming to bridge the gap between the level of knowledge attained with a university degree and that required for starting accelerator research work.

In 1995, it was recognized that the level of the advanced course was possibly too high and not necessarily adapted anymore to the needs of the young scientists entering the field of accelerator physics. Consequently, the advanced level course was replaced by an intermediate level course. The first Intermediate course was organized in 1995 in Eger, Hungary.

Although the present proceedings are for the fifth Intermediate course on general accelerator physics held in Zeuthen, Germany in September 2003, they are the first proceedings to be published as a CERN Report for a course at the intermediate level.

The intermediate level, which is clearly meant to be a logical follow-up of the introductory course, is composed of three distinct parts: first, some core topics (introduced at the introductory level) are reviewed at a more advanced level. Secondly, essential topics, which require some preliminary basic knowledge, are introduced (e.g., non-linear effects, Landau damping, beam–beam effects). Finally, the specificity of the Intermediate course, namely the ‘afternoon courses’, where each participant selects one of three proposed subjects and follows this course every afternoon for a detailed tuition on this specific topic.

The basic objective of the intermediate level is therefore not only that the participants follow regular lectures in the mornings, but that they also have an opportunity to acquire a ‘solid’ introduction to a topic which, ideally, should be new to them. This specialization should not be composed of lectures, but rather of a mixture of theory and experimental applications.

This course was made possible by the active support of several laboratories and many individuals. In particular, the help and collaboration of the DESY-Zeuthen Laboratory management and staff, especially Professor U. Gensch, K. Varschen, M. Mende, A. Hagedorn, and Dr. P. Wegner were most invaluable. Their endless enthusiasm and the provision of a first-class infrastructure for the lecture rooms definitely contributed to the success of the school.

As always, the backing of the CERN management, the guidance of the CAS Advisory and Programme Committees, and the attention to detail of the Local Organizing Committee (DESY-Zeuthen) ensured that the school was held under optimum conditions.

Very special thanks must go to the lecturers for the enormous task of preparing, presenting, and writing up their topics. Similarly, the enthusiasm and positive feedback of the participants was convincing proof that an Intermediate level course with afternoon courses was the appropriate complement to the Introductory level course.

Finally, we thank the CERN Scientific Text Processing Service for their dedication and commitment to the production of this document.
# Programme for the Intermediate CAS School in Zeuthen (September 2003)

<table>
<thead>
<tr>
<th>Time</th>
<th>Monday 15/9/03</th>
<th>Tuesday 16/9/03</th>
<th>Wednesday 17/9/03</th>
<th>Thursday 18/9/03</th>
<th>Friday 19/9/03</th>
<th>Saturday 20/9/03</th>
<th>Sunday 21/9/03</th>
<th>Monday 22/9/03</th>
<th>Tuesday 23/9/03</th>
<th>Wednesday 24/9/03</th>
<th>Thursday 25/9/03</th>
<th>Friday 26/9/03</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:30</td>
<td>Opening Talk</td>
<td>Transverse</td>
<td>Nonlinear</td>
<td>Introduction</td>
<td>Instabilities</td>
<td>Longitudinal</td>
<td>Transverse</td>
<td>Insertion</td>
<td>Students’</td>
<td>Devices</td>
<td>Feedback</td>
<td></td>
</tr>
<tr>
<td></td>
<td>U. Gensch</td>
<td>Dynamics</td>
<td>Imperfections</td>
<td>to Beam</td>
<td>in Linacs</td>
<td>Instabilities</td>
<td>Dynamics</td>
<td>Devices</td>
<td>Feedback</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Instabilities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>09:30</td>
<td>Recap. Transverse</td>
<td>Longitudinal</td>
<td>GS2</td>
<td>T2</td>
<td>T4</td>
<td>Space Charge</td>
<td>T5</td>
<td>GS6</td>
<td>T6</td>
<td>Instabilities</td>
<td>Closing Session</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dynamics</td>
<td>Dynamics</td>
<td>Transverse</td>
<td>Dynamics</td>
<td>Dynamics with</td>
<td>Instabilities</td>
<td>Dynamics</td>
<td>Dynamics with</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Dynamics</td>
<td></td>
<td>Damping</td>
<td></td>
<td></td>
<td>Damping</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:30</td>
<td>Recap.</td>
<td>Insertions</td>
<td>SC - RF Cavities</td>
<td>Landau Damping I</td>
<td>Space Charge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Longitudinal</td>
<td>Nonlinearities</td>
<td></td>
<td></td>
<td>Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dynamics</td>
<td>for LS’s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Landau Damping</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dynamics</td>
<td>Space Charge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:50</td>
<td>Recap.</td>
<td>SC - RF Cavities</td>
<td>Nonlinearities</td>
<td>Landau Damping</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Longitudinal</td>
<td>for</td>
<td>for LS’s</td>
<td>Damping</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dynamics</td>
<td>LS’s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11:50</td>
<td>Intro. Beam</td>
<td>Lattices for</td>
<td>Landau Damping</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Instrumentation</td>
<td>LSs</td>
<td>Damping II</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&amp; Feedback</td>
<td></td>
<td>Space Charge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:50</td>
<td>Recap.</td>
<td>C1</td>
<td>Course 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Transverse</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13:30</td>
<td>LUNCH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14:30</td>
<td>Lattice Cells</td>
<td>C1</td>
<td>Course 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15:30</td>
<td>B. Holzer</td>
<td>C2</td>
<td>Course 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Linac Structures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16:30</td>
<td>N. Pichoff</td>
<td>C3</td>
<td>Course 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17:00</td>
<td>School Organ.</td>
<td>GS1</td>
<td>Seminar</td>
<td>GS5</td>
<td>Seminar</td>
<td>GS5</td>
<td>Seminar</td>
<td>Seminar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Organisation</td>
<td></td>
<td>The Mother of all</td>
<td>Dynamics with</td>
<td>Physics made</td>
<td>Dynamics with</td>
<td>Physics made</td>
<td>Neutrino-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Accelerators</td>
<td>Damping</td>
<td>in Berlin</td>
<td>Damping</td>
<td>in Berlin</td>
<td>Astro-Physics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A. De Rujula</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C. Spiering</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18:00</td>
<td>D. Brandt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18:30</td>
<td>COCKTAIL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19:30</td>
<td>EVENING MEAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MEAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Course 1 = “Optics Design, Measurement and Correction”
Course 2 = “Linac Structures”
Course 3 = “Instrumentation, Beam Diagnostics and Feedback Systems”
## Contents

Foreword  
*D. Brandt* ...................................................... v

Transverse beam dynamics  
*E.J.N. Wilson* .................................................. 1

Lattice design in high-energy particle accelerators  
*B.J. Holzer* .................................................... 31

Introduction to beam instrumentation and diagnostics  
*R. Jones and H. Schmickler* ................................. 75

Introduction to RF linear accelerators  
*N. Pichoff* ..................................................... 105

Linear imperfections  
*O. Br"uning* .................................................... 129

Basic principles of RF superconductivity and superconducting cavities  
*P. Schm"user* .................................................. 183

Non-linearities in light sources  
*A. Streun* ....................................................... 203

Lattices for light sources  
*A. Streun* ....................................................... 217

Sources of emittance growth  
*D. M"ohl* ......................................................... 245

Landau damping  
*A. Hofmann* .................................................... 271

Space charge  
*K. Schindl* ....................................................... 305

Instabilities  
*K. Schindl* ....................................................... 321

Wake fields and instabilities in linear accelerators  
*M. Ferrario, M. Migliorati and L. Palumbo* ......... 343

Concept of luminosity  
*W. Herr and B. Muratori* ..................................... 361

Beam–beam interactions  
*W. Herr* ......................................................... 379

Electromagnetic design of accelerator magnets  
*R. Russenschuck* .............................................. 411

Insertion devices  
*J. Bahrdt* ......................................................... 441

Free-electron lasers  
*P. Schm"user* .................................................. 477

Electron and ion sources for particle accelerators  
*R. Scrivens* ...................................................... 495

A MAD-X primer  
*W. Herr* ......................................................... 505

List of Participants ............................................... 529
Abstract
This contribution describes the transverse dynamics of particles in a synchrotron. It builds on other contributions to the General Accelerator School for definitions of transport matrices and lattice functions. After a discussion of the conservation laws which govern emittance, the effects of closed orbit distortion and other field errors are treated. A number of practical methods of measuring the transverse behaviour of particles are outlined.

1. INTRODUCTION

During the design phase of an accelerator project a considerable amount of calculation and discussion centres around the choice of the transverse focusing system. The lattice, formed by the pattern of bending and focusing magnets, has a strong influence on the aperture of these magnets which are usually the most expensive single system in the accelerator.

Figure 1 shows the lattice design for the SPS at CERN. The ring consists of a chain of 108 such cells. The principles of designing such a lattice have been treated elsewhere in this school by Rossbach and Schmüser [1] and the methods of modifying the regular pattern of cells to make matched insertions where space is needed for accelerating structures, extraction systems and for experiments where colliding beams interact, will be the subject of another talk [2].

Fig. 1 One cell of the CERN 400 GeV Super Proton Synchrotron representing 1/108 of the circumference. The pattern of dipole and quadrupole magnets (F and D) is shown above. Beam particles make betatron oscillations within the shaded envelopes.
In this series of lectures I shall concentrate on those aspects of transverse dynamics which frustrate the designer or plague the person whose job it is to make the machine perform as well as theory predicts. I shall start with Liouville's theorem because this is an inescapable limitation to the beam size. Next I shall explain the distortion of the central orbit of the machine due to errors in the bending field. I shall then move on to explain how errors in quadrupole gradients may lead to instability of the betatron motion at certain $Q$ values known as stopbands. Finally I shall explain why it is necessary to correct the variation of $Q$ with momentum using sextupoles in order to avoid these stopbands.

The more advanced topics of the influence of fields which couple the two transverse phase planes together and fields whose non-linear nature can excite resonances which extract particles are given in two separate talks in this school [3,4]. In those talks we shall move away from the predictable linear behaviour to glimpse the fascinating jungle of effects which stretch our minds at the boundaries of accelerator theory and which will be fully developed in the Advanced Course of this school.

Since the effects I shall mention are annoying and best eliminated I shall try to indicate how measurements may be made on a circulating beam and its transverse behaviour corrected. Many of the effects I shall describe need only to be estimated rather than accurately predicted but it is very important to understand their mechanism. I shall therefore prefer to introduce concepts with approximate but physically revealing theory.

I make no apologies for building the theory from elementary definitions. These have to be restated to give those not fortunate enough to have eaten and slept in phase space a firm basic understanding. I hope that others who become impatient with the redefinition of basic quantities will be eventually gratified by the later stages in the exposition of the theory and the references I have given for further study.

2. LIOUVILLE'S THEOREM

![Diagram of ellipse illustrating Liouville's theorem](image)

Fig. 2 Liouville's theorem applies to this ellipse

Particle dynamics obey a conservation law of phase space called Liouville’s theorem. A beam of particles may be represented in a phase space diagram as a cloud of points within a closed contour, usually an ellipse (Fig. 2).
The area within the contour is proportional to the emittance of the beam. At constant energy we express this as the area \( \epsilon = \int y\,dy \) in units of \( \pi \text{ mm mmiradians} \). We have shown the contour at a place where the \( \beta \) function is at a maximum or minimum and where the major and minor axes of an upright ellipse are \( \sqrt{\epsilon\beta} \) and \( \sqrt{\epsilon/\beta} \).

This emittance is conserved, whatever magnetic focusing or bending operation we do on the beam. This is a consequence of the conservation of transverse momentum or, more strictly, a consequence of Liouville's theorem which states that:

"In the vicinity of a particle, the particle density in phase space is constant if the particles move in an external magnetic field or in a general field in which the forces do not depend upon velocity."

Clearly this definition rules out the application of Liouville's theorem to space charge problems or when particles emit synchrotron light. However, with these exceptions, it allows us to be confident that though the beam cross section may take on many shapes around the accelerator, its phase space area will not change (Fig. 3). Thus if the beam is at a narrow waist (a) its divergence will be large. In an F quadrupole (d) where the betatron function is maximum, its divergence will be small and a small angular deflection can greatly dilute phase space.

Fig. 3 The development of a constant emittance beam in phase space at a) a narrow waist, b) a broad waist or maximum in the beta function, c) place where the beam is diverging and d) at a broad maximum at the centre of an F lens.

In regions where the \( \beta \) function is not at a maximum or minimum the ellipse is tilted representing a beam that is either diverging or converging. In Fig. 4 we see the ellipse portrayed in terms of the Twiss parameters. The equation of the ellipse (often called the Courant and Snyder invariant) is:

\[
\gamma(s)y^2 + 2\alpha(s)y'y + \beta(s)y'^2 = \epsilon.
\]

and the invariance of this quantity as we move to different points in the ring is a consequence of Liouville's theorem.

All of this is true for a beam circulating at constant energy, but we may go further and apply Liouville's theorem to a proton beam during acceleration. To do so we must express emittance in the canonical phase space \((q, p)\), and we must first look carefully at the coordinates: displacement, \( y \), and divergence, \( y' \), we have been using so far. The canonical coordinates of relativistic mechanics are:

\[
q = y \\
p = \frac{my'}{\sqrt{1 - v^2/c^2}} = mc\gamma\beta_y
\]
where we are obliged to redefine the meaning of $\beta$ and $\gamma$:

$$\beta_y = \frac{\dot{y}}{c} ,$$

$$\gamma = \left[ 1 - \left( \dot{x}^2 + \dot{z}^2 + \dot{s}^2 \right) / c^2 \right]^{-1/2} ,$$

$m =$ rest mass,

$c =$ velocity of light.

The divergence is therefore:

$$y' = \frac{dy}{ds} = \frac{\dot{y}}{s} = \frac{\beta_y}{\beta} ,$$

where

$$\beta = \frac{v}{c} .$$

and the canonically conjugate $p$ and $q$ are therefore related to $y$ and $y'$:

$$p = mc \beta \gamma \cdot y'$$

$$q = y$$

In its general form Liouville's theorem states:

$$\int p \ dq = \text{constant} = mc \beta \gamma \int y' \ dy = mc \beta \gamma \cdot \pi \varepsilon = p_0 c \cdot \pi \varepsilon$$
where:

\( p_0 \) is the momentum of the particle,
\( \varepsilon \), the emittance, which is invariant at any given energy.

Accelerator physicists often quote an invariant emittance:

\[
\varepsilon^* = \beta \gamma \varepsilon \left[ \pi \cdot \text{mm.mrad} \right].
\]

As acceleration proceeds the normalized emittance is conserved; \( \beta \gamma \) is just proportional to the momentum or, above a few GeV, to the energy of the proton. Thus, we expect the emittance \( \varepsilon \) to shrink "adiabatically" during acceleration as \( 1/p_0 \) and the beam dimensions to shrink as \( 1/p_0^{1/2} \) (Fig. 5).

As a rule, proton accelerators need their full aperture at injection and it is then that their design is most critical. It is for this reason too, that multistage accelerators such as the Linac-PSB-PS complex are used, since by inserting the PSB the energy of the Linac beam is increased, thus allowing a beam of larger normalized emittance \( \varepsilon^* \) and containing more protons, to be injected into the PS.

2.1 Chains of accelerators

In a chain of proton accelerators such as those at FNAL or CERN the invariant emittance, fixed by space charge at the beginning of the linac may be conserved to several hundred GeV.

Of course, in practice there may be mismatches between machines or non-linear fields which dilate the emittance. There are even techniques, stochastic and electron cooling, which condense the emittance and appear to defeat Liouville's theorem. In fact they merely rearrange and centralise the particles within the emittance contour rather than affecting the mean density.
Another, more mathematical, description of Liouville's theorem is to consider the transformation of a small element of phase space bounded by eight particles, including a ninth and excluding a tenth particle. Suppose the two diagrams of Fig. 6 are related by the transformation, $M$. Then, in this picture area is related by the Jacobian determinant:

$$J = |M| = \begin{vmatrix} \frac{\partial y}{\partial y'} & \frac{\partial y}{\partial y''} \\ \frac{\partial y'}{\partial y''} & \frac{\partial y'}{\partial y''} \end{vmatrix} = 1.$$ 

Since the determinant is 1, area is conserved and Liouville follows. In other words Liouville's theorem implies that the determinant of the transport matrix is unity.

2.2 Exceptions to Liouville's theorem

By now, anyone who comes from a laboratory that has an electron synchrotron will be beginning to experience a rising tide of protest. In fact, the invariance of normalized emittance and the shrinking of physical emittance with energy is quite the opposite of what happens in an electron machine.

As we shall read in later contributions on synchrotron radiation [5], electrons, being highly relativistic, emit quanta of radiation as they are accelerated. The size of the average quantum increases with $E^3$ and the rate of emission as $E$. This quantized emission causes particles to jump around in momentum, changes which couple into transverse phase space. At the same time there is a steady tendency for particles near the edge of the emittance to lose transverse energy and fall back towards the centre. In an electron machine the emittance, determined by the equilibrium between these two effects, grows with $E^2$. Liouville's theorem only applies to particles guided by external fields and does not apply to electron machines where particles emit some of their own energy.

2.3 Definitions of emittance

Within the boundary of transverse phase space, the emittance of the beam, particles are usually distributed in a population which appears Gaussian when projected on a vertical or horizontal plane. The emittance boundary is traditionally chosen to include 95% of a Gaussian beam or $\approx 2\sigma_y$ in a proton machine. In an electron machine a $2\sigma$ boundary would be too close to the beam. An aperture stop placed at this distance would rather rapidly absorb
most of the beam as quantum emission stirs it up in phase space. The physical boundary allowed, depends on the lifetime required but is in the region $6\,\sigma$ to $10\,\sigma$. What is normally quoted as an emittance for an electron beam corresponds to $\sigma$ of the Gaussian projection. There is consequently a factor 4 between emittance defined by electron and proton people. Beware! Just to complicate matters proton people have of late taken to using the $\sigma^2/\beta$ definition of emittance used by electron people.

2.4 Acceptance

In contrast, the acceptance, $A$, the size of the hole which is the vacuum chamber transformed into phase space, is unambiguous. Remembering from the lectures on focusing that particles follow an elliptical trajectory with axes

$$\Delta y = \pm \sqrt{\epsilon \beta}, \quad \Delta y' = \pm \sqrt{\epsilon / \beta}$$

and area $\pi \epsilon$ we can write

$$\epsilon_{\text{protons}} = \frac{(2\sigma)^2}{\beta}$$

$$\epsilon_{\text{electrons}} = \frac{\sigma^2}{\beta}$$

$$A = \frac{r^2}{\beta}$$

where $r$ is the semi-axis of the chamber.

2.5 Measurement of emittance

It is my intention to give you some idea of the means used to measure the various optical quantities as they are mentioned. The methods of observing particle beams are much more limited than theory might suggest and it is as well to bear this in mind and aim for simplicity in arriving at a design rather than to rely upon a complex procedure.

The most reliable and straightforward way to measure proton beam size is to drive a scraper into the beam or move the beam across the scraper and integrate the beam loss curve. Electron machines observe beam size by refocusing the synchrotron light emitted, though the beam dimensions can be so small that special image scanning techniques are needed to resolve the image. A clever, non-destructive device used for proton machines is the Ionisation Beam Scanner [6]. Figure 7 shows the principle in which a zero electrical potential scans across the beam allowing electrons to be collected from its surface as the beam ionises the residual gas. A typical signal is given in Fig. 8.

This IBS has the advantage that it can give a "mountain range" display of how the beam shrinks during acceleration (Fig. 9). Its disadvantage is that the space charge fields of an intense beam can distort the width of the peak which represents the beam. For this reason it is rarely used nowadays except, as in this talk, for pedagogical reasons.

Another diagnostic tool is a wire which is scanned across the beam very rapidly. Secondary particles generated are counted with a scintillation telescope. Here one must be careful not to dilate the beam by scattering (or burn the wire).
Fig. 7  Diagram showing IBS principle of beam scanning
3. A SIMPLIFIED TREATMENT OF BETATRON MOTION

The earlier lectures on focusing treat betatron motion with all the rigour necessary to design a machine. However some readers who are new to the field might find the following sections too confusing if we carry through all the terms from the rigorous theory. We shall therefore use two models of the motion which are approximate but graphic. We shall confine our examination of the motion to maxima or minima in the beta function where the ellipse is upright as in Fig. 10.
In Fig. 10b we see the ellipse, plotted at a point where the amplitude function $\beta$ is large (as it is near F quadrupoles). The ellipse will be very wide and not very high in divergence angle and it is in such positions that a small angular kick has the greatest effect on the beam. Imagine, for example, how little angular displacement is needed to move the ellipse by its own height and increase the emittance by a factor 2. Thus, in a machine with a lattice like Fig. 1 most of the damage done to the beam by bad fields which cause angular kicks, happens near the F quadrupoles.

![Fig. 10 Phase-space diagram at (a) a $\beta$ minimum and (b) a $\beta$ maximum](image)

### 3.1 Circle approximation

So predominant is the effect of perturbations near $\hat{\beta}$ positions that you can often do quite good "back of the envelope" calculations by closing your eyes to what happens to the protons in between F quadrupoles. At F quadrupoles the ellipse always looks the same, i.e. upright, with semi-axes in displacement and divergence

\[ \sqrt{\beta \epsilon}, \sqrt{\epsilon / \beta} \]

This can be reduced to a circle radius $\sqrt{\beta \epsilon}$ by using the new coordinates

\[ y = y \]
\[ p = \beta y' \]

If the machine has 108 periods and a $Q$ of 27.6, the proton advances in phase by $2\pi Q/108$ from one period to the next; this is just the angle subtended at the centre of the circle multiplied by $Q$. After one turn of the machine, it has made 27 revolutions of the circle plus an angle of $2\pi$ multiplied by the fractional part of $Q$, see Fig. 11.

This renormalization of the phase space can be done in a more rigorous way by choosing new variables $(\eta, \psi)$ which transform the distortion of the phase and amplitude so that the motion becomes that of a harmonic oscillator. We must, of course, transform back again to see physical displacements, but the mathematics becomes more transparent. This simplification is discussed in the next section.
3.2 The ($\eta, \psi$) description of AG focusing

A more rigorous renormalization of phase space which does not imply any approximation but which simplifies the problem is the ($\eta, \psi$) transformation to convert Hill's equation into that of a harmonic oscillator:

$$\frac{d^2 \eta}{d\psi^2} + Q^2 \eta = g(\psi)$$

where $g(\psi)$ is the azimuthal pattern of some perturbation of the guide field related to $F(s) = \frac{\Delta B(s)}{B\rho}$.

In the ideal case $g(\psi)$ is everywhere zero.

I will not bother you with how this transformation is found, but just state it. The new coordinates are related to the old:

$$\eta = \beta^{-1/2} \psi$$

$$\psi = \int \frac{ds}{Q\beta}, \quad g(\psi) = Q^2 \beta^{3/2} F(s),$$

where $\psi$ advances by $2\pi$ every revolution. It coincides with $\theta$ at each $\beta$ or $\tilde{\beta}$ location and does not depart very much from $\theta$ in between.

4. THE $Q$ VALUES

A full understanding of transverse dynamics is rather difficult until one has an almost tactile appreciation of the nature of the oscillations. The simple models of the last section are intended to help but even more insight may be had from contemplating the methods of measuring the $Q$ values. A number of methods are possible and each reveals a different aspect of the motion.
4.1 Measurement of $Q$ by kicking

The method is to fire a kicker magnet with a pulse lasting less than one turn and observe the way in which the centre of charge of the beam oscillates as it passes a pick-up on sequential turns (see Fig. 12).

![Fig. 12 Ideal $Q$-measurement signal following a kick which excites coherent betatron motion. $Q =$ integer $\pm 1/6$.
In order to understand Fig. 12 it is convenient to imagine a beam consisting of one short longitudinal bunch. The line density current passing a detector is then a Fourier series

$$\rho(t) = \sum_n a_n \sin 2 \pi nf_0 t .$$

This beam position detector sees the betatron oscillations following the kick as:

$$y(t) = y_0 \cos 2 \pi f_0 Qt$$

but modulated by $\rho(t)$.

An oscilloscope connected to the pick-up will give a display of the modulated signal, the product of $\rho$ and $y$

$$\rho(t)y(t) = \frac{1}{2} \sum_n a_n y_0 \left[ \sin 2 \pi (n + Q)f_0 t + \sin 2 \pi (n - Q)f_0 t \right] .$$

The envelope of the oscilloscope signal will be the slowest of these terms in which $(n - Q)$ is the fractional part of $Q$. The other terms in the series reconstruct the spikes in the signal occurring once per turn. Note that I have made use of the elementary relation:

$$\sin a \cos b = \frac{1}{2} \left[ \sin (a + b) + \sin (a - b) \right] .$$
This will be often used when we come to look at resonances.

4.2 R.F. knockout and $Q$ measurement

A very simple form of resonance can be induced by applying a deflecting field with the frequency of a betatron sideband as the frequencies $(n \pm Q)f_0$ are called. We can easily invert the above treatment to show that if you apply a signal of the form

$$\sin[2\pi(n - Q)f_0 t]$$

then the particle passing the electrodes of the deflector sees the deflector every turn and experiences a kick

$$\sin[2\pi nf_0 t] \sin[2\pi(n - Q)nf_0 t].$$

Using again the elementary relation of Section 4.1 we see that the $(a - b)$ component is in resonance with the betatron motion

$$\sin 2\pi Qf_0 t.$$

Once in resonance, the particle is deflected on each turn by a kick which increases its amplitude of transverse oscillations in phase with the excitation and blows up the beam. Note that this like the previous method gives a value of the fractional part of $Q$ with respect to the nearest integer but gives no information about which integer this is.

4.3 Measurement by analysing the frequencies emitted by the beam

It is also possible nowadays to detect these betatron frequencies among the statistical noise signal detected by a transverse pick-up which can be just a pair of plates and displayed with an integrating spectrum analyser which is really a scanning radio receiver connected to an oscilloscope. Peaks appear as sidebands to the revolution frequency in the display of response versus frequency (Fig. 13). Their separation is $2\Delta Qf_0$ where $\Delta Q$ is the fractional part of $Q$.

![Transverse pickup seen on a spectrum analyser](image)

One may wonder why the particles, evenly spread around the ellipse in phase space, can generate such a signal. The answer is that they are finite in number and the pick-up samples
but a small fraction of them. In a sample there are always significant statistical fluctuations of
the centre of charge, or mean displacement, which the spectrum analyser picks up.

5. CLOSED-ORBIT DISTORTION

As an illustration of the power of \((\eta, \psi)\) coordinates we look at closed-orbit distortions which are produced by field errors. Even the best synchrotron magnets cannot be made absolutely identical. Each magnet differs from the mean by some small error in integrated strength:

\[
\delta(B \ell) = \int B \, d\ell - \left(\int B \, d\ell\right)_\text{ideal}.
\]

These and other machine imperfections, such as survey errors which can be expressed as equivalent to field errors, are randomly spread around the ring.

We can use the \((\eta, \psi)\) coordinates to find out how this perturbs a proton which would otherwise have had zero betatron amplitude. Such a proton no longer goes straight down the centre of the vacuum chamber but follows a perturbed closed orbit about which the normal betatron motion of the other protons can be superimposed.

One of the most important considerations in designing a machine is to keep this closed orbit distortion to a minimum because it eats up available machine aperture. Also, once we have succeeded in getting a few turns round the machine, we want to reduce this distortion with correcting dipole magnets. As a first step let us consider the effect on the orbit of such a correcting dipole located at a position where \(\beta = \beta_K\) and observed at another position.

A short dipole (we shall assume it is a delta function in \(s\)) makes a constant angular kick in divergence

\[
\delta\psi' = \delta(B \ell) / (B \rho),
\]

which perturbs the orbit trajectory which elsewhere obeys

\[
\frac{d^2 \eta}{d\psi^2} + Q^2 \eta = 0
\]

\[
\eta = \eta_0 \cos (Q \psi + \lambda).
\]

We choose \(\psi = 0\) origin to be diametrically opposite the kick. Then by symmetry \(\lambda = 0\) and the "orbit" is that shown in Fig. 14.
Fig. 14  Tracing the closed orbit for one turn in (η,ψ) space with a single kick at ψ = π.  (The $Q$ value is about 5.6.)

Since, by definition, the trajectory is closed, continuity demands that the kick $\delta y'$ matches the change in slope at $\psi = \pi$, the location of the dipole.

Differentiating the orbit equation

$$\frac{d\eta}{d\psi} = -\eta_0 Q \sin Q \psi = -\eta_0 Q \sin Q \pi,$$

at $\psi = \pi$.

To relate this to the real kick we use

$$\frac{dy}{ds} = \beta_K \frac{d\eta}{ds},$$

$$\frac{d\psi}{ds} = \frac{1}{Q \beta_K},$$

therefore

$$\frac{\delta y'}{2} = \frac{\delta (B\ell)}{2 B\rho} = \frac{dy}{ds} = \beta_K \frac{d\eta}{d\psi} \frac{d\psi}{ds} = -\frac{\eta_0}{\beta_K} \sin \pi Q,$$

$$\eta_0 = \frac{\sqrt{\beta_K}}{2 \sin \pi Q} \delta y'.$$

Returning to physical coordinates we can write the orbit's equation in the range $-\pi < \psi(s) < \pi$:

$$y = \beta(s) \eta_0 \cos Q \psi(s) = \left[ \frac{\sqrt{\beta(s)} \beta_K \delta(B\ell)}{2 \sin \pi Q} \frac{B\rho}{|B\rho|} \right] \cos Q \psi(s).$$

The expression in square brackets is the maximum amplitude of the perturbation at $\beta(s)$.

The above expression is rigorous but as an example of the use of the circle approximation consider the special case where the kink and observation are at the same value of beta. We see quite clearly from Fig. 15 how the equation for the amplitude of the distortion appear.
Fig. 15 Tracing a closed orbit for one turn in the circle diagram with a single kick. The path is ABCD.

In estimating the effect of a random distribution of dipole errors we must take the r.m.s. average, weighted according to the $\beta_k$ values over all of the kicks $\delta y_i$ from the N magnets in the ring. The expectation value of the amplitude is:

$$\langle y(s) \rangle = \frac{\sqrt{\beta(s)}}{2\sqrt{2} \sin \pi Q} \sqrt{\sum_i \beta_i \delta y_i^2}$$

$$\approx \frac{\sqrt{\beta(s)\beta}}{2\sqrt{2} \sin \pi Q} \sqrt{\frac{\langle \Delta B \rangle_{rms}}{B \rho}}.$$

The factor $\sqrt{2}$ comes from averaging over all the phases of distortion produced.

The principal imperfections in a synchrotron causing orbit distortion are shown in Table 1 [7]. The first line in the table represents the random variations in the position of quadrupole magnets with respect to their ideal location. A small displacement of a quadrupole gives an effective dipole perturbation, $k_i \Delta y$. The tilt of bending magnets causes a small resultant dipole in the horizontal direction which deflects vertically. Obviously random errors in magnet gap, length or in the coercivity of the steel yoke which determines remanent field contribute to the third line. Both remanent and stray fields in straight sections tend to be constant and their effect scales as $l/B$ as the machine pulses. Their effect should therefore be evaluated where it is worst, i.e. at injection. In a modern superconducting machine the persistent current fields play the role of remanent effects.

**Table 1**

<table>
<thead>
<tr>
<th>Type of element</th>
<th>Source of kick</th>
<th>r.m.s. value</th>
<th>$\langle \Delta B / (B \rho) \rangle_{ rms}$</th>
<th>plane</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gradient magnet</td>
<td>Displacement</td>
<td>$&lt;\Delta y&gt;$</td>
<td>$k_i \delta y$</td>
<td>$x,z$</td>
</tr>
<tr>
<td>Bending magnet</td>
<td>Tilt</td>
<td>$&lt;\Delta&gt;$</td>
<td>$\theta_i \langle \Delta \rangle$</td>
<td>$z$</td>
</tr>
</tbody>
</table>

16
In designing a machine it used to be conventional wisdom to make sure that the vacuum chamber will accommodate twice this expectation value. The probability of no particles making the first turn is thus reduced to a mere 2%. More modern designs rely on closed-orbit steering to thread the first turn and thereafter assume that orbit correction to a millimeter or so will be feasible.

5.1 The Fourier harmonics of the error distribution

One of the advantages of reducing the problem to that of a harmonic oscillator in \((\eta, \psi)\) coordinates is that perturbations can be treated as the driving term of the oscillator, broken down into their Fourier components, and the whole problem solved like the forced oscillations of a pendulum. The driving term is put on the right hand side of Hill's equation:

\[
\frac{d^2 \eta}{d\psi^2} + Q^2 \eta = Q^2 \sum_{n=1}^{\infty} f_k e^{i k \psi} = Q^2 \beta^{3/2} F(s) ,
\]

where \(F(s)\) is the azimuthal pattern of the perturbation \(\Delta B/(B \rho)\); and \(Q^2 \beta^{3/2}\) comes from the transformation from physical coordinates to \((\eta, \psi)\).

The Fourier amplitudes are defined:

\[
f(\psi) = \beta^{3/2} F(s) = \sum_k f_k e^{i k \psi} ,
\]

where

\[
f_k = \frac{1}{2 \pi} \int_0^{2\pi} f(\psi)e^{-i k \psi} d\psi = \frac{1}{2 \pi Q^2} \int f^{1/2} F(s)e^{-i k \psi} ds .
\]

We can then solve Hill's equation as

\[
\eta = \sum_{n=1}^{\infty} \frac{Q^2 f_k}{Q^2 - k^2} e^{i k \psi} \quad \text{(or its real part)} .
\]

But be careful. Before doing the Fourier analysis, \(\Delta B\) must be multiplied by \(\beta^{1/2}\) if the physical variable \(s\) is chosen as an independent variable, or \(\beta^{3/2}\) if \(\psi\), the transformed phase, is used.

Looking carefully at the above expression, we see that this differs from the general solutions

\[
\eta = \eta_0 e^{\pm i Q \psi}
\]
which describe betatron motion about the equilibrium orbit, because the wave number is an integer $k$. In fact it is a closed orbit, a particular solution of Hill’s differential equation, to which we must add the general solutions which describe betatron oscillations about this orbit.

The function $Q^2/(Q^2 - k^2)$ is sometimes called the magnification factor for a particular Fourier component of $\Delta B$. It rises steeply when the wave number $k$ is close to $Q$, and the effect of the two Fourier components in the random error pattern with $k$ values adjacent to $Q$ accounts for about 60% of the total distortion due to all random errors. Figure 16 shows a closed orbit pattern from electrostatic pick-ups in the FNAL ring, whose $Q$ is between 19 and 20. The pattern shows strong components with these wave numbers. If $Q$ is deliberately tuned to an integer $k$, the magnification factor is infinite and errors of that frequency make the proton walk out of the machine. This is in fact an integer resonance driven by dipole errors.

One instructive method of correcting orbit distortion is to apply a pattern of dipole correctors which excite an equal and opposite Fourier component of error at an integer close to $Q$.

![Fig. 16 FNAL main ring electrostatic pick-ups show closed orbit around the ring ($Q \cong 19.2$)](image)

### 5.2 Closed-orbit bumps

It is often important to deliberately distort a closed orbit bump at one part of the circumference without affecting the central orbit elsewhere. A typical example of this is to make the beam ride close to an extraction septum or within the narrow jaws of an extraction kicker magnet placed just outside the normal acceptance of the ring.

If one is lucky enough to find two positions for small dipole deflecting magnets spaced by $\pi$ in betatron phase and centred about the place where the bump is required, the solution is very simple. The distortion produced is:

$$y(s) = \delta \sqrt{\beta(s)\beta_k} \sin \left( \phi - \phi_0 \right)$$

where $\beta(s)$ is the beta function at $s$, $\beta_k$ is the beta function at the deflector and
This half-wave bump has a very simple configuration in normalised phase space (Fig. 17) We can immediately see that the central orbit (at the origin) is not disturbed elsewhere.

Note that the magnitude of the bump is not only proportional to the root of the local \( \beta(s) \) but is proportional to \( \sqrt{\beta_k} \). Since \( \beta \) is largest at F quadrupoles this is clearly where one should locate dipole bumpers.

Very often F-quadrupoles are not \( \pi \) apart in phase but \( 2\pi/3 \). this means the dipoles are slightly less effective (\( \sin \pi/3 = \sqrt{3}/2 \)) but also we must introduce a third dipole to form a triad. The third dipole is best located near the peak of the bump.

Figure 18 shows how the three bumps add up in normalised phase space. The case illustrated is the general one with dipoles of different strengths and spaced differently in phase.

In order to find an exact solution to the problem of a triad bump we use the matrix which transforms a point in phase space from one location to another (Eq. 1 of Ref. [2]).

\[
\begin{pmatrix}
y' \\
y''
\end{pmatrix}
= \begin{pmatrix}
\sqrt{\beta} / \sqrt{\beta_0} & (\cos \Delta \phi + \alpha_0 \sin \Delta \phi) \\
-1 / \sqrt{\beta_0} & (\alpha - \alpha_0) \cos \Delta \phi + (1 + \alpha \alpha_0) \sin \Delta \phi
\end{pmatrix}
\begin{pmatrix}
y \\
y'
\end{pmatrix}
= \begin{pmatrix}
\sqrt{\beta} / \sqrt{\beta_0} \beta_0 \sin \Delta \phi \\
-1 / \sqrt{\beta_0} (\sqrt{\beta} / \sqrt{\beta_0} \beta_0 \sin \Delta \phi)
\end{pmatrix}
\begin{pmatrix}
y_0 \\
y_0'
\end{pmatrix}
\]
It is the element which links $y$ to $y_0$ which describes the trajectory. Following the kick $\delta_i$,

$$y_2 = \delta_1 \sqrt{\beta_2 / \beta_1} \sin (\phi_2 - \phi_1).$$

The same argument can be used to describe the trajectory working back from $\delta_i$,

$$y_2 = -\delta_3 \sqrt{\beta_2 / \beta_3} \sin (\phi_2 - \phi_3).$$

The kick $\delta_2$ must be the change in the derivative

$$\delta_2 = \delta_1 \sqrt{\beta_1 / \beta_2} \cos (\phi_2 - \phi_1) + \delta_3 \sqrt{\beta_3 / \beta_2} \cos (\phi_3 - \phi_2).$$

Here the $\sqrt{\beta_2}$ in the denominator comes from multiplying by $d\phi/ds = 1/\beta_2$ at the location of the cusp.

We can rewrite these relations

$$\sqrt{\beta_1} \delta_1 \sin \phi_{1,2} = \sqrt{\beta_3} \delta_3 \sin \phi_{2,3}$$

$$\sqrt{\beta_2} \delta_2 = \sqrt{\beta_1} \delta_1 \cos \phi_{1,2} + \sqrt{\beta_3} \delta_3 \cos \phi_{2,3}.$$

These relations can be seen to be true for the triangle of sides, $\delta / \sqrt{\beta}$ and angles, $\phi$ which can be solved by the well-known symmetric relation

$$\frac{\delta_1 \sqrt{\beta_1}}{\sin \phi_{2,3}} = \frac{\delta_2 \sqrt{\beta_2}}{\sin \phi_{3,1}} = \frac{\delta_3 \sqrt{\beta_3}}{\sin \phi_{1,2}}.$$

### 5.3 The measurement and correction of closed orbits

Electrostatic plates with diagonal slots are commonly used to measure the transverse position of a bunched beam. We have seen that the predominant harmonic in the uncorrected orbit is close to $Q$ and to establish its amplitude and phase one really needs four pick-ups per
wavelength. Given the present fashion for FODO lattices with about 90° per cell and the need to measure in both planes, the final solution is usually one pickup at each quadrupole. The ones at F-quadrupoles, where beta is large horizontally, are the most accurate for the horizontal plane while others at D-quadrupoles are best for the vertical correction.

Similar arguments lead us to have one horizontally deflecting dipole at each F-quadrupole where $\beta_x$ is large, and one vertically deflecting dipole at each D-quadrupole, where $\beta_z$ is large.

Clearly so many correcting dipoles are unnecessary if only the two principal harmonics are to be corrected. In Fig. 19, one can excite the l9th harmonic with a single dipole, though a pair in opposition are necessary if their bending effect is to cancel and a further pair are needed in quadrature for the phase making four per harmonic and per plane. However, studies show that this simple harmonic correction still leaves about 30% of the initial distortion.

Another method consists in applying a set of superposed beam bumps formed along the lines calculated above; each a triad which compensates the measured orbit position at its centre. Given the power of modern computers this kind of correction can be calculated and applied all round the ring in a few seconds.

![Fig. 19](attachment:image.png)  
**Fig. 19** Diagram showing the sign of correction dipoles necessary to excite or compensate even or odd Fourier components of distortion around the ring of a synchrotron

Some machines do not have dipole correctors which are sufficiently strong to correct an orbit at their top energy and quadrupole magnets must be displaced upwards or sideways a distance $\Delta y$ to apply an effective dipole:

$$\frac{\Delta (B\ell)}{B\rho} = k\Delta y .$$

Such displacements are tedious to apply to all quadrupoles in such large machines as the SPS or LEP and indeed the accumulated effect of errors in moving so many quadrupoles by a few tenths of a millimetre might even make the orbit worse.

An alternative [8] to moving all quadrupoles (or powering all dipoles) is to select those which are most effective in correcting the orbit. Stored in the computer is a large matrix $G$ with as many rows as pick-ups and as many columns as correctors. Each term describes the effect $Y_i$ of a corrector $\Delta_j$ at the ith pick-up

$$Y = G\Delta .$$
Once the orbit is measured the column matrix \( Y \) is entered and can be converted into a set of corrections \( \Delta \):

\[
\Delta = \{- (GG)^{-1} \tilde{G}\} Y
\]

But since we first want to select the most effective correctors, we examine the terms of \( \tilde{Y}G\Delta \) which is a measure of efficiency. The best corrector is retained while a search is made for another one to form a doublet. At the cost of considerable computing time one can find any number of best correctors. The method also has the advantage that it can be used if one or more of the pick-ups is out of action.

One last comment on orbit correction is that while originally invented to save on magnet aperture, orbit corrections now seem an essential procedure to reduce the effects of non-linearities in the dynamics of synchrotrons. One example is that orbit distortion in the sextupoles, which all machines have for correcting chromaticity, will generate a pattern of quadrupole gradient errors and drive half-integer stopbands.

6. GRADIENT ERRORS

Fig. 20 Matrix representation of a small quadrupole, \( m_0 \) subject to an error which is a component of the matrix for the whole ring, \( M \)

Quadrupoles as well as bending magnets may have errors. Understanding the effect of such gradient errors is a useful preparation for the study of non-linear errors. We represent a ring of magnets as a circle in Fig. 20 and the matrix for one turn starting at \( A \) as:

\[
M_0(s) = \begin{pmatrix}
\cos \varphi_0 + \alpha_0 \sin \varphi_0, & \beta_0 \sin \varphi_0 \\
-\gamma_0 \sin \varphi_0, & \cos \varphi_0 - \alpha_0 \sin \varphi_0
\end{pmatrix}.
\]

Now consider a small gradient error which afflicts a quadrupole in the lattice between \( B \) and \( A \). The unperturbed matrix for this quadrupole is \( m_0 \) and when perturbed the quadrupole matrix is

\[
m_0 = \begin{pmatrix} 1 & 0 \\ -k_0(s)ds & 1 \end{pmatrix},
\]

and perturbed, a matrix

\[
m = \begin{pmatrix} 1 & 0 \\ [-k_0(s) + \delta k(s)]ds & 1 \end{pmatrix}.
\]
The unperturbed transfer matrix for the whole machine:

\[
M_0(s) = \begin{pmatrix}
\cos \phi_0 + \alpha_0 \sin \phi_0 & \beta_0 \sin \phi_0 \\
-\gamma_0 \sin \phi_0 & \cos \phi_0 - \alpha_0 \sin \phi_0
\end{pmatrix}
\]

includes \( m_0 \).

To find the perturbed transfer matrix we make a turn, back-track through the small unperturbed quadrupole \( m_0^{-1} \), and then proceed through the perturbed quadrupole \( m \). Translated into matrix algebra,

\[
M(s) = mm_0^{-1}M_0.
\]

Now

\[
mm_0^{-1} = \begin{pmatrix} 1 & 0 \\ -\delta k(s) ds & 1 \end{pmatrix}.
\]

So

\[
M = \begin{pmatrix}
\cos \phi_0 + \alpha_0 \sin \phi_0, & \beta_0 \sin \phi_0 \\
-\delta k(s) ds \left( \cos \phi_0 + \alpha_0 \sin \phi_0 \right) - \gamma \sin \phi_0, & -\delta k(s) ds \beta_0 \sin \phi_0 + \cos \phi_0 - \alpha_0 \sin \phi_0
\end{pmatrix}.
\]

Now \( \frac{1}{2} (Tr M) = \cos \phi \). So the change in \( \cos \phi \) is

\[
\Delta (\cos \phi) = -\Delta \phi \sin \phi_0 = \frac{\sin \phi_0}{2} \beta_0(s) \delta k(s) ds
\]

\[
2\pi \Delta Q = \Delta \phi = \frac{\beta(s) \delta k(s) ds}{2}.
\]

Since betatron phase is not involved in this equation we are tempted to integrate around the ring to obtain

\[
\Delta Q = \frac{1}{4\pi} \int \beta(s) \delta k(s) \, ds.
\]

This result is often used in accelerator theory and is surprising in that we see that the change is independent of the phase of the perturbation. However, this equation is only approximately true since as we add each elemental focusing error it modifies \( \beta(s) \) as well as \( Q \) so that there is a higher-order term which should be included if one wants accurate numerical results [see Ref. [9] Eqs. (4.32) to (4.37)]. Nevertheless, used with discretion it is sufficiently accurate to explain the physical basis of the resonant phenomena we shall be discussing in later sections which can usually only be estimated to within a factor of 2 anyway.

The reason for our concern about the change in tune or phase advance which results from errors is that we must steer \( Q \) well away from certain fractional values which can cause the motion to resonate and result in loss of the beam. To understand how some \( Q \) values are dangerous let us return to the case of closed orbit distortion. Earlier we found the orbit distortion amplitude:
\[ \dot{y} = \frac{\sqrt{\beta_x}}{2 \sin \pi Q} \frac{\delta(B/\rho)}{B \rho} . \]

Clearly this will become infinite if \( Q \) is an integer value. What happens physically is that the beam receives a kick at the same phase on every turn and just spirals outwards. An error in gradient can have the same effect if the \( Q \) value is close to one of the lines:

\[
\begin{align*}
2Q_h &= p, & 2Q_v &= p \\
Q_h &= Q_v = p, & Q_h + Q_v &= p
\end{align*}
\]

where \( p \) is an integer.

At this stage in the description of transverse dynamics we can only hint at the explanation for this. Particles spiral outwards in phase space if the perturbation has the same effect on each turn. The perturbation from a dipole is independent of the transverse displacement and can only build up in this way if the particle returns to the same point in phase space on each turn (\( Q = p \)). A quadrupole error has field proportional to \( x \) and if a particle makes half turns in phase space it will see alternately positive and negative kicks in divergence but both will reinforce the growth. One may extend this argument to understand why sextupole errors which have a quadratic \( x \) dependence excite third-integer "resonances" as they are called near the lines:

\[
\begin{align*}
3Q_h &= p \\
2Q_h + Q_v &= p \\
Q_h + 2Q_v &= p \\
3Q_v &= p \\
2Q_h - Q_v &= p \\
Q_h - 2Q_v &= p
\end{align*}
\]

7. THE WORKING DIAGRAM

This is simply a diagram with \( Q_H \) and \( Q_V \) as its axes. The beam can be plotted on it as a point but because there is a certain \( Q \)-spread among protons of different momenta it is better to give the point a finite radius \( \Delta Q \) (Fig. 21).

We plot on the diagram a mesh of lines which mark danger zones for the protons. We have hinted that if \( Q \) in either the vertical or the horizontal plane is a simple vulgar fraction, then

\[ nQ = p , \]

where \( n \) and \( p \) are integer and \( n < 5 \), a resonance takes over and walks the proton out of the beam. In general this is true when

\[ \ell Q_H + m Q_V = p , \]

where \( |\ell| + |m| \) is the order of the resonance and \( p \) is the azimuthal frequency which drives it.
This equation just defines a set of lines in the $Q$ diagram for each order of resonance and for each value of the integer $p$. Figure 21 shows these lines for the SPS.

Somehow, by careful adjustment of the quadrupoles in the lattice and by keeping the $Q$-spread (chromaticity) small, we must coax the beam up to full energy without hitting the lines. To make things more difficult, each line has a finite width, proportional to the strength of the imperfection which drives it. In some cases we must compensate the imperfections with correction multipoles to reduce this width.

But before discussing resonances and their correction in another contribution to these proceedings [4], a word about chromaticity.

8. CHROMATICITY

This steering of $Q$ depends on careful regulation of quadrupole and dipole power supplies. In fact, much of the setting up time of a large accelerator is devoted to tune $Q$ to be constant as the fields and energy rise. Once beam has been accelerated the problem becomes one of reducing all effects which produce a spread in $Q$ among the particles in the beam. The limit to this is usually reached when beam intensity is high enough to cause space-charge focusing effects whose strength varies with the local beam density. Before reaching this limit one must correct the tune spread due to momentum: the chromaticity. This is exactly equivalent to the chromatic aberration in a lens. It is defined as a quantity $\Delta Q'$

$$\Delta Q = Q' \frac{\Delta p}{p}.$$  

It can be measured by changing the mean momentum of the beam by offsetting the r.f. frequency and measuring $Q$. Figure 22 shows such a measurement. Changing the r.f. frequency or momentum at a given field implies a change in radial position. As we have seen, an off-momentum particle will take up a new orbit following the dispersion function.
The chromaticity arises because the focusing strength of a quadrupole has \((B\rho)\) in the denominator and is therefore inversely proportional to momentum. The focusing strength of the lattice quadrupoles, 

\[
k = \frac{1}{(B\rho)} \frac{dB_z}{dx},
\]

varies inversely with \((B\rho)\), i.e., with momentum, \(p\). A small spread in momentum in the beam, \(\pm \Delta p/p\), causes a spread in focusing strength:

![Fig. 22 Measurement of variation of \(Q\) with mean radius made by changing the r.f. frequency](image)

\[
\frac{\Delta k}{k} = -\frac{\Delta p}{p}.
\]

Since the \(Q\)-value depends on \(k\), we can also write a formula for the \(Q\)-spread:

\[
\Delta Q = Q' \frac{\Delta p}{p},
\]

where the constant \(Q'\) is the chromaticity, analogous to chromatic aberration in an optical system.

An equation we derived earlier in the section on gradient errors

\[
\Delta Q = \frac{1}{4\pi} \int \beta(s) \delta k(s) \, ds.
\]

enables us to calculate \(Q'\) rather quickly:

\[
\Delta Q = \frac{1}{4\pi} \int \beta(s) \delta k(s) \, ds = \left[ \beta(s)k(s) \right] \Delta p \frac{p}{p}.
\]

The chromaticity \(Q'\) is just the quantity in square brackets. To be clear, this is called the natural chromaticity. For the SPS and the PS, indeed most AG machines, its value is about \(-1.3\) \(Q\) for H and V planes.
Imagine the situation at injection where $\Delta p/p$ can be $\pm 2 \times 10^{-3}$. In a large synchrotron with a $Q$ about 25 this can make the working point in the $Q$ diagram into a line of length $\Delta Q = 0.15$ which is too long to avoid the resonances. This must be corrected.

Just to make matters worse the chromaticity one has to correct may well be much greater than that due to the natural chromatic properties of quadrupoles. The remanent field at injection into a large ring may well be half a percent of the guide field and has the parabolic shape of a sextupole. In a superconducting ring the sextupole fields at injection stem from persistent currents and are very much larger still. Storage rings are usually designed with low-beta sections with zero dispersion for the interaction regions and the main low-beta quadrupoles being very strong make enormous contributions to the chromaticity. Since the dispersion is zero at the source of the error the compensation can only be made elsewhere in the lattice where the parameter $D$ is large.

One way to correct this is to introduce some focusing which gets stronger for the high momentum orbits near the outside of the vacuum chamber – a quadrupole whose gradient increases with radial position is needed. Such a magnet has 6 poles, i.e. a sextupole. In a place where there is dispersion it will introduce a focusing

$$\Delta k = \frac{B''(s) \Delta p}{(B \rho(s))} \frac{\beta}{p}.$$ 

We use an earlier expression for the effect of this $\Delta k$ on $Q$ and obtain

$$\Delta Q = \left[ \frac{1}{4 \pi} \int \frac{B''(s) \beta(s) D(s) ds}{(B \rho)} \right] \frac{dp}{p}.$$ 

To correct chromaticity we have merely to make the quantity in the square bracket balance the chromaticity.

However, there are two chromaticities, one affecting $Q_x$, the other $Q_z$ and we must therefore arrange for the sextupoles to cancel both. For this we use a trick which is common in many different contexts. Sextupoles near F-quadrupoles where $\beta_x$ is large affect mainly the horizontal $Q$, while those near D-quadrupoles where $\beta_z$ is large influences $Q_z$. The effects of two families like this are not completely orthogonal but by inverting a simple $2 \times 2$ matrix one can find two sextupole sets which do the job.

The correction of chromaticity is a subject on its own since there is a higher-order term, a parabolic variation of $Q$ with momentum which is not compensated in this way. Sextupole patterns which minimise this, yet do not themselves excite serious non-linear side effects, are not easy to find.

There are two ways of measuring chromaticity apart from the radial steering method shown in Fig. 22. The first of these is to observe the width of the betatron sidebands in the spectrum from a transverse pickup (Fig. 13). Secondly, we can measure the time it takes for a coherent betatron oscillation following a small kick to disappear as the $\Delta Q$ smears out the phase relation between protons of different momenta (Fig. 23). A ringing time of 200 turns signifies a $\Delta Q \approx 1/200$ and is about the best we can hope for using this rather crude method.
9. CONCLUSIONS

We have now covered sufficient of the theory of transverse beam dynamics to understand the basic processes of designing focusing structures for a circular machine, the mechanisms which are produced by errors and how they may be compensated, at least in the case of those due to linear fields. The reader will also have learned enough about the behaviour of off-momentum particles to follow the explanation of longitudinal dynamics which follows. It remains only to list a bibliography of works which the author has found to be useful in understanding this topic.

* * *

REFERENCES


BIBLIOGRAPHY

Livingood, J., 1961, Principles of Cyclic Accelerators (Van Nostrand).
Rosenblatt, J., 1968, Particle Acceleration (Methuen).
The 300 GeV Programme, 1972, CERN/1050 (CERN).
Theoretical Aspects of the Behaviour of Beams in Accelerators and Storage Rings, 1977, CERN 77-13 (CERN).
Wilson, E.J.N., 1977, Proton Synchrotron Accelerator Theory, CERN 77-07 (CERN).
Bryant, P. and Johnsen, K., The Principles of Circular Accelerators and Storage Rings (Cambridge).
LATTICE DESIGN IN HIGH-ENERGY PARTICLE ACCELERATORS

B. J. Holzer
DESY, Hamburg, Germany

Abstract
This lecture introduces storage-ring lattice design. Applying the formalism that has been established in transverse beam optics, the basic principles of the development of a magnet lattice are explained and the characteristics of the resulting magnet structure are discussed. The periodic assembly of a storage ring cell with its boundary conditions concerning stability and scaling of the beam optics parameters is addressed as well as special lattice structures: drifts, mini beta insertions, dispersion suppressors, etc. In addition to the exact calculations indispensable for a rigorous treatment of the matter, scaling rules are shown and simple rules of thumb are included that enable the lattice designer to do the first estimates and get the basic numbers ‘on the back of an envelope’.

1 INTRODUCTION
The study of nuclear and high-energy physics has always been a driving force for the development of high energetic particle beams; while the first experiments in that field were performed using ‘beams’ from natural radioactive particle sources like α- or β- emitters, it soon became clear that to get higher energetic particles, special machines — particle accelerators — had to be developed.

One of the very first and most important experiments in twentieth century physics — the Rutherford scattering experiment in 1911 — was performed using alpha particles from a natural source in the MeV range (Fig. 1, [1]), before particle accelerators had been invented.

![Fig. 1: Result of the Rutherford scattering experiment compared to the angle expected from the Thomson model](image-url)
Soon afterwards, machines to accelerate particle beams artificially, such as electrostatic machines like the Cockroft–Walton generator and the Van de Graaf accelerator as well as the first circular accelerators like the betatron and the cyclotron [2] came into operation.

The basis for a concept called ‘lattice design’, the topic of this lecture, was laid out in 1952 when Courant, Livingston, and Snyder developed the theory of strong focusing accelerators (or alternating-gradient machines) [3].

Lattice design in the context described here is the design and optimization of the principal elements — the lattice cells — of a (circular) accelerator, and it includes the dedicated variation of lattice elements (for example, position and strength of the magnets in the machine) to get well-defined and predictable parameters of the stored particle beam.

It is therefore closely related to the theory of linear beam optics described in Ted Wilson’s lecture in this school [4] and in the introductory CAS lecture by Rossbach and Schmueser [5].

Machine lattice design is an application of linear beam optics and has high practical relevance. Lattice examples of present-day storage rings and plots from optics calculations of real machines are shown throughout the paper.

In Fig. 2 the beam optics in a part of a lattice structure is shown. The plot shows the most important beam parameters in a typical high-energy storage ring. In the upper part, the beta function in the horizontal ($x$) and vertical ($y$) plane is plotted (or the square root proportional to the beam size), in the middle part the position of the lattice elements is shown, and the lower part shows the so-called dispersion function in $x$ and $y$. We will discuss these parameters in some detail in this paper.

![Lattice and beam optics](image)

**Fig. 2:** Lattice and beam optics in a part of a typical high-energy accelerator. The curves in the upper part refer to the square root of the beta function, the lower part shows the dispersion function.

### 1.1 Geometry of the ring

The first step in the layout of a storage ring is fixing the geometry or, the closely related task, determining the momentum of the particles to be stored in it.
For the bending force and focusing of the particle beam, magnetic fields are applied in a circular accelerator. In principle, electrostatic fields would be possible as well but at high momenta (i.e. if the particle velocity is close to the speed of light) magnetic fields are much more efficient. The force acting on the particles, the Lorentz force, is given by

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

or neglecting the $\vec{E}$ field,

$$\vec{F} = q (\vec{v} \times \vec{B}) .$$

In a constant transverse magnetic field $\vec{B}$, the particle will see a constant deflecting force and the trajectory will be a part of a circle. In other words, the condition for a circular orbit is that the Lorentz force be equal to the centrifugal force:

$$e* v * B = \frac{mv^2}{\rho} .$$

Dividing by the velocity $v$ we get a relation between the magnetic field and the momentum of the particle

$$e * B = \frac{mv}{\rho} = \frac{p}{\rho} .$$

The term $B^\star \rho$ is called beam rigidity,

$$B^\star \rho = \frac{p}{e}$$

and connects the magnetic dipole field needed for a circular orbit of radius $\rho$ to the particles’ momentum and charge. (Note that here we often refer to protons or electrons and the charge is just the elementary charge $e$).

With an ideal circular orbit, for each segment of the path we get the relation

$$\alpha = \frac{ds}{\rho} , \quad \alpha = \frac{B* ds}{B* \rho} ,$$

and integrating along the path $ds$ in all dipole magnets in the ring we require

$$\alpha = \int \frac{Bds}{B* \rho} = 2\pi \quad \rightarrow \quad \int Bds = 2\pi * \frac{p}{q} . \quad (1)$$

As in any ‘circular’ accelerator the angle swept in one turn for the design particle is $2\pi$. Equation (1) tells us that the integral of all bending magnets in the ring has to be $2\pi$ times the momentum of the beam. If the path length inside the dipole magnet does not differ much from the length of the magnet itself, the integral in Eq. (1) can be approximated by $\int Bdl$, where $dl$ refers to the magnet length.

In Fig. 3 the field of a typical bending magnet used in a storage ring is shown. On the vertical scale, the magnetic induction $B$ is shown in tesla, and is measured between the two pole faces of the dipole magnet. The plateau of constant field is easily seen inside the magnet as are the decreasing edge fields before and after the magnet.
For the lattice designer, the integrated $B$ field along the particles’ design orbit (roughly sketched in the figure) is the most important parameter, as it is the value that enters Eq. (1) and defines the field strength and how many of these magnets are needed for a full circle.

Figure 4 shows a photograph of a small storage ring [6] with only eight dipole magnets used to define the design orbit. The magnets are powered symmetrically and therefore each magnet corresponds to a bending angle of exactly $\alpha = 45^\circ$ of the beam. The field strength in this machine is in the order of $B = 1$ T.

In general, for a high-energy storage ring or synchrotron, a large number of bending magnets with very high magnetic fields are needed to determine the design orbit.

As an example, the HERA storage ring is presented in Fig. 5. It accelerates and stores proton beams of an energy of 920 GeV and collides them with $e^+$ or $e^-$ beams of about 27.5 GeV. In the HERA proton ring, 416 dipole magnets are used to guide the beam on an orbit of 6.3 km circumference. The length of each dipole magnet is $l = 8.8$ m.

At an energy of 920 GeV the particles are ultra-relativistic and we can put

$$E \approx p^* c$$
to calculate the momentum

\[ \int Bdl = N \pi B = 2\pi p/q. \]  

(2)

Using formula (2), for the magnetic field we get

\[ B \approx \frac{2\pi \cdot 920 \cdot 10^9 \text{eV}}{416 \cdot 3 \cdot 10^8 \text{m/s} \cdot 8.8 \cdot 10^4 \text{e}} \approx 5.15 \text{T}. \]

It is immediately clear that the machine has to be built with superconducting magnets if the required energy of 920 GeV is to be achieved.

Fig. 5: HERA storage ring: 416 superconducting magnets are needed to bend the protons on a circular path of 6.3 km length

1.2 Equation of motion and matrix formalism

If the geometry and specification of the arc is determined and the layout of the bending magnets is accomplished, the next step is the focusing properties of the machine. In general, we have to keep more then \(10^{12}\) particles in the machine, distributed in a certain number of bunches. And these particles have to be focused to keep their trajectories close to the design orbit.

As mentioned in the linear beam optics lecture, gradient fields are used to do the job. They generate a magnetic field that increases linearly as a function of the distance to the magnet centre;

\[ B_y = -g * x, \quad B_x = -g * y \]

where \(x\) and \(y\) refer to the horizontal and vertical plane and the parameter \(g\) is called the gradient of the magnetic field.

It is the custom to normalize the magnetic fields to the momentum of the particles. In the case of the dipole fields we get from Eq. (1)

\[ \alpha = \frac{\int Bdl}{B \cdot \rho} = \frac{L_{\text{eff}}}{\rho}, \]
where \( L_{\text{eff}} \) is the so-called effective length of the magnet; the term \( 1/\rho \) is the bending strength of the dipole. In the same way the field of the quadrupole lenses is normalized to \( B^*\rho \): the strength \( k \) is defined by

\[
k = \frac{g}{B^*\rho}
\]

and the focal length of the quadrupole is given by

\[
f = \frac{1}{k^*\ell}.
\]

Under the influence of the focusing properties of the quadrupole and dipole fields in the ring, the particle trajectories are described by a differential equation. In the lecture about linear beam optics this equation is derived in all its glory, so here we just state that it is given by the expression

\[
x'' + K^* x = 0
\]

\( x \) describes the horizontal coordinate of the particle with respect to the design orbit, the derivative in linear beam optics is taken with respect to the orbit coordinate \( s \), and the parameter \( K \) combines the focusing strength \( k \) of the quadrupole and the weak focusing term \( 1/\rho^2 \) of the dipole field. (Note: A negative value of \( k \) means a horizontal focusing magnet.)

\[
K = -k + 1/\rho^2.
\]

In most accelerators, the term \( 1/\rho^2 \) is missing in the vertical plane, the design orbit is in the horizontal plane and no vertical bending strength is present. So we get in the vertical plane

\[
K = k.
\]

Remember when designing a magnet lattice to simplify as much as possible in the beginning. Clearly the exact solution of particle motion has to be calculated in full detail and if the beam optics is optimized on a linear basis, higher-order multipole fields and their effect on the beam have to be taken into account. By completing the first steps we can make life easier and ignore terms small enough to be neglected.

In many cases the weak focusing term \( 1/\rho^2 \) can be neglected in favour of a rough estimate making the formula much shorter and symmetric in the horizontal and vertical plane. Referring to the example of the HERA proton ring, the basic parameters of the machine are as in Table 1.

<table>
<thead>
<tr>
<th>Table 1: Basic parameters of the HERA proton storage ring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circumference ( C_0 )</td>
</tr>
<tr>
<td>Bending radius ( \rho )</td>
</tr>
<tr>
<td>Quadrupole gradient ( G )</td>
</tr>
<tr>
<td>Particle momentum ( p )</td>
</tr>
<tr>
<td>Weak focusing term ( 1/\rho^2 )</td>
</tr>
<tr>
<td>Focusing strength ( k )</td>
</tr>
</tbody>
</table>

The weak focusing contribution \( 1/\rho^2 = 2.97 \times 10^6 /\text{m} \) is much smaller than the quadrupole strength \( k \), and can generally be neglected for initial estimates in the lattice of large accelerators.
1.3 Single-particle trajectories

The differential Eq. (3) describes the transverse motion of the particle with respect to the design orbit. In linear approximation it can be solved and the solutions for the horizontal and vertical planes are independent of each other.

If the focusing parameter $K$ is constant, if we refer the situation to a place inside a magnet where the field is constant along the orbit, the general solution for the position and angle of the trajectory can be derived as a function of the initial conditions $y_0$ and $y'_0$. In the case of a focusing lens we obtain

$$x(s) = x_0 \cos(\sqrt{K} \cdot s) + \frac{x'_0}{\sqrt{K}} \sin(\sqrt{K} \cdot s)$$

$$x'(s) = -x_0 \sqrt{K} \sin(\sqrt{K} \cdot s) + x'_0 \cos(\sqrt{K} \cdot s)$$

Or, written in a more convenient matrix form:

$$\begin{pmatrix} x \\ x' \end{pmatrix} = M \begin{pmatrix} x \\ x' \end{pmatrix}_0$$

The matrix $M$ depends on the properties of the magnet, and for a number of typical lattice elements we get

focusing quadrupole

$$M_{QF} = \begin{pmatrix} \cos(\sqrt{K} \cdot \ell) & \frac{1}{\sqrt{K}} \sin(\sqrt{K} \cdot \ell) \\ -\sqrt{K} \sin(\sqrt{K} \cdot \ell) & \cos(\sqrt{K} \cdot \ell) \end{pmatrix} \quad (4a)$$

defocusing quadrupole

$$M_{QD} = \begin{pmatrix} \cosh(\sqrt{K} \cdot \ell) & \frac{1}{\sqrt{K}} \sinh(\sqrt{K} \cdot \ell) \\ \sqrt{K} \sinh(\sqrt{K} \cdot \ell) & \cosh(\sqrt{K} \cdot \ell) \end{pmatrix} \quad (4b)$$

drift space

$$M_{drift} = \begin{pmatrix} 1 & \ell \\ 0 & 1 \end{pmatrix} \quad (4c)$$

1.4 The Twiss parameters $\alpha, \beta, \gamma$

In the case of periodic conditions in the accelerator there is another more convenient way to describe the particle trajectories. The above-mentioned formalism is valid only within a single element.

Note that in a circular accelerator the focusing elements are periodic in the orbit coordinate $s$ after one revolution. In addition, storage ring lattices have, in most cases, an inner periodicity. They often are built — at least partly — of sequences where identical magnetic cells, the lattice cells, are repeated several times in the ring and lead to periodically repeated focusing properties.

In this case the transfer matrix from the beginning of the structure to the end is expressed as a function of the periodic parameters $\alpha, \beta, \gamma, \varphi$:
The parameters \( \alpha \) and \( \gamma \) are related to the \( \beta \)-function by the equations

\[
\alpha(s) = -\frac{1}{2} \beta'(s) \quad \text{and} \quad \gamma(s) = \frac{1+\alpha^2(s)}{\beta(s)}.
\]

The matrix is clearly a function of the position \( s \), as the parameters \( \alpha, \beta, \gamma \) depend on \( s \). The variable \( \phi \) is called phase advance of the trajectory and is given by

\[
\phi = \int_s^{s+L} \frac{dt}{\beta(t)}.
\]

In such a periodic lattice, for stability of the equation of motion the relation

\[ |\text{trace}(M)| < 2 \]

setting boundary conditions for the focusing properties of the lattice, has to be valid.

Given that correlation, the solution of the trajectory of a particle can be expressed as a function of these new parameters:

\[
y(s) = \sqrt{\epsilon} \ast \sqrt{\beta(s)} \ast \cos(\varphi(s) - \delta) \\
y'(s) = \frac{-\sqrt{\epsilon}}{\sqrt{\beta(s)}} \ast \{\sin(\varphi(s) - \delta) + \alpha(s) \cos(\varphi(s) - \delta)\}.
\]

The position and angle of the transverse oscillation of a particle at a point \( s \) is given by the value of the \( \beta \)-function at that location and \( \epsilon \) and \( \delta \) are constants of the particular trajectory.

Finally, in terms of linear beam optics, the Twiss parameters at a position \( s \) in the lattice are defined by the focusing properties of the complete storage ring. They are transformed through the lattice from one point to another by the matrix elements of the corresponding magnets. Without proof, if matrix \( M \) is given by

\[
M(s_1,s_2) = \begin{pmatrix} C & S \\ C' & S' \end{pmatrix}
\]

then the transformation rule from point \( s_1 \) to \( s_2 \) in the lattice is given by

\[
\begin{pmatrix} \beta \\ \alpha \\ \gamma \end{pmatrix}_{s_2} = \begin{pmatrix} C^2 & -2SC & S^2 \\ -CC' & SC' + S'C & -SS' \\ C'^2 & -2S'C' & S'^2 \end{pmatrix} \begin{pmatrix} \beta \\ \alpha \\ \gamma \end{pmatrix}_{s_1}.
\]

The terms \( C, S \) etc. correspond to the focusing properties of the matrix. For one single element, see the expressions given in the Eq. (4) of the previous section.
2 LATTICE DESIGN

A simple drift space in a lattice is the easiest case to investigate.

In Eq. (4) the matrix for a drift space is given by

\[
M = \begin{pmatrix}
C & S \\
C' & S'
\end{pmatrix} = \begin{pmatrix}
1 & \ell \\
0 & 1
\end{pmatrix}.
\]

(7)

Starting with position \(x_0\) and angle \(x'_0\) the trajectory after the drift therefore will be

\[
\begin{pmatrix}
x \\
x'
\end{pmatrix}_\ell = \begin{pmatrix}
1 & \ell \\
0 & 1
\end{pmatrix}^{*} \begin{pmatrix}
x \\
x'
\end{pmatrix}_0
\]

or written explicitly:

\[
x(\ell) = x_0 + \ell x'_0 \\
x'(\ell) = x'_0.
\]

If the drift is located in a circular accelerator or within a periodic part of a lattice, the Twiss parameters are well-defined at the start of the drift, and they will be transformed according to Eq. (6) from their initial values \(\alpha_0, \beta_0, \gamma_0\) via

\[
\begin{pmatrix}
\beta \\
\alpha \\
\gamma
\end{pmatrix}_\ell = \begin{pmatrix}
1 & -2\ell & \ell^2 \\
0 & 1 & -\ell \\
0 & 0 & 1
\end{pmatrix}^{*} \begin{pmatrix}
\beta \\
\alpha \\
\gamma
\end{pmatrix}_0.
\]

The beta function in the drift develops as

\[
\beta(\ell) = \beta_0 - 2\ell \alpha_0 + \ell^2 \gamma_0.
\]

Calculating the trace of the matrix (7) we see that the condition for stability is unfulfilled.

\[
|\text{trace}(M_\ell)| = 1 + 1 = 2.
\]

So we know that a circular accelerator built exclusively out of drift spaces is not a stable machine. (This is a pity, since it would have been a cheap machine.)

N.B., in any storage ring there will be a large number of drift spaces between the focusing elements. The stability criterion requires the magnetic elements and the drift spaces in between them to be arranged so that the resulting lattice cell describes a stable solution in both the horizontal and vertical plane.

Figure 6 shows the lattice and the beam optics of a typical high-energy storage ring, the HERA proton ring already mentioned. In the upper part of Fig. 6, the square root of the \(\beta\) function is shown in both transverse planes. The broad band in the middle is the lattice. As the ring has a circumference of about 6.3 km the single lattice elements cannot be clearly distinguished in the figure. In the lower part, the dispersion function is shown.
Initially, two sections of quite different characteristics can be identified in the accelerator and, in fact, most if not all, high-energy colliders are designed with:

- A section where the beta function shows a regular pattern: These are the arcs where the main bending magnets of the ring are located. They define the geometry of the ring and the maximum magnetic fields of the bends limit the energy (or momentum) of the particle beam. In addition, the main focusing elements of the ring are located in the arcs where quadrupole lenses for tune control and sextupole magnets for the compensation of the chromaticity of the optics are usually found.

- Regular arcs connected by straight sections in the ring: These are long lattice parts where the optics is modified to establish conditions needed for particle injection, to reduce the dispersion function, or where the beam dimensions are reduced to increase the particle collision rate in the case of a collider ring. Beyond that, all kind of devices have to be installed in these long sections like RF cavities, beam diagnostic tools, and even high-energy particle detectors (...if they cannot be avoided in the machine).

Concerning the arcs’ structure it is advantageous to configure them on the basis of small elements — called cells — repeated many times in the ring. One of the most widespread lattice cells used for this purpose is the FODO cell.

### 2.1 The FODO cell

A magnet structure consisting of focusing and defocusing quadrupole lenses in alternating order with basically nothing in between, is called a FODO lattice, and the elementary cell a FODO cell. ‘Basically nothing’ in that context means any element with negligible effect on the focusing properties, as, for example, drift spaces, RF-structures or, under certain circumstances, even bending magnets. A FODO cell is shown in Fig 7.
To calculate the Twiss parameters $\alpha$, $\beta$, $\gamma$ of a FODO cell, we start our calculations in the middle of the focusing quadrupole. And to start with the most simple configuration, the drift spaces between the two quadrupole magnets will really be empty and of equal length.

In Fig. 8 the optical solution of a typical arc structure is shown. Many FODO cells are connected with each other and the optical functions are calculated with a beam optics program. The plot shows the $\beta$ function in both planes and below the position of the magnet lenses, the lattice. The solid line shows the horizontal beta function, the dashed one the vertical $\beta$. Obviously, the solution for both $\beta$’s is periodic.

![Fig. 7: Schematic drawing of a symmetric FODO-cell](image)

![Fig. 8: Left: Beam optics in a periodic structure of nine FODO cells. Right: the corresponding situation calculated for a single cell.](image)

Can we understand what the optics code is doing? To answer this question we refer to a single cell: Qualitatively Fig. 8 clearly illustrates that the horizontal $\beta_x$ reaches its maximum value at the centre of the (horizontal) focusing quadrupoles and its minimum value at the defocusing lenses and vice versa for the vertical function $\beta_y$. Table 2 shows the numerical result of the optics code.

Table 2: Result of an optics calculation in a FODO lattice

<table>
<thead>
<tr>
<th>Element</th>
<th>$l$ (m)</th>
<th>$k$ (1/m²)</th>
<th>$\beta_x$ (m)</th>
<th>$\alpha_x$</th>
<th>$\phi_x$(rad)</th>
<th>$\beta_y$ (m)</th>
<th>$\alpha_y$</th>
<th>$\phi_y$(rad)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start</td>
<td>0</td>
<td>–</td>
<td>11.611</td>
<td>0</td>
<td>0</td>
<td>5.295</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>QFH</td>
<td>0.25</td>
<td>-0.0541</td>
<td>11.228</td>
<td>1.514</td>
<td>0.0110</td>
<td>5.488</td>
<td>-0.78</td>
<td>0.0220</td>
</tr>
<tr>
<td>QD</td>
<td>3.251</td>
<td>0.0541</td>
<td>5.4883</td>
<td>-0.78</td>
<td>0.2196</td>
<td>11.23</td>
<td>1.514</td>
<td>0.2073</td>
</tr>
<tr>
<td>QFH</td>
<td>6.002</td>
<td>-0.0541</td>
<td>11.611</td>
<td>0</td>
<td>0.3927</td>
<td>5.295</td>
<td>0</td>
<td>0.3927</td>
</tr>
<tr>
<td>End</td>
<td>6.002</td>
<td>–</td>
<td>11.611</td>
<td>0</td>
<td>0.3927</td>
<td>5.295</td>
<td>0</td>
<td>0.3927</td>
</tr>
</tbody>
</table>
The \( \alpha \)-function in the middle of the quadrupole is zero and, as \( \alpha(s) = -\beta'(s)/2 \), the \( \beta \) function is maximum or minimum at that position.

We know from the linear optics lecture that the phase advance of the complete machine is called the working point, and it is counted in units of \( 2\pi \). In our case we chose \( \phi = 45^\circ \) or 0.3927 rad as phase advance of one single cell and the corresponding working point would be \( Q_x = \frac{\int \phi ds}{2\pi} = 0.125 \).

Lastly, as we have chosen equal quadrupole strengths in both planes, i.e. \( k_x = -k_y \) and uniform drift spaces between the quadrupoles, the lattice is called a symmetric FODO cell. And therefore we expect symmetric optical solutions in the two transverse planes.

In linear beam-optics the transfer matrix of a number of optical elements is given by the product of the matrices of the single elements introduced in Eq. (4). In our case we get

\[
M_{FODO} = M_{QFH} * M_{L_\phi} * M_{QD} * M_{L_\phi} * M_{QFH} \, .
\]

However, as we have decided to start the calculation in the middle of a quadrupole magnet, the corresponding matrix has to take that into account. The first matrix will be that of a half-quadrupole. Putting in the numbers for the length and strength, \( k = \pm 0.54102/m^2 \), \( l_q = 0.5 \) m, \( l_d = 2.5 \) m we get

\[
M_{FODO} = \begin{pmatrix}
0.707 & 8.206 \\
-0.061 & 0.707
\end{pmatrix}.
\]

This matrix describes uniquely the optical property of the lattice and defines the beam parameters.

2.1.1 The most important point: stability of the motion

The trace of \( M \) gives

\[
|trace(M_{FODO})| = 1.415 < 2 \, .
\]

A lattice built out of these FODO cells provides stable conditions for the particle motion. However, if we introduce new parts in the lattice we have to go through the calculation again.

In addition the matrix can be used to determine the optical parameters of the system:

2.1.2 Phase advance per cell

Writing \( M \) as a function of \( \alpha, \beta, \gamma \) and the phase advance \( \phi \) we get:

\[
M(s) = \begin{pmatrix}
\cos(\phi) + \alpha \sin(\phi) & \beta \sin(\phi) \\
-\gamma \sin(\phi) & \cos(\phi) - \alpha \sin(\phi)
\end{pmatrix}
\]

and we immediately see that

\[
\cos(\phi) = \frac{1}{2} \times trace(M) = 0.707
\]

or \( \phi = 45^\circ \), which corresponds to the working point of 0.125 calculated above.

2.1.3 The \( \alpha \) and \( \beta \) functions are calculated in a similar way

For \( \beta \) we use the relation
\[ \beta = \frac{M(1,2)}{\sin(\phi)} = 11.611 \text{ m} \]

and we obtain \( \alpha \) through the expression

\[ \alpha = \frac{M(1,1) - \cos(\phi)}{\sin(\phi)} = 0 \, . \]

To complete this look at the optical properties of a lattice cell, I want to give a rule of thumb for the working point: Defining an average \( \beta \) function of the ring we put

\[ \int \frac{ds}{\beta} = \frac{L}{\bar{\beta}} \, . \]

With \( L = 2\pi \bar{R} \) and \( \bar{R} \) being the average bending radius of the ring (which is not the bending radius of the dipole magnets) for the working point \( Q \) we can write

\[ Q = N^* \frac{\phi_c}{2\pi} = \frac{1}{2\pi} \int \frac{ds}{\beta(s)} = \frac{1}{2\pi} \frac{2\pi \bar{R}}{\bar{\beta}} \, , \]

where \( N \) is the number of cells and \( \phi_c \) denotes the phase advance per cell. So we get

\[ Q \approx \frac{\bar{R}}{\bar{\beta}} \, . \]

A rough estimate for the working point is obtained by the ratio of the mean radius of the ring and the average \( \beta \) function of the lattice.

### 2.2 Thin-lens approximation

As discussed, a first estimate of the parameters of a lattice can and should be carried out at the beginning of a magnet lattice design.

If we want fast answers and require only rough estimates we can simplify. Under certain circumstances the matrix of a focusing element can be written in the thin-lens approximation. Given, for example, the matrix of a focusing lens

\[ M_{QF} = \begin{pmatrix} \cos(\sqrt{K} \ast \ell) & \frac{1}{\sqrt{K}} \sin(\sqrt{K} \ast \ell) \\ -\sqrt{K} \sin(\sqrt{K} \ast \ell) & \cos(\sqrt{K} \ast \ell) \end{pmatrix} \, , \]

we can simplify the trigonometric terms if the focal length of the quadrupole magnet is much larger than the length of the lens:

\[ f = \frac{1}{k\ell_Q} \gg \ell_Q , \]

the transfer matrix can be approximated using \( kl_Q = \text{const}, l_Q \rightarrow 0 \) and we get

\[ M_{QF} = \begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix} \, . \]
Referring to the notation used in Fig. 7 we can calculate the transfer matrix first from the middle of the focusing to the middle of the defocusing quadrupole and get the matrix for half the cell:

\[
M_{\text{halfCell}} = M_{QD/2}M_{\ell_d}M_{QF/2}
\]

\[
M_{\text{halfCell}} = \begin{pmatrix}
1 & 0 \\
\frac{1}{\ell_f} & 1
\end{pmatrix} \begin{pmatrix}
1 & \ell_D \\
0 & 1
\end{pmatrix} \begin{pmatrix}
1 & 0 \\
-\frac{1}{\ell_f} & 1
\end{pmatrix}
\]

\[
M_{\text{halfCell}} = \begin{pmatrix}
1 - \frac{\ell_D}{\ell_f} & \ell_D \\
-\frac{\ell_D}{\ell_f^2} & 1 + \frac{\ell_D}{\ell_f}
\end{pmatrix}.
\] (10)

Note that the thin-lens approximation implies that \( l_0 \to 0 \), therefore the drift between the magnets has to be \( \ell_d = L/2 \). As we are now dealing with half-quadrupoles, for the focal length of a half-quadrupole we have set \( \ell_f = 2f \). We get the second half of the cell by replacing \( \ell_f \) by \(-\ell_f\), and the matrix for the complete FODO in thin-lens approximation is

\[
M = \begin{pmatrix}
1 + \frac{\ell_D}{\ell_f} & \ell_D \\
-\frac{\ell_D}{\ell_f^2} & 1 - \frac{\ell_D}{\ell_f}
\end{pmatrix} \begin{pmatrix}
1 - \frac{\ell_D}{\ell_f} & \ell_D \\
-\frac{\ell_D}{\ell_f^2} & 1 + \frac{\ell_D}{\ell_f}
\end{pmatrix}
\]

or multiplying

\[
M = \begin{pmatrix}
1 - \frac{2\ell_D^2}{\ell_f^2} & 2\ell_D(1 + \frac{\ell_D}{\ell_f}) \\
2\ell_D(1 + \frac{\ell_D}{\ell_f}) - \frac{2\ell_D}{\ell_f^2} & 1 - \frac{2\ell_D^2}{\ell_f^2}
\end{pmatrix}.
\] (11)

The matrix is now much easier to handle than the equivalent formula (4) and (8). And the approximation is, in general, not bad.

Going briefly through the calculation of the optics parameters again, according to (9) and (11) we immediately get

\[
\cos(\phi) = 1 - \frac{2\ell_D^2}{\ell_f^2},
\]

and with a little bit of trigonometric gymnastics

\[
1 - 2\sin^2(\phi/2) = 1 - \frac{2\ell_D^2}{\ell_f^2}
\]

we can simplify this expression and get

\[
\sin(\phi/2) = \ell_D / \ell_f = \frac{L_{\text{Cell}}}{2f}.
\]
and finally
\[ \sin\left(\frac{\varphi}{2}\right) = \frac{L_{\text{cell}}}{4f}. \]  

(12)

In thin-lens approximation the phase advance of a FODO cell is given by the length of the cell \( L_{\text{cell}} \) and the focal length of the quadrupole magnets \( f \).

For the parameters of the example given above, we get a phase advance per cell of \( \varphi \approx 47.8^\circ \), and in full analogy to the calculation presented previously, we calculate \( \beta \approx 11.4 \text{ m} \), which is very close to the result of the exact calculations \( (\varphi = 45^\circ, \beta = 11.6 \text{ m}) \).

2.2.1 Stability of the motion

In thin-lens approximation, the condition for stability \(|\text{trace}(M)| < 2\) requires that

\[ 2 - \frac{4f^2}{f^2} < 2 \]

or

\[ f > \frac{L_{\text{cell}}}{4}. \]  

(13)

We have the important and simple result that for stable motion the focal length of the quadrupole lenses in the FODO has to be larger than a quarter of the length of the cell.

2.3 Scaling optical parameters of a lattice cell

After the discussion on stability in a lattice cell and the first estimates and calculations of the optical functions \( \alpha, \beta, \gamma \) and \( \varphi \), we would like to concentrate a little more on a detailed analysis of a FODO with these parameters.

We can calculate the \( \beta \)-function that corresponds to the periodic solution — provided we know the strength and length of the focusing elements in the cell. But can we optimize somehow?

In other words, for a given lattice, what would be the ideal magnet strength to get the smallest beam dimensions? To answer this question, look at the transfer matrix of half a FODO cell in Eq. (10), i.e. the transfer from the middle of a QF quadrupole to the middle of a QD (see Fig. 7).

From linear beam optics we know that the transfer matrix between two points in a lattice can be expressed not only as a function of the focusing properties of the elements in that section of the ring, but in an equivalent way as a function of the optical parameters between the two reference points. For a full turn or, within a periodic lattice for one period, we have used that relation already in Eq. (5).

The general expression — in the non periodic case — reads [5]

\[
M_{1\rightarrow2} = \begin{pmatrix}
\frac{\beta_2}{\sqrt{\beta_1}} (\cos \Delta \varphi + \alpha_1 \sin \Delta \varphi) & \sqrt{\beta_2} \beta_1 \sin \Delta \varphi \\
(\alpha_1 - \alpha_2) \cos \Delta \varphi - (1 + \alpha_1 \alpha_2) \sin \Delta \varphi & \frac{1}{\sqrt{\beta_2}} (\cos \Delta \varphi - \alpha_2 \sin \Delta \varphi)
\end{pmatrix}.
\]  

(14)
The indices refer to the starting point \( s_1 \) and the end point \( s_2 \) in the ring and \( \Delta \phi \) is the phase advance between these points. It is evident that this matrix is reduced to the form given in Eq. (5) if the periodic conditions \( \beta_1 = \beta_2, \alpha_1 = \alpha_2 \) are fulfilled.

In the middle of the focusing quadrupole \( \beta \) reaches its highest value and in the middle of the defocusing magnet its lowest one (referring to the vertical plane the argument is valid ‘vice versa’) and the \( \alpha \) functions at that position are zero. Therefore the transfer matrix can be written in the form

\[
M = \begin{pmatrix} C & S \\ C' & S' \end{pmatrix} = \begin{pmatrix} \sqrt{\beta} \cos \Delta \phi & \sqrt{\beta} \sin \Delta \phi \\ -1 & \sin \Delta \phi \end{pmatrix} \begin{pmatrix} \sqrt{\beta} \cos \Delta \phi & \sqrt{\beta} \sin \Delta \phi \\ -1 & \sin \Delta \phi \end{pmatrix}.
\]

Using this expression and putting the terms that we have developed in thin-lens approximation in Eq. (10) for the matrix elements, we get

\[
\hat{\beta} = \frac{S'}{C'} = \frac{1 + \ell_D \bar{f}}{1 - \ell_D \bar{f}} = \frac{1 + \sin \frac{\phi}{2}}{1 - \sin \frac{\phi}{2}},
\]

\[
\hat{\beta} = \frac{-S}{C'} = \bar{f}^2 = \frac{\ell_D^2}{\sin^2 \frac{\phi}{2}}.
\]

where we have set \( \Delta \phi = \phi/2 \) for the phase advance of half the FODO cell. The two expressions for \( \hat{\beta} \) and \( \beta \) can be combined to calculate both parameters

\[
\hat{\beta} = \frac{(1 + \sin \frac{\phi}{2})L}{\sin \phi}, \quad \beta = \frac{(1 - \sin \frac{\phi}{2})L}{\sin \phi}.
\]

We get the simple result that the maximum (and minimum) value of the \( \beta \)-function and therefore the maximum dimension of the beam in the cell is determined by the length \( L \) and the phase advance \( \phi \) of the complete cell.

Figure 9 shows a three-dimensional picture of a proton bunch for typical conditions in the HERA storage ring. The bunch length is about 30 cm and determined by the momentum spread and the RF potential [7]. The values of \( \hat{\beta} \) and \( \beta \), as determined by the cell characteristics, are normally 80 m and 40 m, respectively.
2.3.2 Optimization of the FODO phase advance

From Eq. (15) we see that — given the length of the FODO — the maximum $\beta$ depends only on the phase advance per cell. Therefore, we may ask whether there is an optimum phase that leads to the smallest beam dimension.

If we assume a Gaussian particle distribution in the transverse plane and denote the beam emittance by $\varepsilon$, the transverse beam dimension $\sigma$ is given by
\[
\sigma = \sqrt{\varepsilon \beta}.
\]
In a typical high-energy proton ring, $\varepsilon$ is in the order of some $10^{-9}$ m*rad (for example, the HERA proton ring at $E = 920$ GeV, $\varepsilon \approx 6 \times 10^{-9}$ m*rad) and as the typical $\beta$ functions are about $\beta \approx 40 \ldots 100$ m in the arc, the resulting beam dimension is roughly one millimetre. At the interaction point of two counter-rotating beams even beam radii in the order of a micrometre are obtained.

Figure 10 shows the result of a beam scan used to measure the transverse beam dimension.

![Fig. 10: Transverse beam profile of a HERA proton bunch at the interaction point. The measurement is performed by scanning the colliding beams against each other.](image)

In general both emittances are equal for a proton beam, $\varepsilon_x \approx \varepsilon_y$. A proton beam is ‘round’ even if the varying beta function along the lattice leads to beam dimensions in two transverse planes that can be quite different.

Optimizing the beam dimension in the case of a proton ring therefore means searching for a minimum of the beam radius:
\[
r^2 = \varepsilon_x \beta_x + \varepsilon_y \beta_y
\]
and therefore optimizing the sum of the maximum and minimum beta functions at the same time:

$$\hat{\beta} + \beta = \frac{(1 + \sin \frac{\phi}{2})L}{\sin \phi} + \frac{(1 - \sin \frac{\phi}{2})L}{\sin \phi}.$$  

(16)

The optimum phase $\phi$ is obtained by

$$\frac{d}{d\phi}(\hat{\beta} + \beta) = \frac{d}{d\phi}(\frac{2L}{\sin \phi}) = 0$$

which gives

$$\frac{L}{\sin^2 \phi} \cos \phi = 0 \rightarrow \phi = 90^\circ .$$

As for the aperture requirement of the cell, a phase advance of $\phi = 90^\circ$ is the best value in a proton ring. The plot of Fig. 11 shows the sum of the two $\beta$’s (Eq. 16) as a function of the phase $\phi$ in the range of $\phi = 0 \ldots 180^\circ$.

![Fig. 11: Sum of the horizontal and vertical $\beta$ function as a function on the phase advance $\phi$](image)

Optimization of the beam radii can be a critical issue in accelerator design. Large beam dimensions need big apertures of the quadrupole and dipole magnets in the ring. Running the machine at the highest energy can lead to limitations in the focusing power as the gradient of a quadrupole lens scales as the inverse of its squared aperture radius $r_0^2$ and increases the cost for the magnet lenses. Therefore it is recommended not to tune the lattice too far away from the ideal phase advance.

Here, for completeness, I have to make a short remark on electron machines. Unlike the situation in proton rings, electron beams are flat in general: On account of synchrotron radiation [8], the vertical emittance of an electron or positron beam is only a small fraction of the horizontal one $\varepsilon_y \approx 1\ldots10 \% \varepsilon_x$. For the optimization of the phase advance, the calculation can and should be restricted to the horizontal plane only and the condition for smallest beam dimension is

$$\frac{d}{d\phi}(\hat{\beta}) = \frac{d}{d\phi}(\frac{L(1 + \sin \frac{\phi}{2})}{\sin \phi}) = 0 \rightarrow \phi \approx 76^\circ .$$

Figure 12 shows the horizontal and vertical $\beta$ as a function of $\phi$ in that case. In an electron ring the typical phase advance is in the range of $\phi \approx 30 \ldots 90^\circ$. 

48
2.4 Dispersion in a FODO lattice

In the design and description of the optical parameters of a magnet lattice, we have discussed the scenario of particles with ideal momentum, and as such, have not deviated from standard information presented in the literature on linear beam dynamics. In general, the energy (or momentum) of the particles stored in a ring deviate from the ideal momentum $p_0$ of the beam.

In linear beam-optics, the differential equation for the transverse movement gets an additional term if the momentum deviation is not zero: $\Delta p/p \neq 0$. This gives an inhomogeneous equation of motion

$$x^\prime\prime + K(s)x = \frac{1}{\rho} \frac{\Delta p}{p}.$$  \hspace{1cm} (17)

The left-hand side of (17) is the same as in the homogeneous Eq. (3) and the parameter $K$ describes the focusing strength of the lattice element at position $s$ in the ring. As usual, the general solution of Eq. (17) is the sum of the complete solution $x_{h}$ of the homogeneous equation and a special solution of the inhomogeneous one $x_{i}$.

$$x^*_{h} + K(s)x_{i} = 0$$
$$x^*_{i} + K(s)x_{i} = \frac{1}{\rho} \frac{\Delta p}{p_0}.$$  

The special solution $x_{i}$ can be normalized to the momentum error $\Delta p/p$ giving the dispersion function $D(s)$

$$x_{i}(s) = D(s) \Delta p/p.$$ \hspace{1cm} (18)

Starting from the initial conditions $x_{0}$ and $x^\prime_{0}$, the general solution for the particle trajectory now reads

$$x(s) = C(s)x_{0} + S(s)x_{0}^\prime + D(s) \Delta p/p$$

or, including the expression for the angle $x^\prime(s)$,

$$\begin{pmatrix} x \\ x' \end{pmatrix}_{s} = \begin{pmatrix} C & S \\ C' & S' \end{pmatrix} \begin{pmatrix} x \\ x' \end{pmatrix}_{0} + \frac{\Delta p}{p} \begin{pmatrix} D \\ D' \end{pmatrix}.$$  

LATTICE DESIGN IN HIGH-ENERGY PARTICLE ACCELERATORS
For convenience, extend the matrix to include the second term and write

\[
\begin{pmatrix} x \\ x' \\ \Delta p/p \end{pmatrix}_S = \begin{pmatrix} C & S & D \\ C' & S' & D' \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ x' \\ \Delta p/p \end{pmatrix}_0 .
\]

The dispersion function \( D(s) \) is (obviously) defined by the focusing properties of the lattice and the bending strength of the dipole-magnets \( 1/\rho \) and it can be shown that [5]

\[
D(s) = S(s) \int_{s_0}^{s} \frac{1}{\rho(t)} C(\delta) d\delta - C(s) \int_{s_0}^{s} \frac{1}{\rho(\delta)} S(\delta) d\delta .
\]

(19)

The variable \( s \) refers to the position where the dispersion is obtained (or measured if you like) and the integration has to be performed over all places \( \delta \) where a non-vanishing term \( 1/\rho \) exists (in general in the dipole magnets of the ring).

2.4.1 Example

The 2 \( \times \) 2 matrix for a drift space is given by

\[
M_{\text{Drift}} = \begin{pmatrix} C & S \\ C' & S' \end{pmatrix} = \begin{pmatrix} 1 & \ell \\ 0 & 1 \end{pmatrix} .
\]

As there are no dipoles in the drift, the \( 1/\rho \) term in Eq. (19) is zero and we get the extended 3 \( \times \) 3 matrix

\[
M_{\text{Drift}} = \begin{pmatrix} 1 & \ell & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} .
\]

To calculate the dispersion in a FODO cell we refer back to the thin-lens approximation used for the beta-function calculation. The matrix for a half-cell has been derived in Eq. (10). In thin-lens approximation, the length \( \ell \) of the drift is just half the length of the cell, as the quadrupole lenses have zero length.

\[
M_{\text{half Cell}} = \begin{pmatrix} C & S \\ C' & S' \end{pmatrix} = \begin{pmatrix} 1 - \frac{\ell}{f} & \ell \\ -\frac{\ell}{f^2} & 1 + \frac{\ell}{f} \end{pmatrix} .
\]

Using this expression we can calculate the terms \( D, D' \) of the 3 \( \times \) 3 matrix

\[
D(s) = S(s) \int_{s_0}^{s} \frac{1}{\rho(\delta)} C(\delta) d\delta - C(s) \int_{s_0}^{s} \frac{1}{\rho(\delta)} S(\delta) d\delta
\]

\[
D(\ell) = \ell \int_{0}^{\ell} \left( 1 - \frac{s}{f} \right) ds - \left( 1 - \frac{\ell}{f} \right) \int_{0}^{s} \frac{1}{\rho} \int_{s}^{s} ds
\]
In an analogous way one derives the expression for $D'$

$$D'(\ell) = \frac{\ell}{\rho} \left( 1 + \frac{\ell}{2f} \right)$$

and we get the complete matrix for a FODO half-cell

$$M_{\text{half\text{cell}}} = \begin{pmatrix} C & S & D \\ C' & S' & D' \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 - \frac{\ell}{f} & \frac{\ell}{f} & \frac{\ell^2}{2f} \\ -\frac{\ell}{f^2} & 1 + \frac{\ell}{f} & \frac{\ell}{f}\left(1 + \frac{\ell}{2f}\right) \\ 0 & 0 & 1 \end{pmatrix}.$$  

Fig. 13: Beta function (upper part). Dispersion (lower part) function in a FODO cell.

Because of symmetry, the dispersion in a FODO lattice reaches its maximum value in the centre of a QF quadrupole and its minimum in a QD, as shown in Fig. 13 where, in addition to the $\beta$ function, the dispersion is shown in the lower part of the figure.

Therefore we get the boundary conditions for the transformation from a QF to a QD lens

$$\begin{pmatrix} \dot{D} \\ 0 \\ 1 \end{pmatrix} = M_{1/2}^* \begin{pmatrix} \dot{D} \\ 0 \\ 1 \end{pmatrix}$$
to calculate the dispersion at these locations.

\[
\hat{\mathcal{D}}(\hat{\mathcal{R}}) = \hat{\mathcal{D}}(1 - \frac{\ell}{f}) + \frac{\ell^2}{2\rho}
\]

\[
0 = -\frac{\ell}{f^2} \hat{\mathcal{D}} + \frac{\ell}{\rho} (1 + \frac{\ell}{2f}) .
\]

Remember that we have to use the focal length of a half-quadrupole

\[
\tilde{f} = 2f
\]

and that the phase advance is given by

\[
\sin(\phi/2) = \frac{L_{\text{Cell}}}{2\tilde{f}} .
\]

For maximum dispersion in the middle of a focusing quadrupole and for the minimum dispersion in the middle of a defocusing lens, we get the expressions

\[
\hat{D} = \frac{\ell^2}{\rho} \left(1 + \frac{1}{2}\sin\frac{\phi}{2}\right) \cos\frac{\phi}{2}
\]

(20)

\[
\hat{D} = \frac{\ell^2}{\rho} \left(1 - \frac{1}{2}\sin\frac{\phi}{2}\right) \cos\frac{\phi}{2}
\]

Note that dispersion depends on the half-length \(\ell\) of the cell, the bending strength of the dipole magnet \(1/\rho\), and the phase advance \(\phi\).

The dependence of \(D\) on the phase advance is shown in the plot of Fig. 14. Both values \(D_{\text{max}}\) and \(D_{\text{min}}\) decrease for an increasing phase \(\phi\) (which is just another way to say ‘for increasing focusing strength’, as \(\phi\) depends on the focusing strength of the quadrupole magnets).

To summarize:

- small dispersion needs strong focusing and therefore large phase-advance;
- there is, however, an optimum phase advance concerning the best (i.e. smallest) value of the \(\beta\) function;
- the stability criterion limits the choice of the phase advance per cell.
It is necessary to find a compromise for the focusing strength in a lattice that takes into account the stability of the motion, the $\beta$ function in both transverse planes, and the dispersion function. In a typical high-energy machine, this optimization is not too difficult as the dispersion does not impact much on the beam parameters (as long as it is compensated at the interaction point of the two beams).

In synchrotron light sources, the beam emittance is usually the parameter to be optimized (that means in nearly all cases minimized), and as the emittance depends on the dispersion $D$ in an electron storage ring, the dispersion function and its optimization is of greatest importance in these machines.

The horizontal beam emittance in an electron ring is given by the expression

$$\epsilon_x = \frac{55}{32\sqrt{3}} \frac{\hbar}{mc} \gamma^2 \frac{1}{J_x} \left\langle \frac{1}{R^2} H(s) \right\rangle$$

where the function $H(s)$ is defined by

$$H(s) = \gamma D^2 + 2\alpha DD' + \beta D'^2.$$  

The optimization of $H(s)$ in a magnet structure is a subject of its own and an introduction to that field of so-called low-emittance lattices can be found in the contribution of A. Streun in this school [9].

### 2.5 Orbit distortions in a periodic lattice

The lattice we have designed consists of a small number of basic elements.

Bending magnets define the geometry of the circular accelerator and for a given particle momentum, the size of the machine. Quadrupole lenses define the phase advance of the single particle trajectory and through this parameter define the beam dimensions and the stability of the motion.

To fill the empty spaces in the lattice cell with some useful other components, we have to talk about the ‘O-s’ of the FODO.
Field errors in storage rings have been discussed in the linear beam optics lecture and we know that in the case of a dipole magnet an error of the bending field is described by an additional kick $\delta$ (typically measured in mrad) on the particles

$$\delta = \frac{ds}{\rho} = \int \frac{Bds}{p/e}.$$  

The beam oscillates in the corresponding plane and the resulting orbit amplitude is

$$x(s) = \sqrt{\frac{\beta(s)}{2\sin \pi Q}} \sqrt{\frac{\beta(s)}{\rho(s)}} \cos(\varphi(s) - \varphi(\tilde{s}) - \pi Q) ds \tag{21}.$$  

This is given by the beta function at the place of the dipole magnet $\beta(s)$ and its bending strength $1/\rho$, and the beta function at the observation point in the lattice $\beta(s)$.

For the lattice designer that means if a correction magnet has to be installed in the lattice cell, it should be placed at a location where $\beta$ is high in the corresponding plane.

At the same time Eq. (21) tells us that the amplitude of an orbit distortion is highest at the place where $\beta$ is high and this is the place where beam-position monitors have to be located to measure the orbit distortion precisely.

In practice, both beam-position monitors and orbit-correction coils are located at places in the lattice cell where the $\beta$ function in the considered plane is large, i.e. close to the corresponding quadrupole lens.

### 2.6 Chromaticity in a FODO cell

The chromaticity $\xi$ describes an optical error of a quadrupole lens in an accelerator. For a given magnetic field, i.e. the gradient of the quadrupole magnet, particles with smaller momentum will feel a stronger focusing force and vice versa.

The chromaticity $\xi$ relates the resulting tune shift to the relative momentum error of the particle

$$\Delta Q = \xi \frac{\Delta p}{p}$$

and as it is a consequence of the focusing properties of the quadrupole magnets, it is given by the characteristics of the lattice. For small momentum errors $\Delta p/p$ the focusing parameter $k$ can be written

$$k(p) = \frac{g}{p/e} = g \frac{e}{p_0 + \Delta p}$$

where $g$ denotes the gradient of the quadrupole lens, $p_0$ the design momentum, and the term $\Delta p$ refers to the momentum error. If $\Delta p$ is small as we have assumed, we can write

$$k(p) \approx g \frac{e}{p_0} \left(1 - \frac{\Delta p}{p} \right) = k + \Delta k.$$  

This describes a quadrupole error

$$\Delta k = -k_0 \frac{\Delta p}{p}.$$
and leads to a tune shift of

$$\Delta Q = \frac{1}{4\pi} \int \Delta k * \beta(s) ds$$

$$\Delta Q = -\frac{1}{4\pi} \frac{\Delta p}{p} \int k_s \beta(s) ds .$$

By definition, the chromaticity $\xi$ of a lattice is given by

$$\xi = -\frac{1}{4\pi} \int \beta(s) * k(s) ds . \quad (22)$$

Assuming the accelerator consists of $N$ identical FODO cells then, replacing the $\beta(s)$ by its maximum value at the focusing and its minimum value at the defocusing quadrupoles permits the approximation of the integral by the sum

$$\xi \approx -\frac{1}{4\pi} \frac{N * \hat{\beta} - \hat{\beta}}{f_Q} ;$$

$f_Q = 1/(k*\ell)$ denotes the focal length of the quadrupole magnet.

$$\xi \approx -\frac{1}{4\pi} \frac{N * \hat{\beta} - \hat{\beta}}{f_Q} \left\{ \frac{L(1 + \sin \frac{\varphi}{2}) - L(1 - \sin \frac{\varphi}{2})}{\sin \varphi} \right\} \quad (23)$$

uses the expressions from Eq. (15) for $\hat{\beta}$ and $\hat{\beta}$. With some useful trigonometric transformations like

$$\sin x = 2 \sin \frac{x}{2} \cos \frac{x}{2}$$

we can transform the right-hand side of (23) to get

$$\xi = -\frac{1}{4\pi} \frac{N}{f_Q} \frac{L * \sin \frac{\varphi}{2}}{\sin \varphi \cos \frac{\varphi}{2}}$$

or for one single cell, $N = 1$,

$$\xi_{Cell} = -\frac{1}{4\pi} \frac{1}{f_Q} \frac{L * \tan \frac{\varphi}{2}}{\sin \frac{\varphi}{2}} .$$

Remembering the relation

$$\sin \frac{\varphi}{2} = \frac{L}{4f_Q}$$
we obtain a surprisingly simple result for the chromaticity contribution of a FODO cell.

\[ \xi_{\text{cell}} = -\frac{1}{\pi} \tan \frac{\varphi}{2} \cdot \]

3 LATTICE INSERTIONS

In Fig. 6, the lattice of a typical machine for the acceleration of high-energy particles consists of two quite different parts: The arcs built out of a number of identical cells, and the straight sections that connect them and that house complicated systems like dispersion suppressors, mini-beta insertions, or high-energy particle detectors.

3.1 Drift space

For an insight into the design of lattice insertions, it is useful to look at a simple drift space embedded in a normal lattice structure.

What happens to the beam parameters \( \alpha \), \( \beta \) and \( \gamma \) if we stop focusing for a while? The transfer matrix for the Twiss parameters from a point ‘0’ to position ‘s’ in a lattice is given by the formula

\[
\begin{pmatrix}
\beta \\
\alpha \\
\gamma \\
\end{pmatrix}
_s =
\begin{pmatrix}
C^2 & -2SC & S^2 \\
-CC' & SC' + S'C & -SS' \\
C'^2 & -2S'C' & S'^2 \\
\end{pmatrix}
\begin{pmatrix}
\beta \\
\alpha \\
\gamma \\
\end{pmatrix}
_0
\]

(24)

where the cosine and sine functions \( C \) and \( S \) are given by the focusing properties of the lattice elements between the two points.

For a drift space of length \( s \) this is according to Eq. (4) as simple as

\[
M = \begin{pmatrix}
C & S \\
C' & S' \\
\end{pmatrix} = \begin{pmatrix}
1 & s \\
0 & 1 \\
\end{pmatrix}
\]

and the optical parameters develop as a function of \( s \)

\[
\beta(s) = \beta_0 - 2\alpha s + \gamma_0 s^2 \\
\alpha(s) = \alpha_0 - \gamma_0 s \\
\gamma(s) = \gamma_0
\]

(25)

We now examine the relations more closely.

3.1.1 Location of a beam waist

The first equation indication that if the drift space is long enough, even a convergent beam at position ‘0’ will become divergent as the term \( \gamma_0 s^2 \) is always positive, see Fig. 15.
Therefore there will be a point in the drift where the beam dimension is smallest, in other words where the beam envelope has a waist. The position of this waist can be calculated by

$$\alpha(s_1) = 0$$;

the second equation of (25) then gives

$$\alpha_0 = \gamma_0 s_1$$

or

$$s_1 = \frac{\alpha_0}{\gamma_0}.$$  

The position of the waist is given by the ratio of the $\alpha$- and $\gamma$-function at the beginning of the drift.

As the $\gamma$ parameter is constant in a drift and $\alpha$ is zero at the waist, we can directly calculate the beam size that we get at the waist.

$$\gamma(s_1) = \gamma_0, \quad \alpha(s_1) = 0$$

$$\beta(s_1) = \frac{1 + \alpha^2(s_1)}{\gamma(s_1)} = \frac{1}{\gamma_0}.$$  \hspace{1cm} (26)

The beta function at the location of the waist is given by the inverse of the gamma function at the beginning of the drift; a nice, simple scaling law.

3.1.2 Beta function in a drift space

To expand on the behaviour of the Twiss parameters in a drift, the scaling of $\beta$ with the length of the drift has a large impact on the design of high-energy machines. Assuming that we are in the centre of a drift and the situation is symmetric, index ‘0’ refers to the position at the starting point, but we now want to have a left–right symmetric optics with respect to it, $a_0 = 0$. From Eq. (25) at the starting point we get

$$\beta(s) = \beta_0 - 2a_0 s + \gamma_0 s^2$$

and as we know from Eq. (26) that at the waist $s = 0$, we have

$$\gamma_0 = \frac{1 + \alpha_0^2}{\beta_0} = \frac{1}{\beta_0}$$

and we get $\beta$ as a function of the distance $s$ from the starting point:
The relation (27) is a direct consequence of Liouville’s theorem: The density of the particle’s phase space is constant in an accelerator. In other words, with conservative forces only, beam emittance $\varepsilon$ is constant, leading immediately to the relation (24). The conservation of $\varepsilon$ is a fundamental law that cannot be avoided and there is no way to overcome the increase of the beam dimension in a drift.

– The behaviour of $\beta$ in a drift has a strong impact on storage-ring design. As large beam dimensions have to be avoided, large drift spaces are forbidden or at least very inconvenient. In Section 3.2, we see that this is one of the major limitations of the luminosity of colliding beams in an accelerator.

At the beam waist we can derive another short relation often used for scaling beam parameters. The beam envelope $\sigma$ is given by the beta function and the emittance of the beam,

$$\sigma(s) = \sqrt{\varepsilon^* \beta}$$

and the divergence $\sigma'$ by

$$\sigma'(s) = \sqrt{\varepsilon^* \gamma}.$$  

Now, as $\gamma = (1 + \alpha^2)/\beta$, wherever $\alpha = 0$ the beam envelope has a local minimum (i.e. a waist) or maximum. At that position the beta function is just the ratio of the beam envelope and the beam divergence.

$$\beta(s) = \frac{\sigma(s)}{\sigma'(s)} \text{ at a waist.}$$

If we cannot fight against Liouville’s theorem, we can at least try to optimize its consequences. Equation (27) for $\beta$ in a symmetric drift can be used to find the starting value that gives the smallest beam dimension at the end of the drift of length $\ell$.

Setting

$$\frac{d\hat{\beta}}{d\beta_0} = 1 - \frac{\ell^2}{\beta_0^2} = 0$$

gives us the value of $\beta_0$ that leads to the smallest $\beta$ after a drift of length $\ell$:

$$\beta_0 = \ell.$$  

For a starting value of $\beta_0 = \ell$, the maximum beam dimension at the end of the drift will be smallest and its value is just double the length of the drift

$$\hat{\beta} = 2\beta_0 = 2\ell.$$  

### 3.2 Mini-beta insertions

Section 3.1 showed that the $\beta$ function in a drift space can be chosen with respect to the length $\ell$ to minimize the beam dimension and accordingly the aperture requirements for vacuum chambers and magnets. In general, the value of $\beta$ is in the order of a few metres and the typical length of drift spaces in a lattice is of the same order.
However, the straight sections of a storage ring are often designed for the collision of two counter-rotating beams. The beta functions at the collision points are therefore very small compared to their values in the arc cells. Typical values are more in the range of centimetres than of metres. Still, the same scaling law of Eq. (28) holds and the optimum length of such a drift would, for example, be \( \ell \approx 36 \text{ cm} \) for the interaction regions of the two beams in the HERA collider.

Modern high-energy detectors, in contrast, are impressive devices consisting of many large components that do not fit into a drift space of a few centimetres. Figure 16 shows the ZEUS detector at the HERA collider. Obviously, to install a huge detector like this, the storage-ring lattice needs special treatment.

The lattice has to be modified before and after the interaction point, to establish a large drift space where the high-energy experiment detector can be embedded. At the same time, the beams have to be strongly focused to get very small beam dimensions in both transverse planes at the collision point, in other words to get high luminosity. This lattice structure is called ‘mini-beta insertion’.

The luminosity of a particle collider is defined by the event rate \( R \) of a special reaction (e.g. a particle produced by beam collision).

\[
R = \sigma_R \cdot L.
\]

The production rate \( R \) of a reaction is given by its cross-section \( \sigma_R \) and a number that is the result of the lattice design: the luminosity \( L \) of the storage ring. It is given by the beam optics at the collision point and the amount of the stored beam currents [10].

\[
L = \frac{1}{4\pi e^2 f_0 b} \cdot \frac{I_1 \cdot I_2}{\sigma_x^* \cdot \sigma_y^*}.
\]

Here \( I_1 \) and \( I_2 \) are the values of the stored beam currents, \( f_0 \) is the revolution frequency of the machine, and \( b \) the number of stored bunches. \( \sigma_x^* \) and \( \sigma_y^* \) in the denominator are the beam sizes in the horizontal and vertical plane at the interaction point. For a high-luminosity collider, the stored beam currents have to be high and at the same time the beams have to be focused at the interaction point to very small values.
Figure 17 shows the typical layout of a mini-beta-insertion scheme. It generally consists of
– a symmetric drift space large enough to house the particle detector and whose beam waist
$\alpha_0 = 0$ is centred at the interaction point of the colliding beams;
– a quadrupole doublet (or triplet) on each side as close as possible;
– additional quadrupole lenses to match the Twiss parameters of the mini-beta insertion to
the optical parameters of the lattice cell in the arc.

As a mini-beta scheme is always a kind of symmetric drift space, we can apply the formula
derived above. For $\alpha = 0$, we get a quadratic increase of the beta function in the drift and at the
distance $\ell_1$ of the first quadrupole lens we get

$$\beta(s) = \beta_0 + \frac{\ell^2}{\beta_0}.$$  

The size of the beam at the position of the second quadrupole can be calculated in a similar way. According to Fig. 18, the transfer matrix of the quadrupole doublet system consists of four parts:

Two drifts with lengths $\ell_1$ and $\ell_2$ and a focusing and defocusing quadrupole magnet. Starting at
the IP we get — again in thin-lens approximation:

$$M_{D_1} = \begin{pmatrix} 1 & \ell_1 \\ 0 & 1 \end{pmatrix}, \quad M_{f_1} = \begin{pmatrix} 1 \\ 1/\ell_1 \\ 1 \end{pmatrix},$$

$$M_{D_2} = \begin{pmatrix} 1 & \ell_2 \\ 0 & 1 \end{pmatrix}, \quad M_{f_2} = \begin{pmatrix} 1 \\ -1/\ell_2 \\ 1 \end{pmatrix}. $$
In general the first lens of this system focuses in the vertical plane and therefore according to the sign convention used in this School the focal length is positive, \( \frac{1}{f_1} > 0 \). The matrix for the complete system is

\[
M = M_{QF} * M_{D2} * M_{QD} * M_{D1}
\]

\[
M = \begin{pmatrix}
1 & 0 & \ell_2 \\
-1/f_2 & 1 & 0 \\
1/f_1 & 0 & 1
\end{pmatrix} * \begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
1 & 0 & 1
\end{pmatrix} * \begin{pmatrix}
1 & \ell_1 \\
0 & 1 \\
1 & 0
\end{pmatrix} .
\]

Multiplying out we get

\[
M = \begin{pmatrix}
\frac{1 + \ell_2}{f_1} & \ell_1 + \ell_2 + \frac{\ell_1 \ell_2}{f_1} \\
-\frac{1}{f_2} & -\frac{\ell_1}{f_1 f_2} - \frac{\ell_2}{f_2} + \frac{\ell_1}{f_2} + 1
\end{pmatrix} = \begin{pmatrix}
C & S \\
C' & S'
\end{pmatrix} .
\]

Remembering the transformation of the Twiss parameters in terms of matrix elements [see Eq. (6)]

\[
\begin{pmatrix}
\alpha \\
\beta \\
\gamma
\end{pmatrix}_s = \begin{pmatrix}
C^2 & -2CS & S^2 \\
-CC' & SC' + CS' & -SS' \\
C'^2 & -2S'C' & S'^2
\end{pmatrix} * \begin{pmatrix}
\alpha \\
\beta \\
\gamma
\end{pmatrix}_0
\]

we put in the above terms and obtain

\[
\beta(s) = C^2 \beta_0 + 2SC\alpha_0 + S^2 \gamma_0 .
\]

Here the index ‘0’ denotes the interaction point and ‘s’ refers to the position of the second quadrupole lens. As we are starting at the IP where \( \alpha_0 = 0 \) and \( \gamma_0 = 1 / \beta_0 \), we can simplify and get

\[
\beta(s) = C^2 \beta_0 + S^2 / \beta_0
\]

\[
\beta(s) = \beta_0 * \left( 1 + \frac{\ell_2}{f_1} \right)^2 + \frac{1}{\beta_0} * \left( \ell_1 + \ell_2 + \frac{\ell_1 \ell_2}{f_1} \right)^2 .
\]

This formula for \( \beta \) at the second quadrupole lens is very useful when the gradient and aperture of the mini-beta quadrupole magnet have to be designed.
3.2.1 Phase advance in a mini-beta insertion

Unlike the situation in the arc where the phase advance is a function of the focusing properties of the cell, in a mini-beta insertion or in any long drift space it is, for the most part, a constant: As we know from linear beam optics, the phase advance is given by

\[ \Phi(s) = \int \frac{1}{\beta(s)} ds \]

and inserting \( \beta(s) \) from Eq.(27), we get

\[ \Phi(s) = \frac{1}{\beta_0} \int_0^{\ell_1} \frac{1}{1 + s^2 / \beta_0^2} ds \]

\[ \Phi(s) = \arctan \frac{\ell_1}{\beta_0} \]

where \( \ell_1 \) denotes the distance of the first focusing element from the IP, i.e. the length of the first drift space. In Fig. 19 we have plotted the phase advance as a function of \( \ell \) for a \( \beta \) function of 10 cm. If the length of the drift is large compared to the value of \( \beta \) at the IP, which is usually the case, the phase advance approximates 90 degrees on each side. In other words, the tune of the accelerator will increase by half an integer within the complete drift space.

![Fig. 19: Phase advance in a symmetric drift space as a function of the drift length](image)

There are more remarks concerning mini-beta sections that we shall mention briefly. As we have seen, large values of the beta function on either side of the interaction point cannot be avoided if a mini-beta section is inserted in the machine lattice. These high-\( \beta \) values have a strong impact on the machine performance:

- According to Eq. (22), the chromaticity of a lattice is given by the strength of the focusing elements and the value of the \( \beta \) function at that position. In a mini-beta insertion, unfortunately, we have both strong quadrupoles and large beam dimensions.

\[ \xi = -\frac{1}{4\pi} \oint \{ k(s) \beta(s) \} ds \]

The contribution of this lattice section to \( \xi \) can be very large, and as it has to be corrected in the ring, it places strong limitations on the luminosity in a collider ring.

- During insertion the beam dimension can reach high values, the aperture of the doublet magnets has to be much larger than in the FODO structure of the arc. Large magnet
apertures limit the strength of the quadrupole. A compromise has to be found between aperture requirement, integrated focusing strength, spot size at the IP, and money, as large magnets are quite expensive.

Last but not least is the problem of field quality and adjustment. Compared to the standard magnets in the arc, the lenses in a mini-beta section have to fulfil stronger requirements. A kick, caused by a dipole error or due to an off-centre quadrupole lens, leads to an orbit distortion proportional to the beta function at the place of the error [Eq. (21)].

The field quality concerning higher multipole components has to be much higher and the adjustment of the mini-beta quadrupoles much more precise than those of quadrupole lenses in the arc. In general, multipole components of the order $\Delta B/B = 10^{-4}$ with respect to the main field and alignment tolerances in the transverse plane of about a tenth of a millimetre are desired.

3.2.2 Guidelines for the design of a mini-beta insertion

1. The periodic solution of the lattice cell in the arc has to be calculated to provide starting values for the insertion.
2. Introduce the drift space needed for the insertion device (e.g. the particle detector).
3. Put the mini-beta quadrupoles as close as possible to the IP—nowadays these lenses are often embedded in the detector to keep the distance $s$ small.
4. Introduce additional quadrupole lenses to match the optical parameters of the insertion to the solution of the arc cell. In general, functions $\alpha_x, \beta_x, \alpha_y, \beta_y$ and horizontal dispersion $D_x, D'_x$ have to be matched. Sometimes, additional quadrupoles are needed to adjust the tunes in both planes, in the case of HERA even the vertical dispersion $D_y, D'_y$ needs to be corrected. So, at least eight additional magnet lenses are necessary.

4 DISPERSION SUPPRESSORS

The dispersion function $D(s)$ has already been mentioned in Section 2.4 where we have shown that it is a function of the focusing properties of the lattice cell, and where we have calculated its size as a function of the cell parameters.

Returning to this topic in the context of lattice insertions: In the interaction region of an accelerator, the straight section of a ring where two counter-rotating beams collide (usually designed as a mini-beta insertion) the dispersion function $D(s)$ has to disappear. A non-vanishing dispersion dilutes the luminosity of the machine and leads to additional stop bands in the working diagram of the accelerator (synchro-betatron resonances), driven by the beam–beam interaction.

Therefore sections have to be inserted in our magnet lattice to reduce the function $D(s)$ to zero, these are known as dispersion suppressing schemes. In Eq. (18) we have shown that the oscillation amplitude of a particle is given by

$$x(s) = x_0(s) + D(s) \frac{\Delta p}{p_0},$$

where $x_0$ describes the solution of the homogeneous differential equation (valid for particles with ideal momentum $p_0$) and the second term — the dispersion term — describes the additional oscillation amplitude for particles with a relative momentum error $\Delta p/p_0$. 

63
For example, take some numbers from the HERA proton storage ring: The beam size at the collision point of the two beams is in the horizontal and vertical direction given by the mini-beta insertion: $\sigma_x \approx 118 \, \mu m$ and $\sigma_y \approx 32 \, \mu m$. The contribution of the dispersion function to the particles’ oscillation amplitude with a typical dispersion in the cell of $D(s) \approx 1.5 \, m$ and a momentum distribution of the beam $\Delta p/p \approx 5 \times 10^{-4}$ amounts to $x_{13} = 0.75 \, mm$.

4.1. Dispersion suppression using additional quadrupole magnets: ‘the straightforward way’

There are several ways to suppress the dispersion, each with advantages and disadvantages, which are not covered in this paper. However, the rationale of the basic idea is presented. Let us assume a periodic lattice is given and one simply wants to continue this FODO structure of the arc through the straight section — but with vanishing dispersion.

Given an optical solution in the arc cells as shown in Fig. 20, we have to guarantee that starting from the periodic solution of the optical parameters $\alpha(s), \beta(s),$ and $D(s)$ we obtain a situation at the end of the suppressor where we get $D(s) = D'(s) = 0$ and the values for $\alpha$ and $\beta$ unchanged.

The boundary conditions

$$D(s) = D'(s) = 0$$
$$\beta_x(s) = \beta_{x\,arc}, \quad \alpha_x(s) = \alpha_{x\,arc}$$
$$\beta_y(s) = \beta_{y\,arc}, \quad \alpha_y(s) = \alpha_{y\,arc}$$

can be fulfilled by introducing six additional quadrupole lenses whose strengths have to be matched individually in an adequate way. This can be done using one of the beam-optics codes available in every accelerator laboratory. An example is shown in Fig. 21, starting from a FODO structure with a phase advance of $\varphi \approx 61 \, ^\circ$ per cell.

The advantages of this scheme are

– it works for arbitrary phase advance of the arc structure;
– matching works also for different optical parameters $\alpha$ and $\beta$ before and after the dispersion suppressor;
– the ring geometry is unchanged as no additional dipoles are needed.
On the other hand there are a number of disadvantages:

- As the strength of the additional quadrupole magnets have to be matched individually, the scheme needs additional power supplies and quadrupole magnet types which can be an expensive requirement.
- The required quadrupole fields are in general stronger than those in the arc.
- The $\beta$ function reaches higher values (sometimes really high values) and so the aperture of The vacuum chamber and of the magnets has to be increased.

There are alternative ways to suppress the dispersion which do not need individually powered quadrupole lenses but instead change the strength of the dipole magnets at the end of the arc structure.

4.2 ‘The clever way’: half-bend schemes

This dispersion-suppressing scheme exits of $n$ additional FODO cells that are added to the periodic arc structure but where the bending strength of the dipole magnets is reduced. As before, we split the lattice into three parts: the periodic structure of the FODO cells in the arc, the lattice insertion where the dispersion is suppressed, followed by a dispersion-free section which can be another FODO structure without bending magnets or a mini-beta insertion etc.

The calculation of the suppressor requires several steps.

4.2.1 Establish the matrix for a periodic arc cell

We have already calculated the dispersion in a FODO lattice in Eq. (20). In thin-lens approximation we have derived a formula for $D$ as a function of the focusing properties of the lattice. Now we have to be a little bit more accurate and instead of the focusing strength and phase advance we have to work with the optical parameters of the system. We know that the transfer matrix in a lattice of a storage ring can be written as a function of the optical parameters of Eq. (14)
The variable \( \Phi \) refers to the phase advance between the starting point ‘0’ and the end point ‘s’ of the transformation. The formula is valid for any starting and end point in the lattice. If, for convenience, we refer the transformation to the middle of a focusing quadrupole magnet (as we usually did in the past) where \( \alpha = 0 \), and if we are interested in the solution for a complete cell, we can write the equation in a simpler form. Extending the matrix to the \( 3 \times 3 \) form to include the dispersion terms [see Section (2.4)] and taking into account the periodicity of the system, \( \beta_0 = \beta_s \), we get

\[
M_{\text{cell}} = \begin{pmatrix}
C & S & D \\
C' & S' & D' \\
0 & 0 & 1
\end{pmatrix} = \begin{pmatrix}
\cos \phi_c & \beta_c \sin \phi_c & D(l) \\
-\frac{1}{\beta_c} \sin \phi_c & \cos \phi_c & D'(l) \\
0 & 0 & 1
\end{pmatrix}.
\]  

(30)

Now \( \Phi_c \) is the phase advance for a single cell and the index ‘c’ reminds us that we talk about the periodic solution (one complete cell).

The dispersion elements \( D \) and \( D' \) are as usual given by the \( C \) and \( S \) elements according to Eq. (19)

\[
D(\ell) = S(\ell)^* \int_{0}^{l} \frac{1}{\rho(\bar{s})} C(\bar{s}) d\bar{s} - C(\ell)^* \int_{0}^{l} \frac{1}{\rho(\bar{s})} S(\bar{s}) d\bar{s}
\]

\[
D'(\ell) = S'(\ell)^* \int_{0}^{l} \frac{1}{\rho(\bar{s})} C(\bar{s}) d\bar{s} - C'(\ell)^* \int_{0}^{l} \frac{1}{\rho(\bar{s})} S(\bar{s}) d\bar{s}.
\]

The values \( C(\ell) \) and \( S(\ell) \) refer to the symmetry point of the cell (the middle of the quadrupole). The integral, however, has to be taken over the dipole magnet, where \( \rho \neq 0 \). Assuming a constant bending radius in the dipole magnets of the arc, \( \rho = \text{const} \) (which is a good approximation in general), we can solve the integral over \( C(s) \) and \( S(s) \) if we approximate their values by those in the middle of the dipole magnet.

---

**Fig. 22:** Schematic view of a FODO: notation of the phase relations in the cell.
4.2.2 Transformation of the optical functions from the centre of the quadrupole to the middle of the dipole, to calculate the $C(\tilde{s})$ and $S(\tilde{s})$ functions

As indicated in Fig. 22 we have to transform the optical functions $\alpha$ and $\beta$ from the centre of the quadrupole lens to the centre of the dipole magnet. The formalism is given by Eq (29) and we get (with $\alpha_0 = 0$)

$$C_m = \frac{\beta_m}{\beta_C} \cos \Delta \phi = \sqrt{\frac{\beta_m}{\beta_C}} \cos \left(\frac{\phi_c}{2} \pm \varphi_m\right)$$

$$S_m = \sqrt{\frac{\beta_m}{\beta_C}} \sin \left(\frac{\phi_c}{2} \pm \varphi_m\right).$$

The index $m$ tells us that we are now dealing with values in the middle of the bending magnets and as our starting point was the centre of the QF quadrupole, the phase advance for this transformation is half the phase advance of the cell, which brings us to the QD lens, plus/minus the phase distance $\varphi_m$ from that point to the dipole centre.

Now we can solve the integrals for $D(s)$ and $D'(s)$

$$D(\ell) = S(\ell) \ast \int_0^\ell \frac{1}{\rho(\tilde{s})} C(\tilde{s}) d\tilde{s} - C(\ell) \ast \int_0^\ell \frac{1}{\rho(\tilde{s})} S(\tilde{s}) d\tilde{s}$$

$$D(\ell) = \beta_C \sin \phi_c \ast \frac{L_B}{\rho} \sqrt{\frac{\beta_m}{\beta_C}} \cos \left(\frac{\phi_c}{2} \pm \varphi_m\right) - \cos \phi_c \ast \frac{L_B}{\rho} \sqrt{\beta_m \beta_C} \sin \left(\frac{\phi_c}{2} \pm \varphi_m\right).$$

(31)

With $L_B$ the length of the dipole magnets and putting $\delta = L_B/\rho$ for the bending angle we get

$$D(\ell) = \delta \sqrt{\beta_m \beta_C} \sin \phi_c \left[ \cos \left(\frac{\phi_c}{2} + \varphi_m\right) + \cos \left(\frac{\phi_c}{2} - \varphi_m\right) \right] -$$

$$- \cos \phi_c \left[ \sin \left(\frac{\phi_c}{2} + \varphi_m\right) + \sin \left(\frac{\phi_c}{2} - \varphi_m\right) \right].$$

Using the trigonometric relations

$$\cos x + \cos y = 2 \cos \frac{x + y}{2} \cos \frac{x - y}{2}$$

$$\sin x + \sin y = 2 \sin \frac{x + y}{2} \cos \frac{x - y}{2}$$

we get

$$D(\ell) = \delta \sqrt{\beta_m \beta_C} \left[ \sin \phi_c \ast \cos \left(\frac{\phi_c}{2} \pm \varphi_m\right) - \cos \phi_c \ast \cos \left(\frac{\phi_c}{2} \pm \varphi_m\right) \right]$$

$$D(\ell) = 2 \delta \sqrt{\beta_m \beta_C} \ast \cos \varphi_m \left[ \sin \phi_c \ast \cos \left(\frac{\phi_c}{2} \pm \varphi_m\right) - \cos \phi_c \ast \sin \left(\frac{\phi_c}{2} \pm \varphi_m\right) \right]$$

and with
we can derive the dispersion at the middle of the quadrupole magnet in its final form

\[
D(\ell) = 2\delta \sqrt{\beta_m/\beta_C} \cos \phi_m \left( 2 \sin \frac{\phi_C}{2} \cos^2 \frac{\phi_C}{2} - \left( \cos^2 \frac{\phi_C}{2} - \sin^2 \frac{\phi_C}{2} \right) \sin \phi_C \right)
\]

\[
D(\ell) = 2\delta \sqrt{\beta_m/\beta_C} \cos \phi_m \left( \frac{2 \cos^2 \phi_C}{2} - \cos^2 \frac{\phi_C}{2} + \sin^2 \frac{\phi_C}{2} \right)
\]

\[
D(\ell) = 2\delta \sqrt{\beta_m/\beta_C} \cos \phi_m \sin \frac{\phi_C}{2} .
\] (32)

This is the expression for the dispersion term of matrix (30) at the centre of the quadrupole magnet, determined from the dipole strength \(1/\rho\) and matrix elements \(C\) and \(S\) at the position of the dipole.

In full analogy, one derives the formula for the derivative of the dispersion, \(D'(s)\)

\[
D'(\ell) = 2\delta \sqrt{\beta_m/\beta_C} \cos \phi_m \cos \frac{\phi_C}{2} .
\] (33)

As this refers to the situation in the middle of a quadrupole, the expressions for \(D(s)\) and \(D'(s)\) are valid for a periodic structure, namely one FODO cell. Therefore, we require periodic boundary conditions for the transformation from one cell to the next:

\[
\begin{pmatrix}
D_C \\
D'_C
\end{pmatrix} = M_C \begin{pmatrix}
D_C \\
D'_C
\end{pmatrix}
\]

and by symmetry

\[
D'_C = 0 .
\] (34)

With these boundary conditions the periodic dispersion in the FODO cell is determined.

\[
D_C = D_C \cos \phi_C + \delta \sqrt{\beta_m/\beta_C} \cos \phi_m \sin \phi_C \frac{\phi_C}{2}
\]

\[
D_C = \delta \sqrt{\beta_m/\beta_C} \cos \phi_m \sin \frac{\phi_C}{2} .
\] (35)

4.2.3 Calculate the dispersion in the suppressor part

In the dispersion suppressor section, \(D(s)\) starting with the value at the end of the cell is reduced to zero. Or turning it around and thinking from right to left: The dispersion has to be created, starting from \(D = D' = 0\). The goal is to generate the dispersion in this section so that the values of the periodic arc cell are obtained.
The relation for $D(s)$ still holds

$$D(\ell) = S(\ell) \int_0^\ell \frac{1}{\rho(\tilde{s})} C(\tilde{s})d\tilde{s} - C(\ell) \int_0^\ell \frac{1}{\rho(\tilde{s})} S(\tilde{s})d\tilde{s}$$

but now we can take several cells into account, (the number of cells inside the suppressor scheme) and we will have the freedom to choose a dipole strength $\rho_{\text{supr}}$ in this section, which differs from the strength of the arc dipoles. As the dispersion is generated in a number of $n$ cells the matrix for these $n$ cells is

$$M_n = M_n^* = \begin{pmatrix}
\cos n\phi_c & \beta_c \sin n\phi_c & D_n \\
\frac{-1}{\beta_c} \sin n\phi_c & \cos n\phi_c & D_n^* \\
0 & 0 & 1
\end{pmatrix}$$

and according to (31) the dispersion created in these $n$ cells is given by

$$D_n = \beta_c \sin n\phi_c \cdot \delta_{\text{supr}} \cdot \sum_{i=1}^{n} \cos \left( i\phi_c - \frac{1}{2} \phi_c \pm \phi_m \right) \cdot \sqrt{\beta_m \beta_c} - \\
- \cos n\phi_c \cdot \delta_{\text{supr}} \cdot \sum_{i=1}^{n} \sqrt{\beta_m \beta_c} \sin \left( i\phi_c - \frac{1}{2} \phi_c \pm \phi_m \right)$$

$$D_n = \sqrt{\beta_m \beta_c} \cdot \sin n\phi_c \cdot \delta_{\text{supr}} \cdot \sum_{i=1}^{n} \cos \left( \frac{2i-1}{2} \phi_c \pm \phi_m \right) - \\
\sqrt{\beta_m \beta_c} \cdot \delta_{\text{supr}} \cdot \cos n\phi_c \sum_{i=1}^{n} \sin \left( \frac{2i-1}{2} \phi_c + \phi_m \right).$$

Remembering the trigonometric gymnastics shown above we get

$$D_n = \delta_{\text{supr}} \cdot \sqrt{\beta_m \beta_c} \cdot \sin n\phi_c \cdot \sum_{i=1}^{n} \cos \left( \frac{2i-1}{2} \phi_c \right) \cdot 2 \cos \phi_m - \\
- \delta_{\text{supr}} \cdot \sqrt{\beta_m \beta_c} \cdot \cos n\phi_c \sum_{i=1}^{n} \sin \left( \frac{2i-1}{2} \phi_c \right) \cdot 2 \cos \phi_m$$

$$D_n = 2\delta_{\text{supr}} \cdot \sqrt{\beta_m \beta_c} \cdot \cos \phi_m \sum_{i=1}^{n} \cos \left( \frac{2i-1}{2} \phi_c \right) \cdot \sin(n\phi_c) - \\
- \sum_{i=1}^{n} \sin \left( \frac{2i-1}{2} \phi_c \right) \cdot \cos(n\phi_c) \right)$$
\[ D_n = 2\delta_{\text{supr}} \sqrt{\beta_m \beta_C} \cos \varphi_m \sin \left( \frac{n\phi_C}{2} \right) \frac{\sin \frac{n\phi_C}{2} \cos \frac{n\phi_C}{2}}{\sin \frac{\phi_C}{2}} - 
\]

\[ -2\delta_{\text{supr}} \sqrt{\beta_m \beta_C} \cos \varphi_m \sin \left( n\Phi_C \right) \frac{\sin \frac{n\phi_C}{2} \cos \frac{n\phi_C}{2}}{\sin \frac{\phi_C}{2}} \]

\[ D_n = \frac{2\delta_{\text{supr}} \sqrt{\beta_m \beta_C} \cos \varphi_m}{\sin \frac{\phi_C}{2}} \left\{ 2\sin \frac{n\phi_C}{2} \cos \frac{n\phi_C}{2} - \cos \frac{n\phi_C}{2} \sin \frac{n\phi_C}{2} \right\} . \]

And finally

\[ D_n = \frac{2\delta_{\text{supr}} \sqrt{\beta_m \beta_C} \cos \varphi_m}{\sin \frac{\phi_C}{2}} \sin^2 \frac{n\phi_C}{2} . \] (36)

This relation gives us the dispersion \( D(s) \) that is created in a number of \( n \) cells with a phase advance of \( \Phi_C \) per cell. \( \delta_{\text{supr}} \) is the bending strength of the dipole magnets located in these \( n \) cells and the optical function \( \beta_m \) and \( \beta_C \) refer to the values at the centre of the dipole and the quadrupole, respectively.

In a similar calculation we get the expression for the derivative \( D'(s) \) of the dispersion:

\[ D'_n = \frac{2\delta_{\text{supr}} \sqrt{\beta_m \beta_C} \cos \varphi_m}{\sin \frac{\phi_C}{2}} \sin n\phi_c . \] (37)

4.2.4 **Determine the strength of the suppressor dipoles**

The last step is to calculate the strength of the dipole magnets in the suppressor section. As for the optimum match of \( D \) the dispersion generated in this section has to be equal to that of the arc cells we equate the expressions (34),(35) and (36),(37) and get for \( D_n \) the condition

\[ D_n = \frac{2\delta_{\text{supr}} \sqrt{\beta_m \beta_C} \cos \varphi_m \sin \frac{n\phi_C}{2}}{\sin \frac{\phi_C}{2}} = \delta_{\text{arc}} \sqrt{\beta_m \beta_C} \cos \frac{\varphi_m}{\sin \frac{\phi_C}{2}} \]

and for \( D' \)
From these last equations we deduce two conditions for the dispersion matching

\[ D'_n = \frac{2\delta_{supr}^* \sqrt{\frac{\beta_m}{\beta_C}} \cos \varphi_m}{\sin \phi_C} * \sin n\phi_C = 0. \]

If the phase advance per cell in the arc fulfils the condition \( \sin(n\Phi_C) = 0 \), the strength of the dipoles in the suppressor region is just half the strength of the arc dipoles. In other words the phase has to fulfil the condition

\[ n\phi_C = k * \pi, \quad k = 1, 3, \ldots. \]

There are a number of possible phase advances that fulfil that relation, but clearly not every arbitrary phase is allowed. Possible constellations would be \( \Phi_C = 90^\circ, n = 2 \) cells or \( \Phi_C = 60^\circ, n = 3 \) cells in the suppressor.

Figure 23 shows such a half-bend dispersion suppressor, starting from a FODO structure with 60° phase advance per cell. The focusing strength of the FODO cells before and after the suppressor are identical, with the exception that — clearly — the FODO cells on the right are ‘empty’, i.e. they have no bending magnets.

Obviously, the beta function is now unchanged in the suppressor region, unlike the suppressor scheme with quadrupole lenses.
– special dipole magnets are needed (with half the strength of the arc types);
– the geometry of the ring is changed.

I want to mention, for purists, that in these equations the phase advance of the suppressor part is equal to the one of the arc structure — which is not completely true as the weak focusing term $1/\rho^2$ in the arc FODO differs from the term $1/(2\rho)^2$ in the half-bend scheme. As, however, the impact of the weak focusing on the beam optics can be neglected in many practical cases, Eq. (38) is nearly correct.

The application of this scheme is very elegant, but it has to be embedded in the accelerator design at an early stage as it has a strong impact on the beam optics and geometry.

### 4.3 The missing-bend dispersion suppressor scheme

Another suppressor scheme is used in a number of storage rings: It consists of a number of $n$ cells without dipole magnets at the end of the arc, followed by $m$ cells identical to the arc cells. The matching condition for this ‘missing-bend scheme’ with respect to the phase advance is

\[
\frac{2n+m}{2} \phi_C = (2k+1) \frac{\pi}{2}
\]

and for the number $m$ of the required cells.

\[
\sin \frac{m\phi_C}{2} = \frac{1}{2}, \quad k = 0, 2 \quad \text{or} \quad \sin \frac{m\phi_C}{2} = -\frac{1}{2}, \quad k = 1, 3 \ldots
\]

An example that is based on $\Phi = 60^\circ$ and $m = n = 1$ is shown in Fig. 24.

![Fig. 24: Dispersion suppressor based on the missing-magnet scheme](image)

There are more scenarios for a variety of phase relations in the arc and the corresponding bending strength needed to reduce $D(s)$, see Refs. [11,12].

In general, combine one of the two schemes (missing or half-bend suppressor) with a certain number of individual quadrupole lenses to guarantee the flexibility of the system with respect to phases changes in the lattice and to keep the size of the $\beta$ function moderate.
References

INTRODUCTION TO BEAM INSTRUMENTATION AND DIAGNOSTICS

R. Jones & H. Schmickler
CERN, Geneva, Switzerland

Abstract
These lectures aim at describing instruments and methods used for measuring beam parameters in circular accelerators. Emphasis will be given to new detection and analysis techniques in each field of accelerator instrumentation. The clear distinction is made between “instrumentation”, i.e. the design and construction of the instruments themselves and “diagnostics”, the use of the data from these instruments for running and improving the performance of the accelerator.

INTRODUCTION
Beam instrumentation and diagnostics combines the disciplines of accelerator physics with mechanical, electronic and software engineering, making it an extremely interesting field in which to work. The aim of the beam instrumentation physicist or engineer is to design, build, maintain and improve the diagnostic equipment for the observation of particle beams with the precision required to tune, operate and improve the accelerators and their associated transfer lines.

This introduction is intended to give an overview of the instrumentation in use in modern synchrotrons. The choice available today is so vast that inevitably it will not be possible to cover them all. Many of the standard instruments have been covered in previous CAS schools (see for example Ref. [1] which also contains a comprehensive list of references) and will therefore be touched upon only briefly, with more emphasis being given to new and innovative measurement techniques and their use in beam diagnostics. The following subjects will be discussed:

1. Beam position measurement
2. Beam current measurement
3. Diagnostics of transverse beam motion (tune, chromaticity and coupling)
4. Emittance measurement
5. Beam loss monitoring
6. Luminosity measurement
7. Some examples of beam diagnostics

1. BEAM POSITION MEASUREMENT
The Beam Position Monitor (BPM) can be found in every accelerator. Its role is to provide information on the position of the beam in the vacuum chamber at the monitor location. For linacs and transfer lines the BPMs are used to measure and correct beam trajectories, while for synchrotrons such monitors are distributed around the ring and used to calculate the closed orbit. In circular machines, their location is usually chosen close to the main quadrupole magnets where the β-functions are largest and so any orbit distortion a maximum. For 90° lattices a typical layout involves placing horizontal monitors near the focusing quadrupoles (where the horizontal β-function is large) and the vertical monitors near the defocusing quadrupoles (where the vertical β-function is large). Apart from closed orbit measurements, the BPMs are also used for trajectory measurements (the first turn trajectory is particularly important for closing the orbit on itself) and for accelerator physics experiments, where turn-by-turn data, and even bunch-to-bunch data is often required.
In the early days a BPM monitoring system simply consisted of an oscilloscope linked directly to the pick-up signals. Since then, enormous advances in the acquisition and processing electronics have been made, turning beam position monitors into very complex systems. Modern BPMs are capable of digitising individual bunches separated by a few nanoseconds, with a spatial resolution in the micron range, while the resulting orbit or trajectory collected from several hundred pick-ups can be displayed in a fraction of a second.

### 1.1 Pick-ups

The measurement of beam position relies on processing the information from pick-up electrodes located in the beam pipe. Five pick-up families are commonly employed:

- Electrostatic – including so-called ‘button’ and ‘shoe-box’ pick-ups
- Electromagnetic – stripline couplers
- Resonant cavity – especially suited for high frequency linacs
- Resistive
- Magnetic

An excellent in depth analysis of most of these pick-ups is presented in Ref. [2]. Here we will briefly describe the two most commonly used, namely the electrostatic and electromagnetic pick-up.

#### 1.1.2 Electrostatic (Capacitive)

The electrostatic or capacitive pick-up is the most widely used in circular accelerators. It consists of metallic electrodes situated on opposite sides of the vacuum chamber at the location where the beam is to be measured. As the beam passes through, electric charges are induced on the electrodes, with more induced on the side which is closer to the beam than the one furthest from the beam. By measuring the difference in the charge induced, the position can be calculated. Let us analyse the properties of button pick-ups (see Fig. 1) since they are the most popular due to their cheapness and ease of construction.

![Fig. 1 Cross-section and photo of an LHC button electrode.](image)

The image current associated with the beam will induce charges on the button which are proportional to the beam intensity and inversely proportional to the position of the beam from the electrode. A schematic representation is given in Fig. 2.

The figure of merit for any electrode is its transfer impedance (the ratio of the pick-up output voltage, \( V \), to the beam current, \( I_b \)). For a capacitive pick-up the signal is proportional to the rate of change of beam current at low frequencies, while for high frequencies the capacitance ‘integrates’ the signal and the transfer impedance tends to its maximum. For the case of a button electrode of area \( A \) and capacitance \( C \) situated at a distance \( d \) from the beam the maximum transfer impedance (i.e. the...
value it tends to at high frequency) can be approximated by:

\[ Z_{fe} = \frac{A}{2\pi d(\beta c) C} \]

Fig. 2 Schematic of a capacitively coupled electrode.

Button electrode capacitances are typically in the 10pF range. Impedance transformation can be used to improve the low frequency response at the expense of that at high frequency. Figure 3(a) shows the frequency response of an 8pF button electrode for the matched 50Ω impedance case (1:1) and after two different impedance transformations. The time response of the button for different bunch lengths can be seen in Fig. 3(b).

When designing such pick-ups care must be taken to limit the impedance variations when the transmission line used for signal extraction passes from the vacuum to a feedthrough or cable dielectric (such as ceramic, glass or air). Any such mismatch will produce unwanted reflections, often at high frequency, which could perturb the processing electronics. For this reason most processing chains introduce a low-pass filter on the button output. Special care must also be taken to pair the electrodes on opposite sides of the chamber to minimise offsets in the position reading. This pairing can be made less sensitive to capacitance variations if the high frequency cut-off for the processing electronics sits on the linear part of the button response, with the disadvantage that the overall signal amplitude is reduced.
1.1.3 Electromagnetic (stripline)

The electromagnetic pick-up is a transmission line (stripline) which couples to the transverse electromagnetic (TEM) field of the beam. The transmission line is formed between the stripline and the wall of the vacuum chamber and is excited by the beam only at the gaps on either end of the stripline where a longitudinal field occurs. Fig. 4 shows the layout of such an electromagnetic stripline electrode.

Consider a bunch travelling from left to right (upstream to downstream). While it is over the upstream port there is a voltage $V_r$ across $R_U$, causing a voltage wave of that amplitude to be launched to the right. The stripline forms a transmission line with the wall of the vacuum chamber of characteristic impedance $Z_0$. The voltage wave is therefore accompanied by a right travelling current wave of amplitude $I_r = V_r / Z_0$. This current flows along the bottom surface of the electrode whilst an equal and opposite current flows along the chamber wall. In addition an image current of amplitude $\eta I_B$ travels along the top surface of the electrode. The voltage $V_r$ across $R_U$ can therefore be expressed as

$$ V_r = (-I_r + \eta I_B) R_U = \eta I_B \frac{R_U Z_0}{R_U + Z_0} \Rightarrow V_r = \frac{1}{2} \eta I_B Z_0 \text{ for a matched stripline (}R_U = Z_0\text{)} $$

When the beam is over the downstream port it produces a voltage $-V_r = - \frac{1}{2} \eta I_B Z_0$ across $R_D$ in the same way as it produced a voltage $+V_r$ across $R_U$. This launches a left-travelling wave of the same magnitude, but different sign to the right-travelling wave, which propagates along the transmission line formed by the stripline and the chamber wall and will produce an inverted signal upon arrival at the upstream port a time $L/c$ later. The final signal observed at the upstream port will therefore be a bipolar pulse with the maxima separated by $2L/c$ (see Fig. 5(a)).

When the RF wavelength of the beam is equal to multiples of $2L$, the reflection and the signal from next bunch will cancel and there will be no net signal from the stripline. A maximum in the frequency response will be observed when $L$ is a quarter of an RF period, and hence the stripline pick-
up length is usually chosen accordingly. The full frequency response of a 60cm long stripline is shown in Fig. 5(b) and has a lobe structure, with the minima located at multiples of \( c/(2L) \).

For a relativistic beam the voltage due to the beam passing the downstream port is produced at the same time as the right-travelling wave propagating between the stripline and the wall arrives at the downstream port. The two equal and opposite voltages therefore cancel producing no net signal at the downstream port. The electromagnetic stripline pick-up is therefore said to be “directional”, i.e. a signal is only observed on the upstream port with respect to the beam direction. These pick-ups are therefore used in all locations where there are two counter rotating beams in the same vacuum chamber. Due to the imperfections in the stripline and feedthrough impedance matching, the best directivity one can hope to obtain for a real stripline is generally around 25-30dB (i.e. the voltage signal of one beam with respect to the other is attenuated by a factor between 18-32).

1.2 Beam Position Acquisition Systems

Once the signals from the opposite electrodes of a pick-up have been obtained, the next step is to convert these signals into a meaningful beam position. The first thing to do is to normalise the position, i.e. to make it independent of the signal amplitude (or beam intensity). This is generally done using one of three algorithms, whose response curves can be seen in Fig. 6.

![Fig. 6 Transfer functions for three commonly used position normalisation algorithms.](image-url)
• **Difference over sum (Δ/Σ)** - The sum and difference can be obtained either using a 0°/180° passive hybrid, a differential amplifier or calculated by software (after digitising), to give:

\[ \text{Normalised Position} = \frac{A - B}{A + B} \]

The transfer function of this algorithm can be seen to be highly linear.

• **Logarithmic ratio** - The two input signals are converted into their logarithmic counterparts and subtracted. In practice this is done using logarithmic amplifiers followed by a differential amplifier. This gives:

\[ \text{Normalised Position} = \log(A) - \log(B) \left( = \log \left( \frac{A}{B} \right) \right) \]

whose response curve is seen to be an reversed S-shape, which becomes highly non-linear when exceeding 70% of the normalised aperture.

• **Amplitude to Phase** - The two input signals are converted by a 90° passive hybrid into signals of equal amplitude but varying phase, with the position dependence of this phase given by:

\[ \text{Normalised Position} = \phi = 2 \times \text{ArcTan} (A/B) \]

Here the transfer function again deviates from the linear in an S form, but does not diverge for large excursions. In addition, the gradient is larger around zero, making it more sensitive towards the middle of the pick-up. A variation on the amplitude to phase algorithm is amplitude to time conversion, which will be discussed in more detail below.

The type of algorithm to be used will depend on the choice of processing electronics. In all cases the non-linearity is taken into account by calibration circuits and correction algorithms. A summary of commonly used beam position acquisition systems is given in Fig. 7.

![Fig. 7 Schematic representation of the various beam position processing families. (courtesy of G. Vismara, CERN)](image-url)
Here we will only briefly mention the various families in passing, but detailed descriptions along with the advantages and disadvantages of each system can be found in ref [3].

**MPX (multiplexed)** – each of the BPM electrodes is multiplexed in turn onto the same, single acquisition electronics chain. This eliminates channel to channel variations, but since the switching is generally quite slow such an acquisition tends to be used in circulating machines where only the average orbit is of importance.

**Hybrid (Sigma & Delta)** – here a 0°/180° passive hybrid is usually used to give the sum (Σ) and difference (Δ) signal from the two electrodes. The position (or ratio of the sum and difference signals) can then be obtained in many ways including: direct digitisation, homodyne detection (mixing the sum and delta signals with the sum signal itself) or heterodyne detection (mixing sum and delta signals with an external reference).

**Individual Treatment** – in this case each electrode is treated separately, but in parallel. The acquisition can either consist of directly digitising each signal or using logarithmic amplifiers as outlined above. The disadvantage of this method is that it requires two (or four depending on the pick-up orientation) very well matched chains of electronics, since the combination of the signals to obtain a position is performed at the very end of the chain.

**Passive Normalisation** – here the amplitude difference (i.e. position information) in the input signals is directly converted into a phase or time. Intensity information is lost in this procedure, but the result is a varying phase or time which is directly proportional to the position.

### 1.2.1 Wide-Band Time Normalisation

The LHC beam position system will be based on a new concept of amplitude to time normalisation, so-called “wide-band time normalisation” or WBTN (see ref [4]). This was developed with two main aims in sight; to provide bunch by bunch beam position information (requiring one measurement every 25ns for the LHC) and to avoid the necessity of gain switching (requiring that one chain of electronics cope with a factor 40 difference in single bunch intensity). The principle of the WBTN technique is outlined in Fig. 8.

![Fig. 8 The wide band time normalisation principle.](image)

The signal from a single pick-up electrode is split and recombined with a delayed signal from the other electrode. This results in two signals where the relative zero crossing time depends on the position of the beam. In the case of the LHC the position information is converted into two pulses separated by 10 ± 1.5ns (the position being encoded into the ±1.5ns). To obtain the required resolution and linearity using such a system requires precision high frequency electronic engineering. For example in order to achieve the 0.1% single shot resolution requested for the LHC (50µm on a 50mm diameter pick-up) a stability and reproducibility at the 3ps level is required. One advantage of this system is that once the position has been encoded into these two pulses, it can be transmitted in a digital manner over a fibre optic network. This is important for machines such as the LHC where the...
amount of electronics in the tunnel, in particular digital electronics, has to be minimised due to radiation levels. Fibre-optic transmission allows all the digital processing electronics, which finally extract the position from the time difference, to be regrouped in surface buildings, where they are accessible and free from radiation concerns. The final implementation layout for the LHC WBTN system is shown in Fig 9.

2. BEAM CURRENT AND INTENSITY MEASUREMENT

The measurement of beam current or bunch intensity is one of the most basic measurements performed at any accelerator. This is usually done by means of a “beam current transformer” or BCT. In order for the transformer to interact with the magnetic field of the beam it has to be placed over a ceramic gap in the vacuum chamber. To keep the impedance seen by the beam as low as possible an RF bypass (either a thin metallic coating or external capacitors on the ceramic) is required for the high frequency wall current components. In addition, to keep the vacuum chamber continuity, an Ohmic bypass external to the transformer is needed.

![Diagram of a beam current transformer](image)

**Fig. 10** The beam current transformer.
2.1 The Beam Current Transformer

The beam current, I_B, can be considered as the primary winding of the transformer, with the output voltage from the secondary windings given by V = L dI_B/dt. An ideal transformer would give a differentiated response, with the integrated charge being zero, which is not of much use as a measuring device. In reality the secondary windings have some stray capacitance, and are terminated by some finite resistance. This leads to signals of the form shown in Fig. 10. The transformer output now closely resembles the beam intensity distribution, with the added inconvenience of a DC offset due to the transformer droop. This DC offset can be corrected for either electronically or, in this modern era, by software treatment of directly digitised data.

A transformer with a bandwidth from 200Hz to 1GHz has recently been installed in the CERN-SPS [5]. Such a bandwidth is obtained by using a ferromagnetic core wound of high permeability metal tape to avoid eddy currents. With this instrument, operators can observe the bunch-by-bunch intensity evolution of beams destined for the LHC throughout the SPS acceleration cycle. In order to obtain the total charge in each bunch, fast integrators are required which are capable of working at repetition frequencies of up to 40MHz. Such an integrator chip has been developed by the Laboratoire de Physique Corpusculaire, Clermont-Ferrand, France, for capturing photomultiplier signals in the LHC-b experiment [6], and is now also used for bunch intensity measurements. A schematic of the integration principle and the resulting signals as measured in the CERN-SPS are shown in Fig. 11. The chip works using two integrators in parallel. As one integrates the other is discharged, with the output switched from one to the other on each clock cycle. The resulting integrated amplitude (voltage) is directly digitised, with all gain linearisation and DC offset subtraction performed by software.

![Fig. 11 The principle and measurement of the CERN-SPS fast beam current transformer.](image)

2.2 The DC Beam Current Transformer

In storage rings and accelerators with cycle times of several seconds, a DC beam transformer can be used to measure the total current. Such an instrument was developed for the CERN-ISR (Intersecting Storage Rings), the first machine to sustain beams for hours [7]. A DC transformer is based on a pair of matched, toroidal, ferromagnetic cores, which are driven into saturation by a modulation current at frequencies of up to a few kHz. The principle of operation is shown in Fig. 12, and makes use of the hysteresis loop of the toroid. If an equal but opposite modulation current (the triangular waveforms in Fig. 12) is applied to both cores with the beam not present, then the voltage induced in the detection windings on each core will also be equal but opposite. When, however, there is a beam current I_B present, the starting point in the hysteresis loop for zero modulation current is offset due to the static magnetic field generated by the beam current. Since the modulation is opposite in each toroid, the time spent in saturation will be different for the two branches of the hysteresis loop. This results in the
generation of voltage pulses at twice the modulation frequency when the induced voltage in the detection windings on each core is combined. The demodulation of this signal gives a train of pulses, with the width of each pulse being a direct measure of the beam current, i.e. by how much the hysteresis curves are offset.

In the “zero flux detector” implementation of the DC beam transformer, the result of the demodulation is fed back into a compensating current loop (see Fig. 12). Once the compensation current and the beam current are identical the net static magnetic field seen by the toroids is zero (hence zero flux) and the output from the demodulator is also zero. The beam current can then be obtained by simply measuring the voltage produced by this compensation current across a known resistor.

For modern DC transformers such a zero flux detector is used to compensate the droop of the simple beam current transformer described in section 6.2. This significantly increases the bandwidth of the system, allowing measurement from DC to a few MHz.

**3. DIAGNOSTICS OF TRANSVERSE BEAM MOTION**

The instrumentation used to look at transverse beam motion is very important to the efficient operation of any circular accelerator [8]. There are three main parameters which can be measured using such diagnostics, namely the betatron tune, chromaticity and coupling, all of which are discussed in detail below.
3.1 Tune Measurement

All betatron tune measurements are based on applying a transverse excitation to the beam and looking at the resulting beam response. The most common methods of performing such measurements are presented in this section.

3.1.1 Fourier Transform (FFT) of beam motion

In the simplest case the beam is given a single kick using a powerful stripline or magnetic kicker, and allowed to oscillate freely (alternatively white noise can be injected onto the beam). The observation of the resulting beam motion is usually carried out using one of the types of position pick-ups covered in Section 1. Once the data has been recorded, the power density spectrum in frequency domain can be computed using a Fast Fourier Transform (FFT). The betatron tune is determined as the frequency which has the highest amplitude response. If there is enough external excitation from other sources (ground motion, power supply ripple etc) or the beam is slightly unstable by itself, the method also gives useful information without any specific beam excitation.

What is usually of most interest for particle colliders is to be able to track the tune evolution during the whole of the accelerator cycle. The simplest way of achieving this is to repeat the tune measurement outlined above at regular intervals. By displaying such data as a spectrograph (Fig. 13(a)) the complete history of the tune during the machine cycle can be tracked.

3.1.2 Chirp Excitation

In order to minimise the frequency range over which power is put into the beam, swept frequency or “chirp excitation” is often used (so-called because if listened to at audio frequencies such a signal sounds like the chirp of a bird). The chirp range is set around the expected betatron tunes and the sweep time is determined depending on the requested time resolution and precision of the tune measurement. The advantage of this technique is that in addition to an amplitude response it also gives phase information, as the phase difference between the observed motion and the applied sine wave is easily measured. This makes it more sensitive than the single kick method and so allows smaller excitation amplitudes to be used. Fig. 13(b) shows a result from the CERN-SPS where a chirp was used.
performed every 30ms. The sine wave can be seen to sweep from low to high frequency, with the main
tune peak and the synchrotron satellites clearly visible.

3.1.3 Swept Frequency Analysis
For this method (often called “Network Analysis”) the beams are excited with a steady sinusoidal
wave. The amplitude and phase of the resulting oscillation are precisely determined by means of
harmonic analysis. Thereafter the excitation frequency is increased in steps until the range of interest is
covered. This represents a very precise measurement yielding the full information of the beam transfer
function. The disadvantage is the long measurement time, which renders the method of little use for
the study of dynamic phenomena.

![Fig. 14 Complete beam transfer function measured using swept frequency analysis in the CERN-LEP.](image)

Such a complete beam transfer function is shown in Fig. 14. Notice how the phase jumps by
180° as the excitation sweeps across the betatron tune frequency. Such a response is typical of any
harmonic oscillator. Since the rate of change of the phase change is a maximum at the peak of the
amplitude response, measurements performed using the phase response of the beam are in general
more sensitive that those relying on amplitude response.

3.1.4 Phase Locked Loop Tune Tracking
In order to have a fully continuous measurement of the tune a Phase Locked Loop (PLL) needs to be
implemented. The basic principle of the PLL is sketched in Fig. 15. A voltage or numerically
controlled oscillator (VCO or NCO) is used to put a sine wave excitation, $A \sin(\omega t)$, on the beam. The
beam response to this signal is then observed using a pick-up, and will be of the form $B \sin(\omega t + \phi)$,
where $\phi$ is the phase difference between the excitation and the observed signal. In the phase detector
the excitation signal and the observed signal are multiplied together, resulting in a signal of the form
$A B \sin(2\omega t + \phi) \cos(\phi)$, which is seen to have a DC component proportional to the cosine of the phase
difference. This will therefore be zero when the phase difference is 90° which, as was seen above, is
where the amplitude response is a maximum, i.e. at the tune frequency. The aim of the PLL is to “lock-
in” to this 90° phase difference between excitation and observed signal by correcting the VCO
frequency until the DC component of the phase detector output is zero. Since the PLL will always try
to maintain this 90° phase difference, the VCO frequency will track any tune changes, so giving a
continuous tune measurement.

In practice things are not quite as simple. Many parameters have to be optimised in order to
for the PLL to find, lock-in and subsequently track the tune peak. The beam spectra and dynamics also
have to be well understood if the PLL is not to lock or jump to a spurious line, resonance, synchrotron
sideband etc. In addition, for hadron machines, the continuous excitation will lead to emittance blow-
up. In order for this to be kept to a minimum the applied excitation has to be small and therefore the
observation pick-up and following electronics very sensitive. This is less of a problem for lepton colliders where radiation damping takes care of any emittance blow-up caused by the excitation, making PLL systems much easier to implement on such machines.

3.2 Chromaticity Measurement

For any high energy synchrotron, the control of chromaticity is very important. If the chromaticity is of the wrong sign (corresponding to positive below the transition energy or negative above it) then the beam quickly becomes unstable due to the head-tail instability. If the chromaticity is too big then the tune spread becomes large and some particles are inevitably lost as they hit resonance lines in tune space. The most common method of measuring the chromaticity of a circular machine is to measure the betatron tune as a function of the beam energy and then to calculate the chromaticity from the resulting gradient. This is usually done by varying the RF frequency, keeping the magnetic field static. The equations of interest are:

$$\Delta Q = (\xi Q) \frac{\Delta p}{p} = Q' \frac{\Delta p}{p} = Q' \frac{\Delta R}{R} = Q \left( \frac{-\gamma_i^2 \gamma^2}{\gamma^2 - \gamma_i^2} \right) \frac{\Delta f}{f} \quad (3.1)$$

where $\Delta Q$ is the change in tune, $\Delta p/p$ the momentum spread (or relative change in momentum), $\Delta R/R$ the relative change in radius, $\Delta f/f$ the relative change in RF frequency and $\xi$ the chromaticity. Please note that the chromaticity, $\xi$, is often expressed as $Q' = Q \xi$, where $Q$ is the total betatron tune including the integer part.

In the CERN-SPS, for example, a chromaticity measurement consists of performing a tune measurement for three different RF frequency settings. Instead of noting the exact RF frequency, what is actually measured is the change in closed orbit, from which the relative change in radius can be calculated. These three points are then plotted, with the gradient giving the chromaticity.

In order to obtain continuous chromaticity measurements this technique of RF modulation is combined with the PLL tune measurement outlined in the previous section. The RF frequency is usually programmed with a small asymmetric function which periodically varies about the mean RF frequency. By tracking the tune during this time using the PLL and knowing the magnitude of the RF change, the chromaticity can be calculated and tracked. An example of such a measurement performed at the CERN-LEP is shown in Fig. 16.

Fig. 15 Principle of a phase locked loop tune tracker.
3.2.1 Head-Tail Chromaticity Measurement

The methods outlined above do not allow instantaneous chromaticity measurements, for instance during energy ramping or beta squeezing and are limited to repetition intervals in the Hz range. In preparation for the LHC a new approach has been developed which uses the energy spread in the beams for a chromaticity measurement. Transverse oscillations are excited with a single kick and the chromaticity is calculated from the phase difference of the individually sampled head and tail motions of a single bunch. Using this method the chromaticity can be calculated using the data from only one synchrotron period (about 15-50 milliseconds in the case of the LHC). In addition, this technique does not rely on an accurate knowledge of the fractional part of the betatron tune and, for a machine operating well above transition, the calculated chromaticity is virtually independent of beam energy.

Assuming longitudinal stability, a single particle will rotate in longitudinal phase-space at a frequency equal to the synchrotron frequency. During this longitudinal motion the particle also undergoes transverse motion. If the chromaticity is zero, then the particle will have the same tune wherever it is in the bucket. As soon as chromaticity is non-zero, however, the particle’s tune will change depending on where it happens to be longitudinally. If a whole bunch of particles is kicked transversely, then the resulting transverse oscillations for a given longitudinal position within the bunch can be shown [9] to be given by

\[ y(n) = A \cos \left( 2\pi n Q_0 + \omega_c \hat{\tau} \left( \cos(2\pi n Q_s) - 1 \right) \right) \]  

(3.2)

where \( n \) is the number of turns since the kick, \( Q_0 \) is the betatron tune, \( Q_s \) is the synchrotron tune, \( \hat{\tau} \) is the longitudinal position with respect to the centre of the bunch and \( \omega_c \) is the so-called chromatic frequency given by

\[ \omega_c = Q' \omega_0 \frac{1}{\eta} \]  

(3.3)

Here \( Q' \) is the chromaticity, \( \omega_0 \) is the revolution frequency and \( \eta = \frac{1}{\gamma^2} - \frac{1}{\gamma_p^2} \). If we now consider the evolution of two longitudinal positions within a single bunch separated in time by \( \Delta \tau \), then from (3.1) it follows that the phase difference in the transverse oscillation of these two positions is given by
\[ \Delta \psi(n) = -\omega_z \Delta \tau \left( \cos(2\pi n Q_s) - 1 \right) \] (3.4)

This phase difference is a maximum when \( nQ_s = \frac{1}{2} \), i.e. after half a synchrotron period, giving

\[ \Delta \psi_{\text{max}} = -2 \omega_z \Delta \tau \] (3.5)

The chromaticity can therefore be written as

\[ Q' = \frac{-\eta \Delta \psi(n)}{\omega_0 \Delta \tau \left( \cos(2\pi n Q_s) - 1 \right)} = \frac{\eta \Delta \psi_{\text{max}}}{2 \omega_0 \Delta \tau} \] (3.6)

A schematic layout of the CERN-SPS Head-Tail monitor [10] set-up is shown in Fig. 17(a). A straight stripline coupler (see section 1.1.3) followed by a 180° hybrid is used to provide the sum and difference signals for a given measurement plane. These signals are fed into a fast-sampling (2GS/s on each channel), high bandwidth (2GHz) digital oscilloscope. A VME front-end acquisition crate then retrieves the data via a GPIB link and provides the bunch synchronous timing. Using the “Fast-Frame” capabilities of the oscilloscope the data from the same bunch can be captured over several hundred turns. Fig. 17(b) shows the result of such a head-tail chromaticity measurement. The top two plots show the transverse movement of the head and tail respectively of a single bunch after the beam is kicked. The lower left plot shows the evolution of the phase of the head and tail and the phase difference. It can be seen that the signals are re-phased after one synchrotron period, with the phase difference a maximum after \( \frac{1}{2} \) a synchrotron period. The final plot (lower right) shows the calculated chromaticity (using equation 3.6) for all turns where the phase difference is well defined.

![Schematic layout of the CERN-SPS Head-Tail monitor](a)

![Head-tail monitor acquisition](b)

Fig. 17(a) Layout of the CERN-SPS head-tail monitor. (b) Head-tail monitor acquisition.

### 3.3 Coupling Measurement

The control of coupling (the degree to which horizontal and vertical betatron motion is linked) is also important for circular accelerators. Excessive coupling will make tune and chromaticity measurements almost impossible, as the information from both planes are mixed-up in the observed signal. A very good and comprehensive summary of linear betatron coupling can be found in [11].

#### 3.3.1 Closest Tune Approach

For this method, both betatron tunes are measured during a linear quadrupole power converter ramp which crosses the values of the horizontal and vertical tunes. The remaining separation of the tune
traces is a direct measure for the total coupling coefficient $|c|$. A measurement example from the CERN-LEP, using a phase locked loop tune measurement is shown in Fig. 18. In order to ensure that the PLL keeps tracking both tunes, even when they approach each other, the measurements are performed on two different bunches.

3.3.2 Kick Method
The above method does not allow for diagnostics during machine transitions. A better tool for the measurement of small coupling coefficients, although demanding quite large beam excitations, consists of applying a single kick in one plane and observing the time evolution of the betatron oscillations in both planes. This method is described in Ref [11].

4. EMITTANCE MEASUREMENT
The ultimate luminosity of any collider is inversely proportional to the transverse emittance of the colliding beams. Preservation of emittance and hence emittance measurements are of particular importance in the long chain of accelerators and storage rings of big hadron colliders as the emittance of a hadron bunch is not appreciably reduced through mechanisms such as the radiation damping associated with lepton machines. Good explanations of emittance can be found in Refs [12, 13].

The emittance which includes about 98% of the beam-particles can be defined as

$$\varepsilon(98\%) = \frac{\text{beamwidth}^2 - \left(\frac{\Delta P}{P} \cdot D_m\right)^2}{\beta_m} = \frac{\text{FWHM}^2 - \left(\frac{\Delta P}{P} \cdot D_m\right)^2}{\beta_m} \quad (4.1)$$

where FWHM is the measured full width at half height ($2.35\sigma$) of the beam, $\Delta P/P$ the FWHM of the momentum spread, $D_m$ the value of the dispersion-function and $\beta_m$ the value of the beta-function at the monitor position.

From this equation one can immediately see that the measurement of emittance depends on many parameters. This limits the accuracy to which emittance can be calculated, which is generally with a precision no better than around 10%. A number of instruments are capable of measuring the beam profile quite precisely, but in calculating the emittance one also relies on knowledge of the beam optical parameters at the place of the instrument and these are often fraught with considerable uncertainties.
4.1 Scintillator and Optical Transition Radiation Screens

Scintillator screens have been used for nearly a century and are the simplest and most convincing device when one has to thread a beam through a transfer line and into and around an accelerator. The modern version consists of a doped alumina screen which is inserted into the beam and can stand high intensities and large amounts of integrated charge. In its simplest form a graticuled screen is observed using a TV-camera. It can deliver a wealth of information to the eye of an experienced observer, but only in a semi-quantitative way. Much can be done about that with modern means of rapid image treatment, but questions concerning the linearity of these screens at high beam densities remain.

Optical Transition Radiation (OTR) screens are a cheap substitute for scintillator screens. OTR radiation is generated when a charged-particle beam transits the interface of two media with different dielectric constants (e.g. vacuum to metal or vice versa) [14]. Since this is a surface phenomenon, the screens can be made of very thin foils which reduces beam scattering and minimises heat deposition. The radiation produced is emitted in two cones around the angle of reflection for backward (vacuum to metal) OTR so that if the foil is placed at 45° to the beam direction, the radiation produced is at 90° to the beam direction. In addition two cones of forward OTR (metal to vacuum) are produced around the beam direction (see Fig. 19). The angular distribution of the emitted radiation has a central hole and a peak located at $1/\gamma$. The higher the value of $\gamma$ the sharper the peaks and the more light can be collected, which is why OTR is generally suited to lepton or high energy hadron machines.

![Fig. 19 (a) Backward and forward OTR patterns with their imaging schemes.](image1)

(b) Example of 2D OTR images taken every four turns at injection in the CERN-SPS.

4.2 SEM-Grids

Secondary Emission (SEM) Grids, also known as harps, consist of ribbons or wires which are placed in the beam. As the beam intercepts the grid, secondary emission occurs leading to a current in each strip which is proportional to the beam intensity at that location. By measuring this current for all strips a beam profile is obtained. SEM-grids are the most widely used means to measure the density profile of beams in transfer lines. In addition, sets of three, properly spaced (i.e. with the right phase advance between monitors), allow a determination of the emittance ellipse. What makes them popular is their simple and robust construction, the fact that there is little doubt about the measured distribution, and their high sensitivity, in particular at low energies and for ions. At higher energies they can be considered semi-transparent. Amongst their drawbacks are the limited spatial resolution (difficult to get the wire spacing much below 0.25mm) and the rather high cost for the mechanisms and electronics.
4.3 Wire Scanners

Of all the instruments used for measuring the emittance of circulating beams, wire-scanners are considered to be the most trustworthy. They come in two different types; rotative and linear. Rotative wire scanners consist of a thin wire (some tens of microns in diameter) mounted on a fork which is attached to a rotating motor (see Fig. 20), while linear scanners use motors which push/pull the wire across the beam. There are two ways of obtaining a beam profile with wire scanners; by measuring the secondary emission current as a function of wire position (similar to the SEM-grid acquisition mentioned above) or by measuring the flux of secondary particles created as the beam interacts with the wire. This latter technique is often used for high intensities, where the heating of the wire produces thermal emission which falsifies the secondary emission results. It relies on the use of radiation detectors, typically scintillators followed by photo-multipliers, placed downstream of the wire scanner to detect the $\gamma$-radiation and secondary particles produced when the wire intercepts the beam. To make the flux collected independent of the wire position may require the summation of the signals from two or more detectors positioned around the beam chamber.

Fast wire scanners are nearly non-destructive over a wide range of energies. Their spatial resolution can reach the micrometer range and, with fast gated electronics, the profiles of individual bunches can be observed. Their great sensitivity also allows them to be used for the study of beam halos.

Fig. 20 Rotative wire scanner and an example of a wire scanner profile measurement.

4.4 Residual Gas and Luminescence Monitors

Rest gas monitors are used in many high energy accelerators in order to reconstruct transverse beam distributions (see e.g. Ref. [15]). The signal results from the collection of either the ions or the electrons produced by the beam ionising the small amount of residual gas in the vacuum chamber. These ions or electrons are accelerated using a bias voltage of several kilovolts and collected on a micro channel plate (MCP). The avalanche of electrons produced by the MCP then hits a phosphor screen, giving an image of the beam profile which can be monitored using a CCD camera (see Fig. 21). Due to their rigidity, ions are less sensitive to the distorting effects of the space charge from the circulating beam, but their slow drift time, even with high bias voltages, means that they spend a long time in this beam field, making it difficult to analyse rms beam dimensions smaller than one millimetre. In order to use electrons to produce an image, a transverse magnetic field needs to be added around which the electrons spiral on their way to the MCP. This eliminates, to a large extent, the space charge effects of the beam and allows sharper images to be produced than with ions. This additional magnetic field, however, is also seen by the beam and has to be compensated by two corrector magnets either side of the ionisation profile monitor.
Luminescence monitors (see e.g. Ref. [16]) also rely on the interaction of the beam with a gas in the vacuum chamber. In this case the gas of interest is nitrogen, in which electrons are excited from the ground state to a higher energy level by the passing beam. Once the beam has passed the electrons return to the ground state and emit photons. In the case of nitrogen the dominant photon wavelength is 391.3nm, corresponding to light at the lower end of the visible range, for which many detectors are available. In general, the residual gas alone does not produce enough photons for accurate imaging and hence a local pressure bump is usually created by injecting a small amount of nitrogen to enhance the photon production. The principle of luminescence monitoring and a schematic layout of such an instrument are shown in Fig. 22. Also shown in Fig. 22 is an example of a continuous measurement performed at the CERN-SPS, showing the ability of such an instrument to track the evolution of the beam size through the various acceleration stages with little effect on the beam.

Most users consider both the residual gas ionisation and luminescence profile monitors to be semi-quantitative and not be relied upon for absolute emittance measurements, even after calibration.
against some other instrument such as a wire scanner. Their virtual transparency for the beam, however, makes them useful for the continuous on-line tracking of beam size.

4.5 Synchrotron Radiation Monitors

Synchrotron radiation monitors are limited to highly relativistic particles and offer a completely non-destructive and continuous measurement of the 2-dimensional density distribution of the beam. These monitors make use of the light produced when highly relativistic particles are deflected by a magnetic field. They are therefore usually positioned to make use of parasitic light produced by a dipole magnet in the machine or behind a purpose built “wiggler” magnet in which the beam is deflected several times to enhance the photon emission.

Fig. 23 The CERN-LEP BEXE detector based on cadmium telluride photo-conductors.

Fig. 24 Tracking vertical beam sizes with the BEXE detector during electron-position collisions.

The most common way of measuring the beam size with synchrotron radiation is to directly image the extracted light using traditional optics and a camera. The spatial resolution for such systems is usually limited by diffraction and depth-of-field effects. If the beam is sufficiently relativistic then the photon emission extends into the hard X-ray region of the spectrum and X-ray detectors can be used, for which diffraction effects can be completely disregarded. Such an instrument, based on
cadmium-telluride (CdTe) photo-conductors, was used at the CERN-LEP [17] to measure rms beam sizes down to 300 μm with a resolution of some 10 μm (see Fig. 23). The detector consisted of 64 voltage biased CdTe photo-conductors of 4 μm thickness and spaced by 100 μm on a ceramic substrate. Each photo-conductor was followed by its own individual charge amplifier. By reading out the signal from each cell the beam profile could be reconstructed. The interesting feature about such photo-conductors is that they allow real-time measurements with data acquisition rates up to 100 kHz. In addition they are extremely radiation resistant, accepting doses beyond 10^{12} Grays. These detectors were heavily used towards the end of CERN-LEP operation to optimise the luminosity by tracking the electron and position vertical beam size during collision (see Fig. 24).

5. BEAM LOSS MONITORING

Beam loss monitors (BLMs) have three main uses in particle accelerators:

- **Damage prevention** - Beam loss may result in damage to accelerator components or the experimental detectors. One task of any BLM system is to avoid such damage. In some accelerators it is an integral part of the protection system, signalling the beam abort system to fire if a certain loss rate is exceeded. This is of vital importance to the new generation of superconducting accelerators, for which even fairly small beam losses in the superconducting components can lead to magnet quenches.

- **Diagnostics** - Another task of BLM systems is to identify the position (and time) of unacceptable beam losses and to keep the radiation level in the accelerator and its surroundings as low as possible.

- **Luminosity optimisation** - BLMs can also help in the tuning of the machine in order to produce the long lifetimes necessary for improved luminosity.

The job of the BLM system is to establish the number of lost particles at a certain position within a specified time interval. Most BLM systems are mounted outside the vacuum chamber, so that the detector normally observes the shower caused by the lost particles interacting in the vacuum chamber walls or in the materials of the magnets. The number of detected particles and the signal from the BLM should be proportional to the number of lost particles. This proportionality depends on the position of the BLM with respect to the beam, the type of lost particles and the intervening material. It also, however, depends on the momentum of the lost particles, which may vary by a large amount during the acceleration cycle. One has to distinguish between two types of losses:

- **Fast losses** – where a large amount of beam is lost over a very few turns.
- **Slow losses** – where partial beam loss occurs over some time (circular machines) or distance (LINAC, transport lines). In storage-rings, the lifetime is defined by slow losses. There are many reasons for these losses and a BLM system is very helpful for finding out what is happening in the machine. In superconducting accelerators a BLM system can also prevent beam loss induced quenches caused by these slow losses.

The fact that BLM systems have to cover both of these cases means that they are required to function over a very large dynamic range, typically in the region of 10^4 to 10^6.

5.1 Long Ionisation Chambers

In 1963, Panowsky [18] proposed a BLM system for SLAC which consisted of one long (3.5 km) hollow coaxial cable filled with Ar (95%) + CO₂ (5%), mounted on the ceiling along the LINAC, about 2 m from the beam. When a beam loss occurs, an electrical signal is produced which propagates to both ends of the cable. Position sensitivity is achieved by comparing the time delay between the direct pulse from one end and the reflected pulse from the other. The time resolution is about 30 ns (~8 m) which, for shorter versions, can be reduced to about 5 ns. This principle of space resolution works for linear accelerators and transport lines with a bunch train much shorter than the machine and with relativistic particles. For particles travelling significantly slower than the signal in the cable the
resolution of multiple hits in the cable becomes difficult. In this case, and for circular machines, it is necessary to split the cable. Each segment has to be read out separately, with a spatial resolution which becomes approximately equal to their length.

5.2 Short Ionisation Chambers

Short ionisation chambers are used in many accelerators (see e.g. Ref. [19]). They are more or less equally spaced along the accelerator with additional units at special positions such as aperture restrictions, targets, collimators, etc. The chamber provides some medium with which the secondary particles created by the beam loss can interact, typically a gas such as nitrogen or argon. This interaction produces electron-ion pairs which are collected by a series of high voltage gaps along the length of the chamber. The resulting current is then measured and is proportional to the beam loss at the location of the monitor. An example of a CERN-SPS ionisation chamber is shown in Fig. 25.

![CERN-SPS ionisation chamber](image)

Fig. 25 A CERN-SPS ionisation chamber used for beam loss monitoring.

5.3 Scintillation Counters

In the case where losses occur in a machine without a full BLM system, a plastic scintillator with photomultiplier readout is often temporarily installed. Such systems have a well known behaviour, but the radiation damage of the plastic scintillator restricts their long term use. Liquid scintillators are not susceptible to such damage and have been installed in some accelerators [20, 21]. Such BLMs can be very fast, with pulse rise times of around 10ns, but suffer from drift in the photomultiplier gain.

5.4 Aluminum Cathode Electron Multipliers

In such detectors the sensitivity of photomultipliers to ionising radiation is increased by replacing the photocathode with an aluminium foil. This foil then works as a secondary electron emitter when irradiated. A BLM system consisting of Aluminum Cathode Electron Multipliers (ACEMs) is installed in the CERN-PS and PS-Booster [22]. It is very fast, with signal rise times in the order of 10ns, but is rather expensive since the ACEM is not a standard tube of photomultiplier manufacturers.

5.5 PIN Photodiodes

For circular electron accelerators which emit hard synchrotron radiation it is difficult to distinguish between the beam loss distributions and the synchrotron radiation background using traditional BLM techniques. In DESY-HERA, an electron-proton collider, the warm electron and a superconducting proton rings are in the same tunnel. Protection of the superconducting proton beam magnets from beam loss induced quenches must therefore rely on a BLM system which sees only the proton beam losses and not the synchrotron radiation background. In this case back to back PIN photodiodes are used to distinguish between the hadronic shower created by beam losses and the synchrotron radiation [23]. The charged particles will interact with both photodiodes, giving a coincidence signal, while the photons will be absorbed by the first diode. In contrast to the charge detection of most other BLM systems, PIN photodiode detection depends on counting coincidences, with the count rate proportional to the loss rate so long as the number of overlapping coincidences is small.
6. LUMINOSITY MONITORING

Luminosity Monitors are specific to colliders, since they measure the collision rate of the two counter-rotating particle beams. The following formulae define luminosity and related quantities:

Luminosity:
\[ L = f_{\text{rev}} \frac{MN^2}{4\pi\sigma^*} \]

Normalized emittance:
\[ \varepsilon_N = \frac{\gamma \sigma^2}{\beta^*} \]

Beam-beam tune shift:
\[ \Delta \nu_{bb} = \frac{N_{r_p}}{4\pi \varepsilon_N} \]

where \( f_{\text{rev}} \) is the revolution frequency, \( M \) the number of bunches, \( N \) the number of particles per bunch, \( \sigma^* \) the rms beam size at the collision point, \( \beta^* \) the beta function at the collision point and \( r_p \) the particle radius.

Since the counting rates of the experiments are directly proportional to the luminosity, the aim of the accelerator operators is to maximise the luminosity. This can be done by having a large number of particles per bunch, many bunches and small beams sizes at the interaction point.

In this section, luminosity monitoring will be taken as an example of beam instrumentation engineering, i.e. the whole process from selecting an appropriate physics process to system design. The system presently under development for the LHC has been chosen for this case study, as all the documents and figures are easily at hand. The following steps will be treated:

- Functional requirements
- Choice of physics process
- Location of the sensor
- Choice of the sensor

6.1 Functional requirements of the LHC luminosity monitor

The monitor under discussion is aimed at giving a relative luminosity reading for machine optimisation, but it is not required to give the absolute luminosity (as defined above) for the calculation of the underlying cross-section of the experimental physics processes. Hence the system does not need an absolute calibration. In addition, the monitor is to be used to study the variation in luminosity between the individual bunches of the LHC. This means that the detector has to have the bandwidth of the individual bunch crossings, which is 40 MHz. The expected difference in luminosity between bunches is very small, so a resolution of 1% is required.

6.2 Choice of the physics process for the detector

Since any count-rate coming from the collision point of the beam particles can be used as a signal for luminosity monitoring there is a wide choice of physics processes that could be used for this measurement. Due to both cost and integration issues the detector has to be small in size, hence huge detector arrays covering a large solid angle of secondary particle production can be excluded. This eliminates all well identified physics processes producing particles with large transverse momenta. Diffractive beam particle interactions, for which at least one of the incoming protons dissociates into a leading (high energy) neutron plus other secondary particles have therefore been considered as a source for the luminosity signal. Due to the nature of this process, most of the secondary neutrons are emitted into a very small solid angle in the forward direction. The properties of this interaction are pretty well known from lower energies, so that the cross section can be assumed to within a 10% accuracy. This process has been chosen as the basis for LHC luminosity monitoring.
6.3 Location of the sensor

Since the neutron production is in the very forward region, a monitor located close to the interaction point would have to be inside the vacuum chamber. In order to measure the forward neutron flux outside the vacuum chamber one also needs at least one intervening deflecting dipole to bend away the charged beam particles. Looking at the design of one of the high luminosity LHC interaction regions (Fig. 26) one can spot two large metallic objects, which are introduced to shield the superconducting elements from the particle flux of the collision products. The TAS is intended to shield the inner triplet from secondary particles, while the TAN is designed to absorb the forward neutron flux, which is just the signal required for luminosity monitoring.

Fig. 26 Layout of one of the LHC interaction regions.

Fig. 27 shows the simulated secondary particle flux at location of the TAN. The circles indicate the location of the two beam pipes. If the neutron flux (centre image) is weighted with the average particle energy, it becomes the dominant signal at this location. The luminosity detector will therefore be located inside the TAN (a 4m long copper block) between the two beam pipes.

Fig. 27 Simulated secondary particle flux distribution at the location of the TAN.

6.4 Choice of the sensor

Having selected the location, the physical dimensions of the detector are limited by installation constraints. Moreover, since the TAN absorber is designed to shield the superconducting elements from an enormous flux of secondary particles, the luminosity detector will have to withstand a very high radiation dose. If one assumes 20 years of operation at nominal luminosity, the integrated neutron flux will be $10^{18}$ n/cm$^2$. This is about 3 orders of magnitude larger than normal so-called “radiation hard” semiconductor sensors can withstand. It should also be kept in mind that after some running time the whole installation will be so radioactive, that human intervention for repair will be nearly
impossible. This, in addition to the other stringent requirement of a bandwidth of 40MHz meant that only two different detector technologies were retained and studied in detail:

1. A polycrystalline cadmium telluride (CdTe) detector array [24]
2. A pressurised ionisation chamber with continuous gas exchange [25]

The first option has the advantage of high bandwidth, but the radiation hardness is not completely demonstrated. The second option is believed to withstand the high radiation levels but, even after some years of optimisation, the bandwidth is still somewhat too low. Due to the high radiation doses expected, however, the most likely choice will be the ionisation chamber. Fig. 28 shows the design of the ionisation chamber presently used with test beams.

### Fig. 28 Schematic of the ionisation chamber for LHC luminosity monitoring.

**Segmented, multi-gap, pressurized gas ionization chamber**

- segmented into quadrants (each 4cm x 4cm)
- 60 gaps, 10 parallel x 6 series
- 0.5 mm gap spacing
- copper, ceramic construction
- 4 atmospheres Ar + 1%N₂ gas mixture, e- drift velocity 2.3 cm/μm

→ Detector location at shower maximum in TAN absorber

**6.5 Electronics**

The most critical element in the system is the detector itself, due to the requirements for bandwidth and radiation hardness. Once a signal is produced, it will be sent out of the TAN block via special radiation hard cables (stainless steel with a mineral insulator) to standard preamplifiers and digitising electronics. The signals will be recorded for each individual bunch crossing and averaged over many machine turns in order to produce the luminosity information for the control room.

**7. SOME EXAMPLES OF BEAM DIAGNOSTICS**

This section is meant to serve as general entertainment for those readers who have made it to here with their reading! Two examples from CERN-LEP operation have been selected, and show how difficult it can be to interpret primary measurements and decide on the right actions for solving a problem in an accelerator.

**7.1 The CERN-LEP beam does not circulate!**

The schedule for the CERN-LEP accelerator had a very regular structure. Every year LEP was used for about 8 months for physics beams followed by a 4 month maintenance and upgrade shutdown. During this shutdown major intervention work was sometimes carried out on the machine. At the next start-up it was therefore often expected that typical problems such as inverted magnet polarities would have to be overcome. One year, the start-up was particularly bad, with neither the electron beam nor the positron beam capable of being injected and made to circulate. Several hours were used to check all vacuum conditions, power supply currents, settings of the radio frequency system, injection deflectors...
and so on, but nothing indicated a severe problem. Finally people started to look in detail at the measured beam trajectory from the injection point onwards. A typical example for the positron beam is shown in Fig. 29.

![Fig. 29 Measurement of the LEP phase advance when beams did not circulate.](image)

What is actually shown in Fig. 29 is the phase advance from one beam position monitor to the next, as calculated from the measured beam trajectory. At a particular quadrupole (QL10.L1) the regular pattern is distorted. Additional measurements also indicated that most of the beam was lost at this point. The first conclusion was to suspect a problem with this quadrupole. People went in, measured the current in the quadrupole, checked its polarity, inspected its coils, but could not find anything abnormal. The indications of the beam measurements, however, clearly pointed to a problem at this location. After many discussions and potential hypotheses it was decided to open the vacuum chamber. It should be noted that this was a major intervention, causing a stop of the accelerator for at least one day. One can understand the surprise of the intervention team when they looked into the open vacuum chamber and saw a beer bottle!!!

During the shutdown intervention, somebody had sabotaged the LEP accelerator and had inserted a beer bottle into the beam pipe (Fig. 30)! What had upset the operation team most at the time was the fact that it was a very unsocial form of sabotage - the bottle was empty!

![Fig. 30 The mystery of the beam circulation problem in LEP is solved!](image)

### 7.2 The beam gets lost during the beta squeeze

This again is one of the stories from LEP operation which took several hours of beam diagnostics to solve. The problem in itself is pretty complex, and therefore requires some additional explanations beforehand.
The acceleration of the particle beams and the change of the lattice function in the insertion regions in order to get smaller values of the beta-function at the crossing point (hence higher luminosity) are so called “dynamic processes”. The presence of the beam requires that all actions are well synchronised. For example, the power converters of all relevant magnetic circuits have to be controlled such that beam parameters like the closed orbit, tunes and chromaticities stay within tolerance during the dynamic process. In order to achieve this, the behaviour of these beam parameters is periodically measured as a function of time and the corresponding power converter tables are updated.

During one period of LEP operation it was found that the beams were lost during the beta squeeze. Shortly before the total loss of the beams a significant beam loss was measured. As standard practice when encountering such problems, the engineer in charge (EIC) launched a new machine cycle with diagnostics facilities such as “tune history” (the measurement of the betatron tunes as a function of time – see Section 3) switched on. This indicated that the vertical tune moved out of tolerance during the beta squeeze. Fig. 31 shows an excerpt from the actual LEP logbook entry of this event.

As a result of this observation, the EIC launched another cycle, but inserted a breakpoint (to stop the accelerator cycle) just before the critical moment in the beta squeeze when the deviation in tune occurred. Having reached that breakpoint the tunes were measured statically and found to be perfectly within tolerance. The beta squeeze was then executed step by step, and to the big surprise of the operations crew, the tunes were found to be correct at all times. The beam had passed the beta squeeze like on an ordinary day! But on the next attempt, without a break in the cycle, the beam was again lost at the same moment, and several people scratched their heads to find an explanation.

Finally, the following measurement was made. The machine was prepared and a breakpoint again inserted just before the critical beam loss. Once this point was reached, the EIC requested the execution of one further step in the beta squeeze. The facility by which one could execute a single step in a dynamic process had the additional feature that one could specify the rate of current change of any machine element. This current rate limitation was changed from 25A/s (nominal) down to 2.5 A/s on consecutive steps. The corresponding tune history (the result from the vertical plane is plotted on the lower graph) is shown in Fig. 32.

One can clearly see that a huge (negative) tune excursion occurred when the step was executed at the nominal rate. This observation led the EIC to the right conclusion, which was that one of the power supplies was able to deliver the demanded current statically, but not dynamically. When this
was discussed with experts from the power converter group, they indicated that the power supplies for the superconducting insertion quadrupoles were built as two blocks in series, each of them able to deliver the necessary current (each block typically 1000 A/10 V). Both of these blocks were required to have enough voltage margins to enforce a current change against the inductance of the quadrupole coil. This then explained the whole story. One of these blocks was faulty, but since the power converter could deliver its (static) current, it was not detected by an alarm or surveillance circuit. In the static case the working, single block could deliver the requested current. If the dynamic rate was too high, however, this single block could not provide enough current leading it to lose synchronism with the other power converters. This resulted in the large tune change observed and ultimately the total beam loss.

These two examples show the enormous potential of beam instrumentation if they are used in the right combination by intelligent people.

REFERENCES


INTRODUCTION TO RF LINEAR ACCELERATORS

N. Pichoff
CEA/DIF/DPTA/SP2A/LFPA, France

Abstract
After a short introduction to applications of RF linacs and their advantages and drawbacks as opposed to circular accelerators, the model of RF resonant cavities and their excitation by RF sources or beam is introduced. Then beam dynamics notions, essential to linacs, such as transit-time factor, synchronism, r.m.s. properties, matching and mismatching in linear or non-linear forces, are presented.

1 INTRODUCTION
A one-hour lecture on RF linear accelerators (linacs) and a ten-hour course were given to the CAS students. The short lecture introduced the students to RF specifics and beam dynamics basics giving them a good understanding of linacs. This paper deals with the notions introduced in the lecture. Students eager to learn more about linacs are advised to read the books in Refs. [1] and [2].

After a short introduction to RF linac applications and their advantages/disadvantages as opposed to circular accelerators, this paper is divided into two parts:

– Section 3 introduces the RF cavity through its basic principle and model, the notion of RF modes, and the way they are excited either from the RF source or by the beam.

– Section 4 gives useful notions of beam dynamics in linacs: the transit time factor; the notion of synchronism; the particle motion in continuous non-linear forces (longitudinal dynamics); and periodic linear forces (transverse dynamics). The notion of beam r.m.s. properties and matching in the linac is discussed. The effects of non-linear forces on emittance growth are introduced.

2 WHY RF LINACS?
The goal of a particle accelerator is to produce a ‘low-cost wanted’ beam. By ‘wanted’, one means a given particle type, with a given intensity, at a given energy within a given emittance (or brightness) in a given time structure. Costs should cover construction, operation, and personnel.

Synchrotrons, cyclotrons, and RF linear accelerators (linacs)1 can all achieve this.

The main advantages of linacs are that

– they can handle high current beams (they are less limited by tune shift),

– they can run in high duty-cycle (the beam passes only once at each position),

– they exhibit low synchrotron radiation losses (no dipoles).

Their main drawbacks are that

– they consume space and cavities,

– the synchrotron radiation damping of light particles (electrons/positrons) cannot be easily used to reduce the beam emittance.

1 Electrostatic machines are also suitable for low-current, low-energy beams.
That is why linacs are mainly used:

- as low-energy injectors (where the space-charge force is more important and the duty-cycle is high),
- with high-intensity/power proton beams (high space-charge level or/and duty cycle),
- in new lepton collider projects at very high energy (no radiation losses).

3 RF CAVITIES

The RF cavity gives energy to the beam. As the cost of the RF generally represents the main expense of the linac structure apart from the building, the choice of the RF structure has to be studied very carefully. This paper presents only the principle of an RF cavity. More precise information can be found in the CAS dedicated to RF [3].

3.1 A standing-wave RF cavity

3.1.1 Field calculation

An RF cavity is simply a piece of conductor enclosing an empty volume (generally a vacuum). Solutions of Maxwell’s equations in this volume, taking into account the boundary conditions on the conductor, allow the existence of electromagnetic field configurations in the cavity. These are called the resonant modes.

<table>
<thead>
<tr>
<th>Maxwell’s equations</th>
<th>Boundary conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ \nabla \cdot \vec{E} = \frac{\rho}{\varepsilon_0} , ]</td>
<td>[ \vec{n} \times \vec{E}_n = \vec{0} , \quad \vec{n} \cdot \vec{B}_n = \vec{0} , ]</td>
</tr>
<tr>
<td>[ \nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} , ]</td>
<td>[ \vec{n} \cdot \vec{E}_n = \frac{\sum \Sigma}{\varepsilon_0} , \quad \vec{n} \times \vec{H}_n = \vec{K} . ]</td>
</tr>
<tr>
<td>[ \nabla \times \vec{B} = \mu_0 \vec{j} + \frac{1}{c^2} \frac{\partial \vec{E}}{\partial t} . ]</td>
<td>[ \vec{n} \times \vec{B}_n = \vec{0} , \quad \vec{n} \cdot \vec{E}_n = \vec{0} , ]</td>
</tr>
</tbody>
</table>

\( \mu_0 = 4\pi \cdot 10^{-7} \text{T} \cdot \text{m} \cdot \text{A}^{-1} \) : permeability of free space,
\( \varepsilon_0 = 1/\mu_0 c^2 \) : permittivity of free space,
\( c = 2.99792458 \cdot 10^8 \text{m} \cdot \text{s}^{-1} \) : speed of light in vacuum.

Each mode, labelled \( n \), is characterized by an electromagnetic field amplitude configuration \( \vec{E}_n(\vec{r})/\vec{B}_n(\vec{r}) \) oscillating with an RF frequency \( f_n \). The electric field amplitude configuration is the solution of the equation:

\[ \nabla^2 \vec{E}_n + \frac{\omega_n^2}{c^2} \vec{E}_n = \vec{0} , \]  

(1)

where \( \vec{E}_n(\vec{r}) \) should satisfy the boundary conditions and \( \omega_n = 2\pi \cdot f_n \) is the mode pulsation.

The electric field in the cavity is a weighted sum of all the modes:

\[ \vec{E}(\vec{r}, t) = \sum e_n(t) \cdot \vec{E}_n(\vec{r}) = \sum a_n \cdot e^{i\omega_n t} \cdot \vec{E}_n(\vec{r}) . \]  

(2)

Here \( a_n \) is a complex number and \( e_n(t) \) is the field variation with time, it is the solution of [4]:

\[ \nabla \cdot \vec{E} = \frac{\rho}{\varepsilon_0} , \quad \nabla \times \vec{E} = \mu_0 \vec{j} + \frac{1}{c^2} \frac{\partial \vec{B}}{\partial t} . \]  

\[ \mu_0 = 4\pi \cdot 10^{-7} \text{T} \cdot \text{m} \cdot \text{A}^{-1} ; \] \[ \varepsilon_0 = 1/\mu_0 c^2 ; \] \[ c = 2.99792458 \cdot 10^8 \text{m} \cdot \text{s}^{-1} ; \]
\[
\hat{e}_n + \alpha_n^2 \cdot e_n = -\frac{\omega_n^2}{\sqrt{\varepsilon \mu}} \int_S \left( \vec{E} \times \vec{H}_n \right) \cdot \hat{n} \cdot dS \\
+ \frac{1}{\varepsilon} \frac{d}{dt} \int_S \left( \vec{H} \times \vec{E}_n \right) \cdot \hat{n} \cdot dS' - \frac{1}{\varepsilon} \frac{d}{dt} \int_\Omega \vec{J}(\vec{r},t) \cdot \vec{E}_n(\vec{r}) \cdot dV
\]

. (3)

Here \( \vec{H} \) is the magnetic induction. It is often used close to the surface in place of \( \vec{B} \), as unlike \( \vec{B} \), it is macroscopically continuous through the surface. \( \vec{J} \) is the current density, of the beam for example. The first term on the right-hand side is an integration over the conductor which is not a perfect conductor. Because of power losses by Joule effects, it can be rewritten as a damping term:

\[
-\frac{\omega_n}{Q_{on}} \cdot \hat{e}_n
\]

. (4)

The calculation of \( Q_{on} \), the quality factor of the mode, can be deduced from power loss considerations:

\[U_n(0)\] is the energy stored by the \( n \)-mode at time \( t = 0 \). For \( t > 0 \), no more power is injected in the cavity. Let us define \( k(t) \) as:

\[k(t) = \frac{e_n(t)}{e_n(t = 0)}\]

. (5)

The energy lost per unit time is the power dissipated in the conductor \( P_n \):

\[
\frac{dU_n(t)}{dt} = -P_n(t).
\]

The average power dissipated in the conductor per cycle is proportional to the square of the current density (and then the magnetic field) close to the surface:

\[
P_n = \frac{R_s}{2} \int S K_n^2 dS = \frac{R_s}{2} \int S H_n^2 dS,
\]

where \( R_s \) is the surface resistance defined as:

\[
-R_s = \frac{\mu_0 \pi}{\sigma}, \text{ for normal conductors}
\]

. (8)

where \( \sigma \) is the conductor conductivity (\( 1/\sigma = 1.7 \cdot 10^{-7} \Omega \cdot m \) for copper).

\[
-R_s \approx R_{res} + 9 \cdot 10^{-8} \frac{f_0^2 (\text{GHz})}{T (\text{K})} \exp \left(-1.92 \cdot \frac{T_c}{T}\right), \text{ for superconducting niobium}
\]

. (9)

where \( R_{res} \) is the residual resistance (\( 10^{-9} – 10^{-8} \Omega \)) depending on the surface imperfections, \( T \) is the working absolute temperature, \( T_c = 9.2 \text{ K} \) is the critical temperature.

From Eqs. (5) and (7) can be deduced:

\[
P_n(t) = k(t)^2 \cdot P_n(t = 0).
\]

The stored energy is proportional to the square of the field:

\[
U_i = \frac{E_0}{2} \int \int \int \vec{E} \cdot \vec{E} = \frac{1}{2\mu_0} \int \int \int \vec{B} \cdot \vec{B} \, dv,
\]

then:

\[
U_n(t) = k(t)^2 \cdot U_n(t = 0).
\]

107
Equation (6) becomes:
\[
\frac{dk^2}{dt} = -\frac{P_n}{U_n} \cdot k^2 = 2 \cdot kk ,
\]  
(13)
giving:
\[
\frac{de_n}{dt} = -\frac{P_n}{2 \cdot U_n} \cdot e_n .
\]  
(14)
A comparison with the damping term written in (4) gives:
\[
Q_{0n} = \frac{\Omega_n \cdot U_n}{P_n}.
\]  
(15)
In the second term, the integration is performed over the open surfaces \( S \) and represents the coupling with the outside system. This coupling can be divided into two contributions:

- the injected power coming from the power generator through the coupler,
- an additional damping, which can be represented by another quality factor \( Q_{\text{ext}} \) known as the external \( Q \), corresponding to power losses through the opened surfaces. The coupling can be calculated from the coupler geometry with electromagnetic codes.

\[
S_n \cdot e^{i(\omega t + \phi_n)} \] is the RF source filling through the coupler.

The last term, represents the field excited by the beam, known as the beam loading. It is proportional to the beam intensity:
\[
k_n \cdot I(t).
\]  
(17)
\( I(t) \) is a complex number (it has a phase) representing the beam current.

Equation (3) can then be modelized by:
\[
\frac{d^2 e_n}{dt^2} + \frac{\omega_{\text{RF}}}{Q_n} \cdot \frac{de_n}{dt} + \omega_n^2 \cdot e_n = S_n \cdot e^{i(\omega t + \phi_n)} + k_n \cdot I(t),
\]  
(18)
which is the equation of a damped harmonic oscillator in a forced regime. \( Q_n \) is the quality factor of the cavity, with
\[
\frac{1}{Q_n} = \frac{1}{Q_{0n}} + \frac{1}{Q_{\text{ext}}},
\]
and
\[
\tau = 2 \cdot \frac{Q_n}{\omega_{\text{RF}}}
\]
is the cavity filling time.

Note that both the coupler or the beam can excite some RF modes.

Equation (18) shows an RLC circuit which is often used to modelize the system. A complete study of this model can be found in Ref. [4].
From these modes, one with a field amplitude along the longitudinal direction on the axis is used to accelerate the beam. The geometry of the cavity is then calculated to match the frequency of this accelerating mode to the RF frequency. This mode is excited in the cavity through a power coupler whose geometry is calculated and adjusted to transfer electromagnetic energy in the cavity to the beam without reflection—a process called coupler matching.

3.1.2 Shunt impedances

To first order, only the accelerating mode is excited in the cavity. The transverse component of the electric field is generally null along the axis. An expression of the \( z \) component of the field on the axis is then:

\[
E_z(s, t) = E_z(s) \cdot \cos(\alpha t + \varphi).
\] (19)

The field amplitude is \( E_z(s) \).

One defines the cavity voltage \( V_0 \) as:

\[
V_0 = \int_{-\infty}^{\infty} |E_z(s)| \, ds.
\] (20)

Then \( q \cdot V_0 \) represents the maximum energy (in eV) that a particle with charge \( q \) could gain if the field was always maximum.

Let \( P_d \) be the power deposition in the cavity:

\[
P_d = \frac{V_0^2}{2 \cdot R}.
\] (21)

The cavity shunt impedance \( R \) is very useful in cavity design. For optimum acceleration, it has to be as high as possible.

Because the electric field changes with time as the particle transits through the cavity, the maximum energy \( q \cdot V \) that can be gained in the cavity by a particle of charge \( q \) is lower than \( q \cdot V_0 \). One defines the transit-time factor \( T \) as:

\[
T = \frac{V}{V_0} \leq 1.
\] (22)

This corrective factor to the energy gain takes the particle transit time in the cavity into account, and is obviously dependent on the particle velocity. The calculation of this factor is described in Section 4.

The effective shunt impedance \( RT^2 \) is then proportional to the ratio between the square of the maximum energy \( \Delta U_{\text{max}} \) that can be gained by the beam and the power lost in the cavity:

\[
RT^2 = \frac{\Delta U_{\text{max}}^2}{2P_d}.
\] (23)

It is some sort of cavity efficiency and has to be maximum.

The shunt impedance is often used to compare the efficiency of different structures at a given energy. Usually, the geometry is different, so one extends the preceding definition per unit length to allow a better comparison.
Let $L$ be the cavity length. The mean cavity electric field $E_0$ is defined as:

$$E_0 = \frac{V_0}{L}.$$  \hfill (24)

The power deposition per unit length in the cavity $P'_d$ is then:

$$P'_d = \frac{E_0^2}{2 \cdot Z},$$  \hfill (25)

where $Z$ is the cavity shunt impedance per unit length.

The effective shunt impedance per unit length $ZT^2$ is then proportional to the ratio between the square of the maximum energy $\Delta U'_{\text{max}}$ that can be gained per unit length by the beam and the power lost per unit length in the cavity:

$$ZT^2 = \frac{\Delta U'_{\text{max}}^2}{2P'_d}.$$  \hfill (26)

As it is depending on the particle velocity, one chooses the structure that maximize $ZT^2$ at a given energy. Figure 1 represents the evolution of the effective shunt impedance per metre for two different structures (SDTL and CCL) with different apertures $\phi$. The higher the aperture (space for beam), the lower the effective shunt impedance. SDTL structures are more efficient at lower energy, CCL structures are more efficient at higher energy. The optimum transition energy is around 100 MeV for protons.

![Fig. 1: Effective shunt impedance per metre of different TRISPAL structures (C. Bourat)](image)

### 3.2 A travelling-wave RF cavity

A travelling-wave cavity is generally used to accelerate ultrarelativistic particles. These cavities generally have two power ports. One where the power enters, and another, at the other end, where the power exits (Fig. 2). The electric field travels through the cavity from the input to the output port. Its phase velocity is adjusted to the beam velocity. The field phase is adjusted to continuously accelerate the beam.

---

2 Owing to the cavity fringe field, $L$ is often arbitrarily defined as the physical length of the cavity.
The RF phase velocity in empty cavities or wave-guides is usually higher than (or equal to) the speed of light in vacuum $c$. As particle velocity cannot exceed $c$, the RF phase velocity should be decelerated to reach the synchronism condition by introducing some periodic obstacles into the guide (such as iris-loaded waveguide). The periodic field can then be expanded into a Fourier series with different wave numbers:

$$E_z(t, z) = \sum_{n=-\infty}^{\infty} e_{zn} \cdot \exp \left[ j \cdot (\omega t - k_n z) \right], \quad (27)$$

with $e_{zn}$ the space harmonic amplitude, $k_n$ the space harmonic wave numbers,

$$k_n = k_0 + \frac{2\pi n}{d}, \quad (28)$$

d the obstacle period, and $k_0$ the waveguide number.

The phase velocity $v_n$ of space harmonic number $n$ is

$$v_n = \frac{\omega}{k_n}. \quad (29)$$

Particles whose velocity is close to the phase velocity of one space harmonic exchange energy with it. Otherwise, the average effect is null.

A complete calculation of these insertion obstacles as well as a large bibliography can be found in Ref. [5]. This kind of travelling-wave accelerating structure is mainly used to accelerate ultra-relativistic electrons.

Moreover, the model of a travelling-wave acceleration, even with acceleration with standing-wave cavities, is often used to simplify the calculation of the longitudinal-motion equations.

4 ELEMENTS OF BEAM DYNAMICS

4.1 The transit-time factor and the particle synchronous phase

A cavity has a finite length $L$. The cavity input abscissa is $s_0$, and $E_z(s)$ is the amplitude of the electric field longitudinal component on axis.

The energy gained by a charged particle on axis in the cavity is

---

This is actually the longitudinal energy, but we can consider that there is no transverse field on the cavity axis.
\[ \Delta W = \int_{s_0}^{s_0 + L} qE_z(s) \cdot \cos[\phi(s)] \cdot ds , \]

(30)

where \( q \) is the particle charge, \( \phi(s) \) is the cavity RF phase when the particle is at abscissa \( s \). It is defined as

\[ \phi(s) = \phi_0 + \frac{\omega}{c} \int_{s_0}^{s_0 + s} \frac{ds'}{\beta_z(s')} , \]

(31)

where \( \phi_0 = \phi(s_0) \) is the RF phase when the particle enters the cavity.

Writing \( \phi(s) = \phi(s) + (\phi_s - \phi_0) \), \( \phi_s \) being an arbitrary phase and using trigonometric relationships, one gets for the energy gain:

\[ \Delta W = \cos \phi_s \cdot \int_{s_0}^{s_0 + L} qE_z(s) \cdot \cos[\phi(s) - \phi_s] \cdot ds - \sin \phi_s \cdot \int_{s_0}^{s_0 + L} qE_z(s) \cdot \sin[\phi(s) - \phi_s] \cdot ds . \]

(32)

By defining \( \phi_s \) as:

\[ \int_{s_0}^{s_0 + L} qE_z(s) \cdot \sin[\phi(s) - \phi_s] \cdot ds = 0 , \]

giving the definition of the synchronous phase \( \phi_s \):

\[ \phi_s = \arctan \left( \frac{\int_{s_0}^{s_0 + L} E_z(s) \cdot \sin[\phi(s)] \cdot ds}{\int_{s_0}^{s_0 + L} E_z(s) \cdot \cos[\phi(s)] \cdot ds} \right) , \]

(33)

one finally gets

\[ \Delta W = \left( q \int_{s_0}^{s_0 + L} E_z(s) \cdot ds \right) \cdot T \cdot \cos \phi_s = qV_0 \cdot T \cdot \cos \phi_s , \]

(34)

with

\[ T = \frac{1}{V_0} \int_{s_0}^{s_0 + L} E_z(s) \cdot \cos[\phi(s) - \phi_s] \cdot ds . \]

(35)

The transit-time factor \( T \), depends on the particle initial velocity as well as on the field amplitude. This definition does not make any assumption about the field shape (no symmetry) resulting from a slightly different synchronous-phase definition, which can be found in the literature (which is often taken as the RF phase when the particle reaches mid-cavity). When the velocity gain in the cavity is much lower than the input-particle velocity, \( T \) depends only on the velocity and can be easily tabulated.
The calculation of $T$ with formula (35) is sometimes difficult to perform, as $\phi_s$ has to be known. In fact, $T$ does not depend on $\phi_s$ when the velocity gain is small and another formula (a little bit more difficult to understand physically) can be used:

$$T = \frac{1}{V_0} \left[ \int E_z(s) \cdot e^{i\phi(s)} \cdot ds \right].$$

(36)

4.2 Notion of synchronism

A linac is designed so that one theoretical particle called the synchronous particle enters successively along the axis of RF cavities with a wanted RF phase law in order to get a wanted energy gain. This very important notion of synchronism allows the understanding of the efficiency and the stability of linacs.

Particles can be accelerated with travelling waves as well as standing waves (Fig. 3).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig3.png}
\caption{Particle accelerated by a travelling wave}
\end{figure}

4.2.1 Acceleration with travelling waves

The on-axis RF accelerating field can be written as:

$$E_z(z,t) = E_0 \cdot \cos(\omega t - kz),$$

(37)

where $\omega$ is the RF pulsation and $k$ is the RF wave number.

The synchronism condition is reached when the particle-longitudinal velocity equals the RF phase velocity:

$$\beta_z c = \frac{k}{\omega}.$$  

(38)

Here $c$ is the speed of light in vacuum, $\beta_z$ is the reduced longitudinal velocity of the synchronous particle. Note that when the paraxial approximation\(^4\) is used, $\beta_z$ is replaced by $\beta$, the reduced total speed of the particle.

\(^4\) As $\beta = \beta_z \cdot \sqrt{1 + x'^2 + y'^2}$, paraxial approximation occurs when $x'<< 1$ and $y'<< 1$.  

113
4.2.2 Standing waves

In most linacs, the beam is accelerated with RF cavities or gaps operating in standing-wave conditions. An RF power, produced by one or many RF sources, is introduced through a coupler in a resonant cavity exciting the wanted standing-wave accelerating mode. The cavity shape has been calculated and adjusted to match the accelerating mode to the power-supply frequencies and to throw the other mode frequencies far from the RF one.

As a first step, let us assume a set of thin independently phased RF cavities along the beam path (Fig. 4).

<table>
<thead>
<tr>
<th>RF phase</th>
<th>$\phi_{i-1}$</th>
<th>$\phi_{i}$</th>
<th>$\phi_{i+1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Particle velocity</td>
<td>$\beta_{i-1}$</td>
<td>$\beta_{i}$</td>
<td>$\beta_{i+1}$</td>
</tr>
<tr>
<td>Distances</td>
<td>$D_{i-1}$</td>
<td>$D_{i}$</td>
<td>$D_{i+1}$</td>
</tr>
<tr>
<td>Synchronous phase</td>
<td>$\phi_{i-1}$</td>
<td>$\phi_{i}$</td>
<td>$\phi_{i+1}$</td>
</tr>
<tr>
<td>Cavity number</td>
<td>$i-1$</td>
<td>$i$</td>
<td>$i+1$</td>
</tr>
</tbody>
</table>

**Fig. 4:** A set of independently phased cavities

- $\phi_i$ is the absolute RF phase in the $i^{th}$ cavity when $t = 0$ (the $t = 0$ instant has been arbitrarily chosen),
- $\beta_{si}$ is the synchronous, particle reduced velocity at the $i^{th}$ cavity output,
- $\phi_{si}$ is the RF synchronous phase of the $i^{th}$ cavity of the synchronous particle,
- $D_i$ is the distance between the $i^{th}$ and the $i+1^{th}$ cavities.

The synchronism condition is reached when:

$$\phi_{si+1} - \phi_{si} = 2\pi \cdot \frac{D_i}{\beta_{si} \lambda} + \phi_{i+1} - \phi_{i} + [2\pi n].$$  \hspace{1cm} (39)

The RF wavelength is $\lambda = c / f$.

One observes that the synchronism condition does not depend on the RF field amplitude. It has a non-intuitive consequence: an increase of the accelerating field amplitude in the cavities without phase change does not induce an increase of the synchronous-particle final energy but a change of the synchronous phase fulfilling the synchronism condition.

Two different kinds of structures exist:

- The coupled-cavity structures where the phase between cavities is fixed. The synchronism condition is achieved by adjusting the distance between the cavities.

In a Drift-Tube Linac (DTL), for example, the phase difference between the cells is fixed (= $2\pi$). The distance between cells is then calculated to have:

$$D_i = \left(\frac{\phi_{si+1} - \phi_{si}}{2\pi} + 1\right) \cdot \beta_{si} \lambda.$$  \hspace{1cm} (40)

---

5 Do not confuse the synchronous-particle phase, the phase of the synchronous particle in a cavity and a particle synchronous-phase, the synchronous phase of a particle (whatever it is) in a cavity.
The independent cavity structures where the distance between cavities is fixed. The synchronism condition is then achieved by adjusting the phase difference between the cavities.

In a Superconducting-Cavity Linac (SCL), for example, the distance between cavities is fixed by the cryogenics mechanism. The phase difference between cavities is then calculated to have:

$$\phi_{i+1} - \phi_i = \phi_{ii+1} - \phi_{ii} - 2\pi \frac{D_i}{\beta_i \lambda} + [2\pi n]. \quad (41)$$

### 4.3 Particle motion in electromagnetic fields

#### 4.3.1 Basis

The electromagnetic field can be divided into two contributions:

- The electric field: $\vec{E}$.
- The magnetic field: $\vec{B}$.

The intensity of these contributions depends on the referential where they are expressed. The equation of motion of a particle of charge $q$ in these fields is:

$$\frac{d\vec{p}}{dt} = q \cdot (\vec{v} \times \vec{B} + \vec{E}), \quad (42)$$

where $\vec{p}$ is the momentum of the particle and $\vec{v}$ is its velocity.

Let us call $s$ the abscissa of the beam in the linac path (rather than $z$ to avoid any confusion with the particle longitudinal position in the bunch), the equation of motion can be rewritten:

$$\frac{d\vec{p}}{ds} = q \cdot \frac{\vec{v} \times \vec{B} + \vec{E}}{v_z}, \quad (43)$$

where $v_z$ is the particle longitudinal velocity.

A projection on the Cartesian axis$^6$ ($x$, $y$, $z$) gives:

$$\begin{align*}
\frac{d\gamma \beta_x}{ds} &= \frac{q}{mc} \left( y' B_z - B_y + \frac{E_x}{v_z} \right) = \frac{d\gamma \beta_x'}{ds} \\
\frac{d\gamma \beta_y}{ds} &= \frac{q}{mc} \left( B_x - x' B_y + \frac{E_y}{v_z} \right) = \frac{d\gamma \beta_y'}{ds} \\
\frac{d\gamma \beta_z}{ds} &= \frac{q}{mc} \left( x' B_y - y' B_x + \frac{E_z}{v_z} \right)
\end{align*}\quad (44)$$

Here

$$x' = \frac{dx}{ds} = \frac{p_x}{p_z} = \frac{\beta_x}{\beta_z}$$

and

$$y' = \frac{dy}{ds} = \frac{p_y}{p_z} = \frac{\beta_y}{\beta_z}$$

$^6$ In general, $x$ and $y$ play the same role in a linac (contrary to their role in a circular accelerator).
are the slopes of the particle, and
\[ \beta_w = \frac{v_w}{c}, \]
is the reduced velocity \( w \)-component, \( w \) being \( x, y \) or \( z \); \( v_w \) is the particle velocity \( w \) component; \( m \) and \( q \) are, respectively, the rest mass and the particle charge; \( c \) is the speed of light.

One clearly observes that longitudinal and transverse motions are coupled. However, for an easier understanding, and because the coupling is often very weak, the longitudinal and the transverse motions are usually treated as uncoupled, the longitudinal velocity \( v_z \) variations are considered separately. To uncouple the transverse and longitudinal motions, the \textit{paraxial approximation} has to be done.

4.3.2 \textbf{Paraxial approximation}

The \textit{paraxial approximation} is based on the assumption that \( x'^2 + y'^2 \ll 1 \).

Its natural consequence is
\[ \beta_z = \beta \cdot \sqrt{1 + x'^2 + y'^2} = \beta. \] (45)

For \( x' < 100 \) mrad and \( y' < 100 \) mrad, the error on \( \beta \) (or \( \beta_z \)) is lower than 1%.

This approximation is quite accurate at high energy where the beam divergence is small, but is more difficult to justify at very low energy.

4.3.3 \textbf{Energy gain calculation}

From Eqs. (44), one can easily obtain the energy gain:
\[ \frac{d\gamma}{ds} = \beta_z \left( x' \cdot \frac{dy\beta_x}{ds} + y' \cdot \frac{dy\beta_y}{ds} + \frac{dy\beta_z}{ds} \right), \] (46)
giving:
\[ \frac{d\gamma}{ds} = \frac{q}{mc^2} \cdot \left( x'E_x + y'E_y + E_z \right). \] (47)

One finds the well-known result that only the electric field contributes to energy gain.

4.4 \textbf{Longitudinal particle dynamics (motion in non-linear force)}

4.4.1 \textbf{The longitudinal variables}

The variables generally used to describe the longitudinal particle motion, as a function of \( s \), are:

- \( \phi \), the absolute particle phase, calculated from the RF frequency, with \( \phi = 0 \) arbitrarily chosen.
- \( W \), the particle kinetic energy.\(^7\)

The evolution of these variables with \( s \) is given by the equations:
\[
\begin{align*}
\frac{d\phi}{ds} &= \frac{\omega_f}{\beta c} = \frac{2\pi}{\beta \cdot \lambda_c \cdot \sqrt{1 - x'^2 - y'^2}} \, . \\
\frac{dW}{ds} &= q \cdot \left( x'E_x(s, \phi, r) + y'E_y(s, \phi, r) + E_z(s, \phi, r) \right)
\end{align*}
\] (48)

\(^7\)This is really a ‘longitudinal’ particle property only in paraxial approximation.
Applying these equations to the synchronous particle, one gets:

\[
\begin{align*}
\frac{d\phi_s}{ds} &= \frac{2\pi}{\beta_s \cdot \lambda_{rf}}, \\
\frac{dW_s}{ds} &= q \cdot E_z (s, \phi_s, 0).
\end{align*}
\]  

(49)

Let us define the reduced phase and energy variables for each particle:

\[
\begin{align*}
\varphi &= \phi - \phi_s, \\
w &= W - W_s.
\end{align*}
\]  

(50)

Late particles have a positive \( \varphi \).

The equations of motion with these new variables become:

\[
\begin{align*}
\frac{d\varphi}{ds} &= \frac{2\pi}{\lambda_{rf}} \left( \frac{1}{\beta_s \sqrt{1-x'^2-y'^2}} - 1 \right), \\
\frac{dw}{ds} &= q \left[ x'E_x (s, \phi, r) + y'E_y (s, \phi, r) + E_z (s, \phi, r) - E_z (s, \phi_s, 0) \right].
\end{align*}
\]  

(51)

When the beam is accelerated by a standing-wave cavity structure, synchronous particles enter successive cavities giving it a strong energy gain, separated by long drift spaces where no acceleration occurs. In order to understand the physics, this periodic acceleration scheme can be replaced by a continuous acceleration one. This scheme assumes that the beam is accelerated by a travelling wave propagating at the same speed as that of the synchronous particle. This scheme allows a mathematical resolution of the dynamics equations with an electric field independent of \( s \).

4.4.2 The electric field model

The electric field, generally a function of \( s \), is the chosen constant. The field amplitude of the travelling wave is \( E_0 T_s \) (mean electric field) on axis. Here \( E_0 \) is defined as the potential gain of one cavity \( V_0 \) divided by the distance between the centres of consecutive cavities. The transit-time factor \( T \) has been included to take into account the variable efficiency of the acceleration in standing-wave cavities with the particle velocity.

The on-axis electric field longitudinal component becomes:

\[
E_z (s, \phi, r = 0) = E_0 T_s \cdot \cos(\varphi + \phi_0),
\]  

(52)

\( \phi_0 \) being the RF synchronous phase of the synchronous particle.

The energy gain per metre of the synchronous particle is then:

\[
G = qE_0 T_s \cdot \cos \phi_0.
\]  

(53)

Here \( T_s \) is the transit-time factor of the synchronous particle.

Let us assume an axisymmetric accelerating field, the off-axis electric field longitudinal component can be written:

\[
E_z (s, \phi, r) = E_0 T_s \cdot R(r) \cdot \cos(\varphi + \phi_0),
\]  

(54)

---

\textsuperscript{8} Equations are smoothed for analytic solutions, then quantified for a numerical solution.
r being the radial position of the particle, \( R(r) \) expressing the radial evolution of the electric field longitudinal component. It can usually be written as \( R(r) = 1 + O(r^2) \). Close to the axis, the Bessel function, solution of the Maxwell equations in axisymmetric geometry in vacuum, can be used to express \( R(r) \) \([2][6]\), but far from the axis, the cavity geometry has a strong influence through the boundary conditions. The radial position \( r \) can be replaced by \((x,y)\) if the cavity is not axisymmetric. Some authors include the variation of the field with \( r \) in the transit time factor: \( T(r) \).

From the relationship \( \vec{V} \cdot \vec{E} = 0 \) and remarking that the electric field transverse component is null on the axis, one gets the electric field transverse component:

\[
E_r(s, \varphi, r) = -\frac{1}{r} \int_0^r \frac{\partial E_z(s, \varphi, r)}{\partial s} r \cdot dr
\]

\[
= -\frac{1}{r} \frac{E_0 T}{\beta_s \lambda} \sin(\varphi + \phi_{s0}) \cdot \int_0^r R(r) \cdot r \cdot dr
\]  

(55)

The electric field radial component can be written:

\[
E_r(s, \varphi, r) = -\frac{E_0 T}{\beta_s \lambda} \sin(\varphi + \phi_{s0}) \cdot \left( \frac{r}{2} + O(r^3) \right)
\]  

(56)

Three assumptions are made to decouple the longitudinal motion from the transverse one:

- In general, we assume:

\[
\frac{L}{2} + O(r^3) \ll \beta_s \lambda,
\]

and the contribution of the transverse electric field to the energy gain can usually be neglected in Eq. (51):

\[
x' E_x + y' E_y \ll E_z - E_{zs}.
\]  

(57)

- Generally, the paraxial assumption occurs, and we consider:

\[
x'^2 + y'^2 \ll 1.
\]  

(58)

- Finally, we assume that the longitudinal field does not depend on the radial position \( r \), by taking:

\[
R(r) = 1.
\]  

(59)

4.4.3 The equations of motion

Using these assumptions, Eqs. (51) become:

\[
\begin{align}
\frac{d\varphi}{ds} &= -\frac{2\pi}{\lambda} \left( \frac{1}{\beta(s)} - \frac{1}{\beta_s(s)} \right), \\
\frac{dw}{ds} &= -q \cdot E_0 T \cdot \left( \cos \phi_{s0} \cdot (1 - \cos \varphi) + \sin \phi_{s0} \cdot \sin \varphi \right),
\end{align}
\]

(60)

which is in fact the equation of motion of on-axis particles.

Moreover, a small longitudinal velocity dispersion assumption can be carried out:

\[
\frac{1}{\beta} - \frac{1}{\beta_s} = \delta \beta^{-1} \ll \frac{1}{\beta_s},
\]  

(61)
and a first order development around synchronous velocity gives:

\[ \delta \beta^{-1} = -\frac{w}{(\beta_s \gamma_s)^3 \cdot mc^2}. \]

If one considers that the transit-time factor does not depend on the beam particles’ energy:

\[ T(w) = T_s, \]  

Eq. (60) becomes:

\[ \begin{aligned}
  \frac{d\varphi}{ds} &= -2\pi \cdot \frac{w}{(\beta_s \gamma_s)^3 \cdot mc^2 \cdot \lambda} = \frac{\partial H_{\varphi w}}{\partial w} \\
  \frac{dw}{ds} &= -q \cdot E_0 T_s \left[ \cos \phi_0 \cdot (1 - \cos \varphi) + \sin \phi_0 \cdot \sin \varphi \right] = -\frac{\partial H_{\varphi w}}{\partial \varphi}.
\end{aligned} \]  

(63)

As \( \varphi \) and \( w \) are canonical variables with the independent variable \( s \), a Hamiltonian \( H_{\varphi w} \) has been used to describe the particle motion:

\[ H_{\varphi w} = -\frac{2\pi}{(\beta_s \gamma_s)^3 \cdot mc^2 \cdot \lambda} \cdot \frac{w^2}{2} - q \cdot E_0 T_s \cdot R(r) \left[ \sin \phi_0 \cdot (\cos \varphi - 1) + \cos \phi_0 \cdot (\sin \varphi - \varphi) \right] \]  

(64)

In the phase space \( (\varphi, w) \), particles follow curves where \( H_{\varphi w} = \text{Cst} \). They are represented in Fig. 5 for on-axis particles. In Fig. 5(a), \( \beta_s \gamma_s = \text{Cst} \), as in Fig. 5(b), an adiabatic acceleration \( (\beta_s \gamma_s \neq \text{Cst}) \) is added and the bucket turns into the well-known golf club shape.

![Fig. 5: Particle trajectories in longitudinal phase-space. (a) \( \beta_s \gamma_s = \text{Cst} \). (b) Adiabatic acceleration: the golf club represents the input acceptance [in red, (1)]. In blue (2) are the trajectories of two particles. They exhibit the damping of the phase oscillation amplitude with acceleration.](image)

A particle entering the cavity after the synchronous particle gets a larger energy gain. A particle entering the cavity in advance (called the early particle) gets a smaller energy gain.
The synchronous phase of the synchronous particle is a stable point situated between $-\pi/2$ and $0$.

The choice of the synchronous phase delimits a phase acceptance:

1. The higher limit $\phi_1$ is the phase where a late particle gets the same energy gain as the synchronous particle:

$$\phi_1 = -\phi_{s0} \quad \Rightarrow \quad \phi_1 = -2 \cdot \phi_{s0} . \quad (65)$$

2. At the lower limit $\phi_2$, the confinement potential equals the potential at the higher limit ($\phi_1$). As the potential is the integral of the force, $\phi_2$ is the phase where the horizontally hatched surface (in Fig. 6) equals the vertically hatched one. It can be calculated from the Hamiltonian given in Eq. (64):

$$H_{qw}(\phi = \phi_2 - \phi_{s0}, w = 0) = H_{qw}(\phi = \phi_1 - \phi_{s0}, w = 0). \quad (66)$$

$\phi_2$ is the solution of

$$\left(\sin \phi_2 - \phi_2 \cos \phi_{s0}\right) + \left(\sin \phi_{s0} - \phi_{s0} \cos \phi_{s0}\right) = 0 . \quad (67)$$

3. The choice of the synchronous phase also determines the energy acceptance $\Delta E$ corresponding to the difference between the potential energy of a particle with a phase $\phi_1$ and the synchronous particle. It can also be calculated from the Hamiltonian given in Eq. (64):

$$H_{qw}(\phi = 0, w = \Delta E) = H_{qw}(\phi = \phi_1 - \phi_{s0}, w = 0) . \quad (68)$$

giving:

\footnote{For positively charged particles, as for negatively charged ones, it depends on convention (is $qE_0 > 0$ or $E_0 > 0$ ?).}
\[
\Delta E = 2 \cdot qE_0 T (\phi_{s0} \cos \phi_{s0} - \sin \phi_{s0}).
\]

\[
\Delta E = \left( \frac{1}{\pi} \left( \frac{\beta_s \gamma_s^3 \cdot \lambda \cdot mc^2}{2 \cdot qE_0 T (\phi_{s0} \cos \phi_{s0} - \sin \phi_{s0})} \right)^{\frac{1}{3}} \right).
\]

The acceptance area in the phase-energy space is called the \textit{bucket}, its limit is called the \textit{separatrix}. The energy acceptance \(\Delta E\) and the phase \(\phi_2\) are represented as a function of the synchronous phase in Fig. 7.

![Fig. 7: Bucket dimensions as a function of the synchronous phase](image)

For small phase amplitude oscillations, Eqs. (63) become:

\[
\begin{align*}
\frac{d\phi}{ds} &= -2\pi \cdot \frac{w}{(\beta_s \gamma_s^3) \cdot mc^2 \cdot \lambda} \\
\frac{dw}{ds} &= q \cdot E_0 T_s \cdot \sin \phi_{s0} \cdot \phi
\end{align*}
\]

giving the second order differential equation of phase evolution:

\[
\frac{d^2 \phi}{ds^2} + \frac{2}{\zeta} \cdot \frac{d\phi}{ds} + k_z^2 \cdot \phi = 0
\]

with:

\[
k_z^2 = \frac{2\pi q \cdot E_0 T_s \cdot \sin(- \phi_{s0})}{(\beta_s \gamma_s^3) \cdot mc^2 \cdot \lambda}.
\]

Here \(k_z\) is the phase advance per metre of the beam core. In periodic structures of period \(L\), \(\sigma_z = k_z L\) is the longitudinal core phase advance per lattice.

\[
\zeta = \frac{2}{3} \cdot \frac{\beta_s \gamma_s}{d(\beta_s \gamma_s)/ds}.
\]

Here \(\zeta\) is the damping length of the core oscillations.

Both \(\zeta\) and the variation of \(k_z\) with \(\beta_s \gamma_s\) contribute to phase oscillation damping with acceleration. The adiabatic damping of the phase amplitude oscillation \(\varphi_a\), defined when the contribution of \(\zeta\) is negligible, can be calculated [7]:

\[
\varphi_a \propto (\beta_s \gamma_s)^{-3/4}.
\]
Liouville’s theorem implies that the energy amplitude oscillation \( w_a \) variation is

\[
w_a \propto \left( \beta_z \gamma_z \right)^{3/4}.
\]

The Hamiltonian in linear force then becomes:

\[
H_{qw} = -\frac{2\pi}{\left( \beta_z \gamma_z \right)^{3/4} \cdot mc \cdot \lambda} \cdot \frac{w^2}{2} + q \cdot E_0 \cdot T_s \cdot \sin \phi_{x0} \cdot \frac{\phi^2}{2}.
\] (76)

The curves where the Hamiltonian is constant are then ellipses.

4.5 Motion in linear force

We have seen that the longitudinal particle motion is basically non-linear, but it can be linearized when the particle phase oscillation amplitude is very small compared to \( \phi \). The transverse forces are much more linear than the longitudinal ones, and the use of the linear focusing force is very close to reality, and can be solved analytically.

4.5.1 Linear transverse forces

In linacs, the main elements used to transport a beam are the cavities and the quadrupoles. Both these elements induce transverse forces.

4.5.1.1 Quadrupoles

In a perfect thick-lens quadrupole the magnetic field is

\[
\begin{align*}
B_x &= G \cdot y \\
B_y &= G \cdot x
\end{align*}
\] (77)

where \( G \) is the quadrupole gradient (in T/m).

With the paraxial approximation and because the magnetic field does not change the particle’s energy, the equations of transverse dynamics in quadrupole are then:

\[
\begin{align*}
\frac{d\gamma_i}{ds} &= \gamma_i \frac{dx}{ds} = \frac{q \cdot G}{mc} \cdot \frac{dx}{ds} \\
\frac{dx}{ds} &= x' = \frac{\beta_x}{\beta_z} \\
\frac{d\gamma_y}{ds} &= \gamma_y \frac{dy}{ds} = \frac{q \cdot G}{mc} \cdot \frac{dy}{ds} \\
\frac{dy}{ds} &= y' = \frac{\beta_y}{\beta_z}
\end{align*}
\] (78)

The transverse perfect quadrupole force is linear. Actually, fringe field and non-perfect hyperbolic poles induce non-linear effects which can generally be neglected at first order in linacs.

4.5.1.2 RF gap

When a particle travels through a cavity, the integration of the effect of the radial electric field and the azimuthal magnetic field can be modeled by a transverse kick, which is linear at second order. This kick modifies the particle transverse momentum:

\[
\Delta (\gamma \beta_y) = -\frac{\pi q E_{TL}}{mc^2 \gamma_z z \cdot \lambda} \cdot \sin \phi \left[ r + O \left( r^3 \right) \right] = \Delta (\gamma \beta_y) \cdot r' + \gamma \beta_y \cdot \Delta r',
\] (79)
with $\beta r^2 = \beta_x^2 + \beta_y^2$. The term in $r'$ shows that the particle transverse oscillation is damped by acceleration in accelerating cavities.

4.5.2 Motion of particle in periodic linear force

At first order, the motion of a particle can be linearized and the motion along all directions can be decoupled. The equation of motion in the $w$ direction ($w$ being $x$, $y$ or $\phi$) is the solution of a second-order equation:

$$\frac{d^2 w}{ds^2} + A_w \frac{d \beta_w}{ds} \frac{dw}{ds} + k_w(s) \cdot w = 0. \quad (80)$$

Here $A_w$ is a constant equal to 1 for $w = x$ or $y$, and 3 for $w = \phi$.

Now, let us consider that the focusing force is periodic with period $S$, i.e. $k_s(s + S) = k_s(s)$. Generally, the damping term given by the acceleration is very small and can be considered as a perturbation:

$$\frac{\{ A_w \frac{d \beta_w}{ds} \frac{dw}{ds} \}}{\gamma \beta_w} \ll \langle k_w(s) \cdot w \rangle_S, \quad (81)$$

where $\langle a \rangle_S$ gives the average value of quantity $a$ over one lattice period. In this assumption, the solution of Eq. (80) is:

$$w(s) = \sqrt{\beta_w(s)} \cdot I_w/\gamma \beta_w \cdot \cos \left[ \psi_w(s - s_0) + \psi_w(s_0) \right], \quad (82)$$

with $\beta_w$ periodic $[\beta_w(s + S) = \beta_w(s)]$, known as the structure beta function, solution of:

$$\frac{d^2 \beta_w}{ds^2} + 2 \cdot k_w(s) \cdot \beta_w - \frac{2}{\beta_w} \left[ 1 + \frac{1}{4} \left( \frac{d \beta_w}{dt} \right)^2 \right], \quad (83)$$

with $I_w/\gamma \beta_z$, known as the Courant–Snyder invariant (which is actually invariant with no acceleration), and $\psi_w$ the particle phase advance, defined as:

$$\psi_w(s) = \int_{s_0}^{s} \frac{ds}{\beta_w(s)}. \quad (84)$$

Particles are turning around periodic ellipses whose equations are:

$$\gamma_w(s) \cdot w^2 + 2 \cdot \alpha_w(s) \cdot w \cdot w' + \beta_w(s) \cdot w'^2 = I_w/\gamma \beta_z, \quad (85)$$

with

$$\alpha_w(s) = -\frac{1}{2} \frac{d \beta_w(s)}{ds}. \quad (86)$$

and

$$\gamma_w(s) = \frac{1 + \left( \alpha_w(s) \right)^2}{\beta_w(s)}. \quad (87)$$

The surface of the ellipses decreases as $1/\gamma \beta_z$ which is close to $1/\gamma \beta$ with the paraxial approximation.

The phase advance per lattice $\sigma_w$ defined as

$$\sigma_w = \psi_w(s + S) - \psi_w(s), \quad (88)$$
gives an idea of how fast the particles turn around the ellipses. The number \( \frac{2\pi}{\sigma_w} \) is the number of lattice periods when the particle has made one turn around the ellipses. One can note that, in linear forces, the phase advance per lattice is the same whatever the particle amplitude.

As an example, have a look at a particle motion along one direction in a FODO channel. In Fig. 8, five FODO lattices have been represented. The particle phase advance per lattice is \( 360^\circ/5 = 72^\circ \). The particle position in 2D phase-space is represented by the red point in four different positions in the lattice. Each line correspond to one position:
- 1st line: middle of focusing quadrupole,
- 2nd line: between focusing and defocusing quadrupoles,
- 3rd line: middle of defocusing quadrupole,
- 4th line: between defocusing and focusing quadrupoles.

One observes that lattice after lattice the particle turns around an ellipse at the same position. The ellipse is different from position to position within the lattice. Its equation is given by (85). It is very important to understand that these ellipses have nothing to do with the beam (no beam has been defined here, just one particle). These ellipses are defined by the transport channel.

To conclude, we should keep in mind that a large number of assumptions have been made to achieve the results. The opportunity of each assumption has to be studied very carefully in practical cases. Nevertheless, the results presented here help to elucidate beam dynamics.
4.6 Beam r.m.s. dimension and Twiss parameters

A bunch is constituted of \( N \) particles. Its dimensions are defined statistically as follows:

- The beam centre of gravity position: 
  \[
  \langle w \rangle = \frac{1}{N} \sum_{i=1,N} w_i .
  \]  
  \( (89) \)

- The beam centre of gravity slope: 
  \[
  \langle w' \rangle = \frac{1}{N} \sum_{i=1,N} w'_i .
  \]  
  \( (90) \)

- The beam r.m.s. size: 
  \[
  \tilde{w} = \sqrt{\left( \langle w \rangle - \langle w \rangle \right)^2} = \frac{1}{N} \sum_{i=1,N} (w_i - \langle w \rangle)^2 .
  \]  
  \( (91) \)

- The beam r.m.s. divergence: 
  \[
  \tilde{w}' = \sqrt{\left( \langle w' \rangle - \langle w' \rangle \right)^2} = \frac{1}{N} \sum_{i=1,N} (w'_i - \langle w' \rangle)^2 .
  \]  
  \( (92) \)

- The beam r.m.s. emittance: 
  \[
  \tilde{\varepsilon}_w = \sqrt{\tilde{w}^2 \tilde{w}'^2 - \left( \langle w \rangle - \langle w \rangle \right) \cdot \left( \langle w' \rangle - \langle w' \rangle \right)^2} .
  \]  
  \( (93) \)

The beam Twiss parameters are then deduced from the beam r.m.s. dimensions:

\[
\tilde{\beta}_w = \frac{\tilde{w}^2}{\tilde{\varepsilon}_w} , \quad \tilde{\gamma}_w = \frac{\tilde{w}'^2}{\tilde{\varepsilon}_w} , \quad \tilde{\alpha}_w = -\frac{\left( \langle w \rangle - \langle w \rangle \right) \cdot \left( \langle w' \rangle - \langle w' \rangle \right)}{\tilde{\varepsilon}_w} .
\]  
(94)

Generally, at least 90% of the particles in the bunch occupy an ellipse of equation \(^{10}\):

\[
\tilde{\gamma}_w \cdot \tilde{w}^2 + 2 \cdot \tilde{\alpha}_w \cdot \tilde{w}' \cdot \tilde{w}^2 + \tilde{\beta}_w \cdot \tilde{w}'^2 = \tilde{\varepsilon}_w .
\]  
(95)

The parameter \( w \) can be \( x, y, z \) or \( \varphi \). The phase-space 2D projections of a beam with \(~100000\) particles are represented in Fig. 9. Ellipses in red correspond to ellipses calculated with Eq. (95). They contain, in this example, 92% of the particles.

4.7 Matched/mismatched beam

A beam is matched when its Twiss parameters at a given position \( s \) correspond to the transport channel periodic Courant–Snyder parameters. In this condition, the same beam phase-space shape is

---

\(^{10}\) If the bunch were uniform, 100% of the particles would occupy this ellipse.
reproduced period after period. The envelope evolution with $s$ is periodic and as smooth as possible. In Fig. 10 the evolution of beams in the same FODO channel as before (upper line) has been represented. In the middle line, the ellipses represent the beams in the phase-space at the focusing quadrupole centre. The dashed-black circle represents a particle motion in this channel. One matched (in continuous red) and two mismatched (in dashed pink and dotted blue) beams have been represented. One particle of each beam has also been represented.

The matched beam ellipse is periodic, as one particle is replaced by another one. Its envelope (last line) is periodic with the lattice period $L$.

The mismatched beam ellipses sweep a bigger area (dashed-black circle) than the beam ellipse surfaces. Their envelope period is greater than the lattice period. Its oscillation is a combination of two oscillations with two different periods: one is the lattice period $L$, the other is $2\pi/\sigma_w \cdot L$, with $\sigma_w$ being the channel phase advance per lattice.

![Fig. 10: Matched and mismatched beam in the FODO channel](image)

When the force is linear (Fig. 11), all particles turn in the phase-space with the same period (i.e. the same phase advance per lattice). The beam phase-space distribution changes lattice after lattice, but its emittance is kept constant.

![Fig. 11: Matched (left) and mismatched (right) beam in linear forces](image)
When the force is non-linear (Fig. 12) [external force or force induced by space-charge (Coulomb interactions between beam particles)], the particle phase-advance per lattice depends on its oscillation amplitude. Beam particles no longer turn all at the same speed, and an apparent emittance growth is observed\(^\text{11}\). This effect is known as beam filamentation (Fig. 13). After a long time (many particle betatron periods), the phase-space swept by the beam is completely full of particles. The apparent emittance is higher.

---

**Fig. 12**: Matched (left) and mismatched (right) beam in non-linear forces

**Fig. 13**: Filamentation of mismatched beam in non-linear force

\(^{11}\) Even if the phase-space area occupied by the particle is constant (Liouville’s theorem applies).
5 CONCLUSION

This paper is a short introduction with basic notions on linacs. A better understanding cannot be obtained without tackling subjects like the existing structures, the RF control, the space-charge effects, or the resonances. Motivated students are strongly advised to read Tom Wangler's book [2].

ACKNOWLEDGEMENTS

Many thanks to Tom Wangler for his wonderful and very useful book on linacs, his encouragement and to my colleagues Romuald Duperrier and Didier Uriot for their help.

REFERENCES


Linear imperfections

O. Brüning, CERN
CERN, Geneva, Switzerland

Abstract
The course consists of a one hour lecture followed by one hour of guided studies and one hour of discussions. It features a short summary of the equation of motion in a circular storage ring, the introduction of the sine- and cosine- like solutions of the unperturbed linear system, a discussion of linear resonances and the impact of dipole and quadrupole type field errors and presents various examples from the operation of existing storage rings. The final part of the course addresses options for a correction of the closed orbit in a storage ring.

1 Summary of the equation of motion

In the following we present a brief summary of the main aspects of the equation of motion in a storage ring. Our goal here is not to present a complete review of the linear theory of accelerator optics but rather to focus on those aspects that are required for a discussion of linear imperfections in a storage ring. For a more detailed description of the theory we refer the reader either to the proceedings of the general CERN accelerator school [1], the excellent summary article by M. Sands [2] or the original work by Courant and Snyder [3].

1.1 Lorentz force
Dipole fields guide the particle trajectory in a storage ring onto a circular orbit. The average trajectory of the particles over many revolutions in the storage ring is referred to as the closed orbit of the storage ring. The equation of motion of a charged particle in a magnetic field is governed by the Lorentz force:

\[ m \ddot{\vec{r}} = \vec{F}_L = q \cdot (\vec{E} + \vec{v} \times \vec{B}) ; \quad \vec{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} ; \quad \vec{v} = \frac{d\vec{r}}{dt} \] (1)

where \( m \) is the particle mass, \( q \) its charge, \( \times \) denotes the vector product and \( \vec{r}, \vec{v}, \vec{E} \) and \( \vec{B} \) the vector presentations of the particle position, velocity and the electric and magnetic fields in the storage ring respectively. \(^1\) The force due to constant magnetic field lines is always perpendicular to the velocity vector and thus, can not change the energy of the particles. A storage ring therefore employs magnetic field lines only for a deflection of the particle trajectories.

1.2 Lorentz force for dipole fields
In a uniform dipole field \( B_0 \) the Lorentz force is balanced by the centrifugal force. Assuming that the velocity vector lies perpendicular to the magnetic field lines the equilibrium condition can be written as:

\[ q \cdot v \cdot B_0 = \frac{m \cdot v^2}{\rho} , \] (2)

where \( \rho \) is the equilibrium radius of curvature. Solving Eq. (2) for \( \rho \) one obtains the cyclotron equation:

\[ \rho = \frac{m \cdot v}{q \cdot B_0} . \] (3)

\(^1\)In general we will identify vector components with a vector arrow and matrices with an underscore in the following.
The deflection angle due to a single dipole magnet can be written as
\[ \chi = \frac{l}{\rho} \]  
(4)
where \( l \) is the magnet length. The force acting on the particle depends not only on the dipole field strength but also on the ratio of the particle charge and momentum. The quantity
\[ B\rho = \frac{p}{q} \]  
(5)
is referred to as the **beam rigidity** and highlights the impact of the particle charge and momentum in the force term. For the following analysis it is convenient to introduce a normalized dipole strength that combines the beam rigidity and the dipole field strength to an effective force term:
\[ k_0 = \frac{q \cdot B}{p} = \frac{1}{\rho}. \]  
(6)
The normalized dipole strength has the units \([k_0] = 1/m\) and can be conveniently written as
\[ k_0 = 0.2998 \frac{B[T]}{p[GeV/c]}. \]  
(7)

### 1.3 Rotating coordinate system in a storage ring

The particle motion in a storage ring is commonly described within a coordinate system that moves with a reference particle along the design closed orbit of the storage ring. Fig. 1 shows a schematic sketch for such a rotating Cartesian coordinate system. The axis tangential to the design orbit of the storage ring is denoted as the \( s \) coordinate and points into the direction of the rotation. The axis pointing out-wards in the radial direction is denoted by \( x \) and the axis pointing up-wards by \( y \). This labeling generates a right handed coordinate system and corresponds to the conventions used in the control room of a storage ring.\(^2\) In the following we will assume that the storage ring lies in one fixed plane (\( \rightarrow \) the magnetic field lines of the main bending magnets point into the vertical direction of the rotating coordinate system).\(^3\)

Looking at the particle motion from within the moving coordinate system we have to add explicitly the centrifugal force to the equation of motion. Assuming a uniform magnetic dipole field with field lines in the vertical direction the equation of motion in the horizontal plane become
\[ m \frac{d^2 x}{dt^2} = \frac{m \cdot v^2}{r} - q \cdot v \cdot B_0, \]  
(8)
with
\[ r = \rho + x. \]  
(9)

As the particles circulate in the storage ring they encounter at every turn the same sequence of elements. The equations of motion describing the oscillation around the closed orbit have therefore a fundamental periodicity corresponding to one revolution in the storage ring. In order to highlight this

\(^2\)It should be noted here that magnet builders often use a different coordinate system: looking into the direction of the beam the horizontal axis points to the right (and thus in-wards in our moving accelerator coordinate system) and the vertical axis up-wards (identical to the moving accelerator coordinate system). This choice of axis orientation follows the conventions used in the complex plane. However, while this choice is well adopted for the magnet design it creates different signs for some of the magnet field errors and good care must be taken in order to transfer the magnet field errors from one system to the other.

\(^3\)In practice, however, this is not always the case. For example, the geological boundary conditions of an underground storage ring can impose a storage ring structure that features a variation of the bending field direction along the storage ring. Such a variation of the bending field orientation can couple the particle motions in the horizontal and vertical plane of the rotating coordinate system and the effect of coupling must be very carefully analyzed in this case.
Fig. 1: The rotating coordinate system used for describing the particle motion inside a storage ring.

Fig. 2: The effect of geometric focusing in an ideal machine with a homogeneous dipole field all the way around the storage ring. The picture on the left shows the trajectory response to injection angle errors seen from above the storage ring and the picture on the right shows the same cases seen from within the rotating coordinate system. The trajectory response intersects the ideal closed orbit twice on its passage around the storage ring and appears to perform one complete oscillation around the closed orbit when looked at from within the rotating coordinate system. The points $P_1$ and $P_2$ indicate the positions where the trajectory response intersects the design closed orbit.

Periodicity it is convenient to replace the time variable $t$ in the equations of motion by the arc length $s$ along the design orbit in the storage ring. Assuming the particles are ultra relativistic ($v \approx c \rightarrow d^2s/dt^2 = 0$) we have

\[
\begin{align*}
\frac{ds}{dt} &= vt; \\
\frac{dx}{ds} &= \frac{dx}{dt} \frac{dt}{ds} \\
\frac{d^2x}{ds^2} &= \frac{d^2x}{dt^2} \left(\frac{ds}{dt}\right)^2 \\
\frac{d^2x}{dt^2} &= \frac{d^2x}{ds^2} v^2.
\end{align*}
\]
\( v = ds/dt \) is the particle speed in the tangential direction of the closed orbit and can be written as

\[
\frac{ds}{dt} = L \cdot f_{\text{rev}}
\]  

(11)

where \( f_{\text{rev}} \) is the revolution frequency and \( L \) the circumference of the design orbit. The derivatives with respect to \( s \) are often written as

\[
x' = \frac{dx}{ds}
\]  

(12)

\[
x'' = \frac{d^2x}{ds^2}
\]  

(13)

Using the path length \( s \) as the free parameter one obtains for the canonical momentum \( x' = dx/ds \):

\[
\frac{dx}{dt} = \frac{ds}{dt} \cdot \frac{dx}{ds} \rightarrow \frac{dx}{ds} = \frac{v_x}{v_s} \quad \text{or} \quad \frac{dx}{ds} = \frac{p_x}{p_s}
\]  

(14)

where \( \chi \) is the angle between the total momentum vector and the \( s \) coordinate. In other words, the canonical momentum \( dx/ds \) is the ratio of the transverse and longitudinal momentum and corresponds to an angle. We will come back to this later when we discuss the effect of dipole perturbations in a storage ring.

Equation (8) can now be written with \( s \) as the new free parameter:

\[
x'' = \frac{1}{r} - \frac{F_{x,\text{Lorentz}}}{p \cdot v},
\]  

(15)

with \( p = m \cdot v \). In the case of a uniform dipole field with field components in the vertical axis of the rotating coordinate system and perpendicular to the direction of motion one obtains for the equation of motion in the horizontal plane:

\[
x'' = \frac{1}{r} - \frac{q \cdot B}{p}.
\]  

(16)

Assuming further that the dipole field and the momentum can be written as the sum of the design values \( (B_0 \text{ and } p_0) \) and a small error component \( (B = B_0 + \Delta B \text{ and } p = p_0 + \Delta p) \) we can expand the equation of motion to first order in the error components. Using:

\[
\frac{1}{1 + \epsilon} = 1 - \epsilon + O(\epsilon^2)
\]  

(17)

we can expand the momentum dependence of the denominator around the design momentum:

\[
x'' = \frac{1}{r} - \frac{q \cdot B_0}{p_0} \cdot (1 - \delta p) - \frac{q \Delta B}{p_0},
\]  

(18)

where we have introduced the notation

\[
\delta p = \frac{\Delta p}{p_0}.
\]  

(19)

Expanding next the radial position \( r \) for small deviation around \( \rho \) one obtains

\[
r = \rho + x \approx \rho \cdot (1 + \frac{x}{\rho})
\]  

(20)

\[ \rightarrow \frac{1}{r} \approx \frac{1}{\rho} \cdot \left(1 - \frac{x}{\rho}\right).
\]  

(21)

Substituting Eqs. (3) and (21) into Eq. (18) one finally obtains

\[
x'' = -\frac{1}{\rho^2} \cdot x + \frac{1}{\rho} \cdot \frac{\delta p}{p_0} - \Delta k_0, \quad \text{with} \quad \Delta k_0 = \frac{q \cdot \Delta B}{p_0}.
\]  

(22)
We will treat the effect of the field and momentum errors in Section 2 and assume for the moment that equation of motion does not have any error components.

Equation (22) shows that force always acts on the particle with the opposite sign of the particle amplitude. The force therefore always deflects the particle motion back to the ideal closed orbit and the horizontal particle motion inside a uniform dipole field without errors is stable. This stabilizing effect is called **geometric focusing**. The effect of geometric focusing can be nicely illustrated by looking at a machine that features a continuous homogeneous dipole field all the way around the storage ring [1]. The closed orbit of such a storage ring is a perfect circle with a radius given by the cyclotron equation (3). The trajectory of a particle that enters the storage ring with initial conditions that lie not on the closed orbit still follows a circle with a radius given by the cyclotron equation (3). However, the circle of the trajectory response is rotated with respect to the ideal closed orbit. The trajectory comes back to the same initial condition after one revolution around the storage ring and intersects the closed orbit twice during its passage around the machine. Observed from within the rotating coordinate system the motion appears to perform oscillations around the design closed orbit. The deflection of the dipole field therefore look like a focusing force within the rotating coordinate system. Figure 2 illustrates this effect for two different initial errors. In both cases the initial position of the particle lies on the closed orbit but the particle momentum is not tangential to the closed orbit. In one case the initial conditions have a positive and in the other a negative error in the angle between the particle momentum and tangent of the closed orbit.

### 1.4 Lorentz force for a quadrupole field

Equation (22) shows that the stabilizing effect decreases with increasing radius of curvature and the effect can therefore be negligible for large storage rings. Furthermore, the above mechanism provides only focusing in one plane. Additional quadrupole field components are required to obtain also a focusing in the orthogonal plane. While the dipole fields are uniform (at least in an ideal storage ring) and are independent from the transverse particle coordinates inside the magnet the quadrupole field depends linearly on the distance from the magnet center at which the particle passes through the magnet. The magnetic field of a quadrupole magnet is characterized by its gradient:

\[
g = \frac{\partial B_y}{\partial x}
\]

and the magnetic field experienced by the particle during its passage through the quadrupole magnet is given by

\[
B_x = +g \cdot y, \quad B_y = +g \cdot x
\]

where \(x\) and \(y\) are the horizontal and vertical offsets from the magnet center respectively. The plus sign in Eq. (24) comes from our choice of the rotating coordinate system. Figure 3 shows a schematic sketch of a quadrupole cross section and the corresponding magnetic field lines using the rotating coordinate system of a storage ring as a reference system. However, magnet designers commonly present the magnetic field lines in a coordinate system that resembles the complex plane: the horizontal \(x\) axis goes to the right and the vertical \(y\) axis upwards. The horizontal axis of this coordinate system has the opposite sign with respect to the rotating coordinate system. Expressing the quadrupole field lines within this 'magnet' coordinate system results in a sign change in Eq. (24) and one has to be careful to specify the underlying coordinate system whenever one talks about a positive or negative quadrupole gradient (we come back to this point later in this section). Here we choose the rotating coordinate so that the magnetic field presentation is well adopted for discussing the equations of motion in a storage ring.

In analogy to the normalized dipole field strength we introduce a normalized quadrupole field strength:

\[
k_1 = \frac{qq}{p},
\]
The normalized quadrupole gradient has the units $[k_1] = 1/m^2$ and can be conveniently written as

$$k_1 = 0.2998 \cdot \frac{g[T/m]}{p[GeV/c]}.$$  \hfill (26)

Inserting the expression of the magnetic quadrupole field into Eq. (1) one obtains for the Lorentz force acting on a positively charged particle inside a quadrupole magnet

$$F_x(Quadrupole) = -k_1 \cdot x \cdot v \cdot p, \quad F_y(Quadrupole) = +k_1 \cdot y \cdot v \cdot p.$$  \hfill (27)

Using the $s$ variable as the free parameter in the Lorentz force one obtains for the new variables the following equations of motion for a quadrupole magnet with positive gradient:

$$x'' = -k_1 \cdot x, \quad y'' = +k_1 \cdot y.$$  \hfill (28)

For a positive quadrupole gradient the horizontal force acts always on the particle motion with the opposite sign as the particle position. As for the the case of the dipole field the particle motion is therefore stable. The vertical force, however, acts on the particle motion with the same sign as the particle position and deflects the particle trajectory away from the ideal closed orbit. The particle motion is therefore unstable in the vertical plane. An overall focusing of the particle motion in both transverse planes requires either a negative quadrupole gradient that focuses the particle motion in the vertical plane and generates in the horizontal plane a defocusing force which is smaller than the focusing force due the dipole fields (principle of weak focusing) or an alternating sequence of quadrupole polarities (principle of strong focusing). Most modern storage rings use the principle of strong quadrupole focusing. The reference trajectory inside a quadrupole magnet is a straight line and the closed orbit in a storage ring is a collection of curved passages in the dipole magnets and straight sections inside the quadrupole magnets and drift sections (regions without any active magnetic elements).

The quadrupole magnets 'focus' the particles in a storage ring and, in analogy to optical lenses, one can define a 'quadrupole focal length'

$$f = \frac{1}{lk_1}.$$  \hfill (29)
where $l$ is the quadrupole magnet length. We will come back to the concept of a 'quadrupole focal length' when we discuss the transport matrices for quadrupole magnets.

### 1.5 Hills equation and Floquet theorem

In the following we assume that the motion in the horizontal and vertical planes are not coupled and that we can solve the equations of motion in the two transverse planes independently. This is only true as long as we consider dipole magnets with magnetic field lines in the vertical direction and quadrupole magnets that are not rotated with respect to the schematic layout in Fig. 3. For other magnetic field configurations the particle motion in the horizontal and vertical directions is coupled. The problem of coupling is beyond the scope of this course. However, it is very important for the design and operation of a storage ring and deserves a full curse on its own. We refer the interested reader for a detailed treatment of this special topic to the excellent papers by F. Willeke and G. Ripken [4] and D. A. Edwards and L.C. Teng [5]. Omitting the effect of coupling we can limit our treatment in the following to one transverse plane only. Choosing the horizontal plane the equation of motion can now be written in the form of a second order linear differential equation with periodic coefficients:

$$ x'' + K(s)x = \frac{1}{\rho} \cdot \frac{\Delta p}{p_0} \quad (30) $$

with $K(s + L) = K(s)$. The periodic coefficient $K(s)$ depends on the type of magnet the particle is passing through (see Eqs. (22) and (28)):

$$ K(s) = \begin{cases} 
0 & \text{drift space} \\
\frac{1}{\rho} & \text{dipole field} \\
k_1 & \text{quadrupole field} 
\end{cases} \quad (31) $$

For the moment we assume that the particle has its design momentum and set the inhomogeneous part of Eq. (30) to zero. We come to the solution of the inhomogeneous equations of motion in Section 2. Equations of the form (30) are referred to as 'C-W' or 'Hills' equation. The Hills equation has its origin in celestial mechanics where it has been used by Hill (1878) and Clohessy and Wiltshire (1960) for studying the planetary and satellite motion in our solar system.

For a constant coefficient ($K(s) = \text{const} \neq 0$) the Hills equation corresponds to the equations of motion of a harmonic oscillator and has the solutions

$$ x_a(s) = A \cdot e^{+i(\sqrt{\beta(s)} + \phi_0)} \quad \text{and} \quad x_b(s) = A \cdot e^{-i(\sqrt{\beta(s)} + \phi_0)} \quad (32) $$

For a vanishing coefficient ($K(s) = 0$) the Hills equation can be written as

$$ x'' = 0 \quad (33) $$

and the solution becomes

$$ x_0(s) = A + B \cdot s \quad (34) $$

$A$ and $B$ are arbitrary constants with units $[A] = m$ and $[B] = 1$ in order to generate the proper units for the solutions. The Floquet theorem states that the solutions of the Hills equation with non-constant coefficient $K(s)$ can also be written in terms of sine and cosine functions. But this time the amplitude and the derivative of the phase term are not constants but are functions of the $s$ coordinate:

$$ x_{Ha}(s) = \sqrt{A} \cdot \sqrt{\beta(s)} \cdot e^{+i\phi(s)} \quad (35) $$

$$ x_{Hb}(s) = \sqrt{A} \cdot \sqrt{\beta(s)} \cdot e^{-i\phi(s)} \quad (36) $$
The $s$ dependent amplitude term $\beta(s)$ is called the optical beta function and has the unit meter ($[\beta] = m$). The phase term $\phi(s)$ depends on the beta function via

$$\phi(s) = \int_{s_0}^{s} \frac{1}{\beta(\tau)} d\tau + \phi_0. \quad (37)$$

Substituting Eq. (35) or Eq. (36) back into the equation of motion (30) one obtains a nonlinear differential equation for the beta function:

$$\frac{1}{2} \beta(s) \cdot \beta''(s) - \frac{1}{4} \beta'^2(s) + K(s) \cdot \beta^2(s) = 1. \quad (38)$$

Equation (38) indicates that a change in the focusing strength $K(s)$ changes the beta function and via Eq. (37) the phase advance in the storage ring. Using Eq. (37) the phase advance for one full revolution in the storage ring can be written as

$$2\pi Q = \int_{s_0}^{s_0+L} \frac{1}{\beta(\tau)} d\tau$$

where we have introduced the tune 'Q' in a storage as the number of oscillations over one turn in the storage ring. For a constant coefficient $K(s)$ we get

$$K(s) \cdot \beta^2(s) = 1 \rightarrow \beta = \frac{1}{\sqrt{K}} \quad (39)$$

and Eqs. (39) and (39) state that a storage ring with weak focusing (small $K(s)$) features large beta functions and a small betatron oscillation frequency (small 'tune' value).

The solutions (35) and (36) can be combined to form two independent solutions for $x(s)$:

$$x_{H1}(s) = -\frac{i}{2} \cdot (x_{Ha}(s) - x_{Hb}(s)) = \sqrt{A} \cdot \sqrt{\beta(s)} \cdot \sin(\phi(s) - \phi_0) \quad (40)$$

$$x_{H2}(s) = \frac{1}{2} \cdot (x_{Ha}(s) + x_{Hb}(s)) = \sqrt{A} \cdot \sqrt{\beta(s)} \cdot \cos(\phi(s) - \phi_0) \quad (41)$$

### 1.6 Sine and cosine like solutions and the one-turn map

In the following it will be useful to have a transformation that propagates any initial conditions of a trajectory to another $s$ coordinate of the storage ring. To this end it is useful to define the functions

$$\alpha(s) = -\frac{1}{2} \frac{d}{ds} \beta(s) \quad (42)$$

$$\gamma(s) = \frac{1 + \alpha^2(s)}{\beta(s)}. \quad (43)$$

The functions $\beta(s), \alpha(s)$ and $\gamma(s)$ are referred to as the 'twiss parameters' and are periodic function of $s$ satisfying

$$\alpha(s + L) = \alpha(s); \quad \beta(s + L) = \beta(s); \quad \gamma(s + L) = \gamma(s). \quad (44)$$

For the construction of a transfer matrix we rewrite the second order differential equation of motion as a system of two first order differential equations. Introducing the vector notation

$$\vec{z} = \begin{pmatrix} x \\ x' \end{pmatrix} \quad (45)$$
one obtains for the equations of motion
\[
\frac{d\vec{z}}{ds} + \begin{pmatrix}
0 & -1 \\
K(s) & 0
\end{pmatrix} \cdot \vec{z} = 0.
\] (46)

The solutions (40) and (41) can be written as
\[
\vec{z}_{H1}(s) = \sqrt{A} \cdot \begin{pmatrix}
\sqrt{\beta(s)} \sin (\phi(s) - \phi_0) \\
[\cos (\phi(s) - \phi_0) - \alpha(s) \sin (\phi(s) - \phi_0)] / \sqrt{\beta(s)}
\end{pmatrix}
\] (47)
\[
\vec{z}_{H2}(s) = \sqrt{A} \cdot \begin{pmatrix}
\sqrt{\beta(s)} \cos (\phi(s) - \phi_0) \\
-[-\sin (\phi(s) - \phi_0) + \alpha(s) \cos (\phi(s) - \phi_0)] / \sqrt{\beta(s)}
\end{pmatrix}.
\] (48)

The solutions for the particle position and momentum can be combined with the twiss functions to form an invariant of the motion:
\[
A = \gamma x^2 + 2 \alpha x' x + \beta x'^2.
\] (49)

Equation (49) is the equation of an ellipse in the \(x x'\) plane and the particle motion lies on an ellipse in phase space. For \(\alpha = 0\) the axes of the ellipse are parallel to the \(x\) and \(x'\) axes. The area of this ellipse is an invariant of the particle motion and is given by
\[
\text{area} = \pi \cdot a \cdot b = \pi \cdot A.
\] (50)

For a particle ensemble one can define a beam invariant along the storage ring by taking the average of all single particle invariants in the distribution. This beam invariant is referred to as the 'beam emittance' \(\epsilon\):
\[
\epsilon = \langle A \rangle_{\text{distribution}}.
\] (51)

Sometimes, the beam emittance is also defined as the area of the smallest ellipse in phase space that contains a given percentage of the total beam distribution (e.g. 95.5%) and an emittance value should best be quoted together with its definition.

At any position in the storage ring the RMS beam size of the distribution can now be expressed in terms of the optics \(\beta\)-function and the beam emittance:
\[
\sigma(s) = \sqrt{\epsilon \cdot \beta(s)}.
\] (52)

Similarly, the beam divergence is given by
\[
\sigma'(s) = \sqrt{\epsilon / \beta(s)}.
\] (53)

Next we define linear combinations of the \(z_{H1}\) and \(z_{H2}\) solutions
\[
\vec{S} = \begin{pmatrix} S \\ S' \end{pmatrix} = a \cdot z_{H1}(s) + b \cdot z_{H2}(s)
\] (54)
\[
\vec{C} = \begin{pmatrix} C \\ C' \end{pmatrix} = c \cdot z_{H1}(s) + d \cdot z_{H2}(s)
\] (55)
such that they satisfy the initial boundary conditions
\[
S(s_0) = 0 \quad C(s_0) = 1
\] (56)
\[
S'(s_0) = 1 \quad C'(s_0) = 0.
\] Setting \(s = s_0\) in Eqs. (54) and (55) and using the initial conditions in Eq. (56) one obtains for the coefficients:
\[
a = \sqrt{\beta(s_0)} \quad b = 0
\] (57)
\[ c = \frac{\alpha(s_0)}{\sqrt{\beta(s_0)}} \quad d = \frac{1}{\sqrt{\beta(s_0)}}. \]

The solutions \( \vec{S} \) and \( \vec{C} \) can now be written as
\[ \vec{S} = \left( \sqrt{\beta_0 \beta} \cdot \sin (\Delta \phi) \sqrt{\beta_0} \cdot \left( \cos (\Delta \phi) \right) \right) \]
\[ \vec{C} = \left( \sqrt{\beta} \cdot \left[ \cos (\Delta \phi) + \alpha_0 \cdot \sin (\Delta \phi) \right] / \sqrt{\beta_0} \right) \]

where we have introduced the notations
\[ \beta_0 = \beta(s_0) \quad \alpha_0 = \alpha(s_0) \quad \Delta \phi = \phi(s) - \phi(s_0) \]

and where we have suppressed all \( s \) arguments for \( \beta(s) \) and \( \alpha(s) \).

The solutions (58) and (59) define again a complete set of solutions to the homogeneous Hill’s equation and are called the 'sine' and 'cosine' like solutions of the Hill’s equation. Any solution can be expressed as a linear combination of the 'sine' and 'cosine' like solutions and defining a matrix \( \Phi \) with the 'sine' and 'cosine' like solutions as columns
\[ \Phi(s_0, s) = \begin{pmatrix} C(s) & S(s) \\ C'(s) & S'(s) \end{pmatrix} \]

one can write the trajectory values at any location in the storage ring as a function of the initial conditions:
\[ \vec{z}(s) = \Phi(s_0, s) \cdot \vec{z}(s_0). \]

An interesting special case is the transformation matrix for one revolution, the so called 'one turn map'. Setting \( s = s_0 + L \) \((\Rightarrow \Delta \phi = 2\pi Q)\) in Eq. (62) we obtain:
\[ \Phi(s_0, s_0 + L) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \cos (2\pi Q) + \begin{pmatrix} \alpha(s_0) & \beta(s_0) \\ -\gamma(s_0) & -\alpha(s_0) \end{pmatrix} \cdot \sin (2\pi Q). \]

From Eq. (63) it follows that the trace of the one-turn map is given by
\[ \text{Tr}(\Phi(s_0, s_0 + L)) = 2 \cdot \cos (2\pi Q). \]

We will use this result later when we calculate the closed orbit response to dipole field errors.

### 1.7 Transfer matrices for special elements

The 'one turn' map can be either constructed with the help of the 'sine' and 'cosine' like functions or by multiplication of the individual transport matrices for the elements along the storage ring. While the first options requires already a knowledge of the Hill’s solutions the second requires only a knowledge of the transfer functions for the individual elements in the storage ring. This method is, in fact, used by most numerical codes for the calculation of the twiss parameters and the closed orbit in a storage ring. For example, we have seen in the previous section how the linear matrices can be used for calculating the 'twiss parameters' along a storage ring. In the next Sections we will use this method also for estimating the effect of quadrupole perturbations on the machine tune and the optical \( \beta \)-function. We will therefore briefly summarize the transport matrices for some of the most common linear elements in a storage ring.

In all cases the transfer matrix can be obtained via the solution (32) of Hill’s equation (30) with constant coefficient and remembering
\[ x'(s) = \frac{dx}{ds}. \]
1.7.1 Transfer matrix for a drift section

For a drift section we have

\[ K(s) = 0 \]  

(66)

and the solution of Hill’s equation is given by Eq. (34). The solution at the end of the drift section can thus be written as a function of the initial conditions at the beginning of the drift section \((x_{ini} \text{ and } x'_{ini})\):

\[
\begin{align*}
    x_{end} &= x_{ini} + l \cdot x'_{ini} \\
    x'_{end} &= x'_{ini}
\end{align*}
\]  

(67)

where \(l\) is the length of the drift section. In matrix form the propagation can be written as

\[
\vec{z}_{end} = \begin{pmatrix} 1 & l \\ 0 & 1 \end{pmatrix} \cdot \vec{z}_{ini}
\]  

(68)

where we have used the vector notation defined in Eq. (45).

The extended transport matrix is easily obtained by observing that the particle momentum does not change over a drift space and that the final coordinates are entirely determined by the initial coordinates \(x_{ini}\) and \(x'_{ini}\) and do not depend on the momentum error \(\delta p\):

\[
\begin{pmatrix}
    x_{end} \\
    x'_{end} \\
    \delta p_{end}
\end{pmatrix}
= \begin{pmatrix} 1 & l & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}
\cdot
\begin{pmatrix}
    x_{ini} \\
    x'_{ini} \\
    \delta p_{ini}
\end{pmatrix}
\]  

(69)

1.7.2 Transfer matrix for a section with constant dipole field

For a constant dipole field we have

\[ K(s) = \frac{1}{\rho^2} \]  

(70)

in the plane orthogonal to the dipole field direction and the solution of Hill’s equation is given by Eq. (32):

\[
\begin{align*}
    x(s) &= a \cdot \cos \left( \frac{s}{\rho} \right) + b \cdot \sin \left( \frac{s}{\rho} \right) \\
    x'(s) &= -a \cdot \sin \left( \frac{s}{\rho} \right) \cdot \frac{1}{\rho} + b \cdot \cos \left( \frac{s}{\rho} \right) \cdot \frac{1}{\rho}
\end{align*}
\]  

(71)\hspace{1cm}(72)

For the initial conditions at the entrance of the dipole field we have thus:

\[
\begin{align*}
    x_{ini} &= a \\
    x'_{ini} &= \frac{b}{\rho}
\end{align*}
\]  

(73)\hspace{1cm}(74)

and we can write the solutions at the end of the dipole field as

\[
\begin{align*}
    x_{end} &= \cos (\chi) \cdot x_{ini} + \rho \cdot \sin (\chi) \cdot x'_{ini} \\
    x'_{end} &= -\frac{1}{\rho} \cdot \sin (\chi) \cdot x_{ini} + \cos (\chi) \cdot x'_{ini}
\end{align*}
\]  

(75)\hspace{1cm}(76)

where we have used the definition (4) for the integrated deflection angle \(\chi\) of a dipole magnet. In matrix form the propagation can now be written as

\[
\vec{z}_{end} = \begin{pmatrix} \cos (\chi) & \rho \cdot \sin (\chi) \\ -\frac{1}{\rho} \cdot \sin (\chi) & \cos (\chi) \end{pmatrix} \cdot \vec{z}_{ini}
\]  

(77)

In the plane parallel to the dipole field direction we have the same equation of motion as for a drift space and we obtain again the solution in Eq. (68).
The extended transport matrix can be obtained by observing that the trajectory inside a dipole magnet depends on the particle momentum. Figure 4 shows schematically the particle trajectory in a dipole magnet (top view with the magnet being indicated by the red shape). A particle entering the magnet from the left (blue arrow) on a straight trajectory will follow a curved trajectory inside the magnet. The height $h$ and the segment $k$ of the dipole deflection can be obtained via geometric analysis of Fig. 4 and are given by

\begin{align}
  k &= \rho \cdot \sin \chi, \\
  h &= \rho \cdot (1 - \cos \chi),
\end{align}

where the radius of curvature is given by Eq. (3) and the angle is given by the dipole deflection angle in Eq. (4). For a particle having the design momentum the segment $k$ and the height $h$ correspond to the change of the origin of the rotating coordinate system and the particle position within the rotating coordinate system does not change during the passage through the dipole magnet. A particle having a momentum error will exit the magnet with different coordinates compared to the on-momentum particle. The difference in the particle position is given by the change in the height $h$. The momentum dependence of the radius of curvature can be written as

\begin{equation}
  \rho = \rho_0 \cdot \left(1 + \frac{\Delta p}{p_0}\right)
\end{equation}

and, using Eq. (79), we get for the change in the particle position as a function of the momentum error

\begin{equation}
  \Delta h = \Delta x = \rho_0 \cdot (1 - \cos \chi) \cdot \frac{\Delta p}{p_0}.
\end{equation}

The change in the canonical momentum can be evaluated by observing $s = \rho \cdot \chi$ and

\begin{align}
  x' &= \frac{dx}{ds} \\
  &= \frac{1}{\rho} \cdot \frac{dx}{d\chi}.
\end{align}

Taking the derivative of Eq. (81) with respect to $\chi$ and inserting the result into Eq. (82) yields for the change in the canonical momentum as a function of the momentum error

\begin{equation}
  \Delta x' = \sin \chi \cdot \frac{\Delta p}{p_0}.
\end{equation}
The particle momentum does not change within a dipole magnet and we can write the extended dipole transport matrix as:

\[
\begin{pmatrix}
    x_{\text{end}} \\
    y_{\text{end}} \\
    \delta_{\text{pend}}
\end{pmatrix}
= \begin{pmatrix}
    \cos (\chi) & \rho \cdot \sin (\chi) & \rho_0 \cdot (1 - \cos \chi) \\
    -\frac{1}{\rho} \cdot \sin (\chi) & \cos (\chi) & \sin \chi \\
    0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
    x_{\text{ini}} \\
    y_{\text{ini}} \\
    \delta_{\text{pini}}
\end{pmatrix}.
\] (84)

### 1.7.3 Thin lens approximation

Next we want to briefly introduce the 'thin lens' approximation of the dipole map. The interesting feature of the 'thin lens' approximation is that the particle position does not change during the map. Furthermore, it is generally not possible to calculate the map of a non-linear magnetic field in a closed form. But is is always possible to write down a thin lens approximation for deflection due to any magnetic field. Setting

\[
\Delta x = 0 \quad \text{(85)}
\]
\[
\Delta y' = \frac{F_{\text{Lorentz}}}{v \cdot p_0} \cdot l, \quad \text{(86)}
\]

where \( l \) is the length of the non-linear element and \( F_{\text{Lorentz}} \) the Lorentz force experienced by a moving, charged particle inside the field, one can write a 'thin lens' approximation for an arbitrary magnetic field perturbation. Reducing the length of the dipole magnet while keeping the total deflection angle constant we can obtain a 'thin lens' approximation for the dipole transport map. Using the definition of the deflection angle in Eq. (4) and using the approximation

\[
\sin \chi \approx \chi \quad \text{(87)}
\]

for small arguments of the sin function we obtain for the 'thin lens' approximation of a dipole magnet:

\[
\vec{z}_{\text{end}} = \begin{pmatrix}
    1 & 0 \\
    -\frac{l}{\rho^2} & 1
\end{pmatrix} \cdot \vec{z}_{\text{ini}}. \quad \text{(88)}
\]

### 1.7.4 Transfer matrix for a section with constant quadrupole field

For a constant dipole field we have

\[
K_x(s) = k_1; \quad K_y(s) = -k_1 \quad \text{(89)}
\]

in the horizontal and vertical planes respectively. The solution of Hill’s equation is given by Eq. (32) and depending on the sign of the normalized quadrupole one obtains:

\[
x(s) = a \cdot \cos (s \sqrt{|k_1|}) + b \cdot \sin (s \sqrt{|k_1|}) \\
x'(s) = -a \cdot \sin (s \sqrt{|k_1|}) \cdot \sqrt{|k_1|} + b \cdot \cos (s \sqrt{|k_1|}) \cdot \sqrt{|k_1|} \quad \text{(90)}
\]

for a positive quadrupole gradient and

\[
x(s) = a \cdot \cosh (s \sqrt{|k_1|}) + b \cdot \sinh (s \sqrt{|k_1|}) \\
x'(s) = a \cdot \sinh (s \sqrt{|k_1|}) \cdot \sqrt{|k_1|} + b \cdot \cosh (s \sqrt{|k_1|}) \cdot \sqrt{|k_1|} \quad \text{(92)}
\]

for a negative quadrupole gradient. For the initial conditions at the entrance of the quadrupole field we have in both cases

\[
x_{\text{ini}} = a \quad \text{(94)}
\]
\[
x_{\text{ini}} = b \cdot \sqrt{|k_1|} \quad \text{(95)}
\]
and we can write the solutions at the end of the quadrupole field as
\[
x_{\text{end}} = \cos (l \sqrt{|k_1|}) \cdot x_{\text{ini}} + \frac{1}{\sqrt{|k_1|}} \cdot \sin (l \sqrt{|k_1|}) \cdot x'_{\text{ini}} \tag{96}
\]
\[
x'_{\text{end}} = -\sqrt{|k_1|} \cdot \sin (l \sqrt{|k_1|}) \cdot x_{\text{ini}} + \cos (l \sqrt{|k_1|}) \cdot x'_{\text{ini}}, \tag{97}
\]
for a positive normalized gradient and
\[
x_{\text{end}} = \cosh (l \sqrt{|k_1|}) \cdot x_{\text{ini}} + \frac{1}{\sqrt{|k_1|}} \cdot \sinh (l \sqrt{|k_1|}) \cdot x'_{\text{ini}} \tag{98}
\]
\[
x'_{\text{end}} = \sqrt{|k_1|} \cdot \sinh (l \sqrt{|k_1|}) \cdot x_{\text{ini}} + \cosh (l \sqrt{|k_1|}) \cdot x'_{\text{ini}}, \tag{99}
\]
for a negative normalized gradient. In matrix form the propagation can now be written as
\[
\begin{pmatrix}
\delta_{\text{end}} \\
x_{\text{end}} \\
x'_{\text{end}}
\end{pmatrix} =
\begin{pmatrix}
\cos (l \sqrt{|k_1|}) & \frac{1}{\sqrt{|k_1|}} \cdot \sin (l \sqrt{|k_1|}) & 0 \\
-\sqrt{|k_1|} \cdot \sin (l \sqrt{|k_1|}) & \cos (l \sqrt{|k_1|}) & 0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
x_{\text{ini}} \\
x'_{\text{ini}} \\
\delta_{\text{pini}}
\end{pmatrix} \tag{100}
\]
for a positive normalized quadrupole gradient and
\[
\begin{pmatrix}
\delta_{\text{end}} \\
x_{\text{end}} \\
x'_{\text{end}}
\end{pmatrix} =
\begin{pmatrix}
\cosh (l \sqrt{|k_1|}) & \frac{1}{\sqrt{|k_1|}} \cdot \sinh (l \sqrt{|k_1|}) & 0 \\
\sqrt{|k_1|} \cdot \sinh (l \sqrt{|k_1|}) & \cosh (l \sqrt{|k_1|}) & 0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
x_{\text{ini}} \\
x'_{\text{ini}} \\
\delta_{\text{pini}}
\end{pmatrix} \tag{101}
\]
for a negative normalized quadrupole gradient.

Even though the deflection inside a quadrupole magnet depends on the particle momentum a quadrupole does not have a nominal deflection angle without a quadrupole offset and thus, does not contribute to first order to the dispersion of a storage ring. We will come back to this point in Section 2 when we use perturbation treatment for calculating the dispersion function in a storage ring. The extended quadrupole transfer matrices can therefore be written as:
\[
\begin{pmatrix}
\delta_{\text{end}} \\
x_{\text{end}} \\
x'_{\text{end}}
\end{pmatrix} =
\begin{pmatrix}
\cos (l \sqrt{|k_1|}) & \frac{1}{\sqrt{|k_1|}} \cdot \sin (l \sqrt{|k_1|}) & 0 \\
-\sqrt{|k_1|} \cdot \sin (l \sqrt{|k_1|}) & \cos (l \sqrt{|k_1|}) & 0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
x_{\text{ini}} \\
x'_{\text{ini}} \\
\delta_{\text{pini}}
\end{pmatrix} \tag{102}
\]
for a focusing quadrupole and
\[
\begin{pmatrix}
\delta_{\text{end}} \\
x_{\text{end}} \\
x'_{\text{end}}
\end{pmatrix} =
\begin{pmatrix}
\cosh (l \sqrt{|k_1|}) & \frac{1}{\sqrt{|k_1|}} \cdot \sinh (l \sqrt{|k_1|}) & 0 \\
\sqrt{|k_1|} \cdot \sinh (l \sqrt{|k_1|}) & \cosh (l \sqrt{|k_1|}) & 0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
x_{\text{ini}} \\
x'_{\text{ini}} \\
\delta_{\text{pini}}
\end{pmatrix} \tag{103}
\]
for a defocusing quadrupole magnet.

Reducing the length (29) of the quadrupole magnet while keeping the total focal length constant we can obtain a ‘thin lens’ approximation for the quadrupole transport map. One obtains:
\[
\begin{pmatrix}
\delta_{\text{end}} \\
x_{\text{end}} \\
x'_{\text{end}}
\end{pmatrix} =
\begin{pmatrix}
1 & 0 & 0 \\
\frac{1}{f} & 1 & 0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
x_{\text{ini}} \\
x'_{\text{ini}} \\
\delta_{\text{pini}}
\end{pmatrix}
\tag{104}
\]
for a positive normalized quadrupole gradient and
\[
\begin{pmatrix}
\delta_{\text{end}} \\
x_{\text{end}} \\
x'_{\text{end}}
\end{pmatrix} =
\begin{pmatrix}
1 & 0 & 0 \\
\frac{1}{f} & 1 & 0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
x_{\text{ini}} \\
x'_{\text{ini}} \\
\delta_{\text{pini}}
\end{pmatrix} \tag{105}
\]

142
for a negative normalized quadrupole gradient. Most quadrupole magnets of a particle storage ring have a focal length \( f = \frac{1}{lk} \gg l \). (106)

The arguments of the trigonometric functions in the transfer matrices (100) and (101) are usually quite small and the quadrupole 'thin lens' approximation are usually quite good approximations to the real transport matrices.

Comparing the dipole 'thin lens' approximation (88) with the 'thin lens' approximation of a quadrupole magnet we observe that it corresponds to a focusing quadrupole kick with a focal length \( f = \frac{\rho^2}{l} \) (107) and an effective normalized quadrupole gradient

\[ k_1 = \frac{1}{\rho^2}. \] (108)

1.8 Transfer matrices for the 'Twiss parameters'

For the operation of a storage ring one usually likes to know the beam size and divergence at any position in the machine. To this end one needs to know the beam emittance (51), the 'twiss parameters' and the energy spread and the dispersion functions \( D(s) \) and \( D'(s) \) at any position in the storage ring. The differential equation (38) provides in principle the means to calculate the \( \beta \)-function, and thus the 'twiss parameters' as a function of the \( s \) coordinate. However, in practice it is often easier to calculate the twiss parameters by means of transfer matrices.

Assuming the transformation of the particle coordinates through the machine from point \( s_0 \) to \( s \) is given by the matrix:

\[ M(s, s_0) = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \] (109)

we can write the particle coordinates at \( s \) as a function of the initial conditions:

\[ \tilde{z}(s) = M(s, s_0) \cdot \tilde{z}(s_0). \] (110)

The matrix \( M(s_0, s) \) is in principle identical to the matrix \( \phi(s_0, s) \) with the 'sine' and 'cosine' like solutions as columns. However, we have introduced here a different notation in order to emphasize that the matrix can be constructed using the transfer matrices for the magnetic elements in the storage ring without knowledge of the 'sine' and 'cosine' like solutions. We can use therefore the element transfer matrices for calculating the 'sine' and 'cosine' like solutions along the storage ring.

Inverting Eq. (110) we get

\[ \begin{pmatrix} x(s_0) \\ x'(s_0) \end{pmatrix} = \begin{pmatrix} m_{22} & -m_{12} \\ -m_{21} & m_{11} \end{pmatrix} \cdot \begin{pmatrix} x(s) \\ x'(s) \end{pmatrix} \] (111)

where we have used \( \text{det}(M) = 1 \). Inserting the expressions for \( x(s_0) \) and \( x'(s_0) \) into Eq. (49) for the invariant of the particle motion we get:

\[ I = \gamma_0 \dot{x}_0^2 + 2 \alpha_0 x_0 \dot{x}_0 + \beta_0 \dot{x}_0^2 \]

\[ = \gamma_0 (m_{22} x - m_{12} x')^2 + 2 \alpha_0 (m_{22} x - m_{12} x') (-m_{21} x - m_{11} x') + \beta_0 (-m_{21} x - m_{11} x')^2 \]

\[ = \gamma x^2 + 2 \alpha xx' + \beta x'^2 \] (113)
Comparing the coefficients in Eqs. (112) and (113) we obtain for the twiss parameters at $s$:

$$
\begin{pmatrix}
\beta \\
\alpha \\
\gamma \\
\end{pmatrix} =
\begin{pmatrix}
m_{11}^2 & -2m_{11}m_{12} & m_{12}^2 \\
-m_{11}m_{21} & m_{12}m_{21} + m_{11}m_{22} & -m_{12}m_{22} \\
m_{21}^2 & -2m_{21}m_{22} & m_{22}^2 \\
\end{pmatrix}
\cdot
\begin{pmatrix}
\beta_0 \\
\alpha_0 \\
\gamma_0 \\
\end{pmatrix}.
$$

(114)

For the special case where the matrix (109) corresponds to the 'one-turn' map one can calculate the 'twiss parameters' by calculating the eigenvector of Eq. (114) for the eigenvalue 1. Once the 'twiss parameters' are known at one position in the storage ring on can use the transfer matrices of the individual elements in the storage ring and matrix multiplication to propagate the 'twiss parameters' along the storage ring.

Using the transport matrix of a drift section (68) with $m_{11} = 1$ and $m_{12} = s (l = s)$ one obtains for the $\beta$-function at the end of the drift section:

$$
\beta(s) = \beta_0 - 2 \cdot \alpha_0 \cdot l + \gamma_0 \cdot s^2.
$$

(115)

In case the initial conditions are given for a symmetry point with $\alpha_0 = 0$, like the interaction point of an experimental insertion for example, the above expression simplifies to

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

(116)

The experimental detector in a collider requires a minimum space where no quadrupole magnets can be placed. The $\beta$-function value at the IP and the closest distance from the IP where a quadrupole magnetic can be placed are also referred to as $\beta^*$ and $L^*$. Equation (116) indicates that the $\beta$-function inside the first quadrupole magnets left and right from the IP increases quadratically with $L^*$ and inversely with $\beta^*$. Designing a so called low $\beta$ insertion therefore either requires the installation of large aperture quadrupole magnets left and right from the 'low $\beta$' point or a compact detector design so that the quadrupole magnets can be placed as close as possible to the 'low $\beta$' point. The first solution is limited by magnet technology and the fact that the peak field at the magnet coils increases for a given field gradient linearly with the quadrupole aperture. The second option implies a strong focusing and thus large quadrupole gradients which are limited again by the magnet technology and the resulting large chromatic errors (see Section 2.2.1).

The dispersion function is not part of the 'twiss parameters' but it can be calculated in a similar fashion using an 'extended one-turn map'. For the 'extended one-turn map' the relative momentum error of a particle is treated as an additional degree of freedom of the particle motion:

$$
\begin{pmatrix}
x(s) \\
x'(s) \\
\delta p(s) \\
\end{pmatrix} = \tilde{M} \cdot \begin{pmatrix}
x(s_0) \\
x'(s_0) \\
\delta p(s_0) \\
\end{pmatrix}.
$$

(117)

The dispersion functions can now be calculated as the eigenvector components of the extended one turn map:

$$
\begin{pmatrix}
\delta \cdot D_x(s + L) \\
\delta \cdot D'_x(s + L) \\
\delta p(s + L) \\
\end{pmatrix} = \tilde{M} \cdot \begin{pmatrix}
\delta \cdot D_x(s) \\
\delta \cdot D'_x(s) \\
\delta p(s) \\
\end{pmatrix}.
$$

(118)

The extended transformation matrices for individual magnet elements can then be used to propagate the dispersion function along the $s$ coordinate of the storage ring.

2 Perturbed equation of motion

So far we have assumed that all elements in the storage ring are installed and powered according to their design values. However, in a real storage ring there will always be deviations from these ideal settings.
In the following we will look at dipole and quadrupole field perturbations to the equations of motion and study to what extend these perturbations will change the closed orbit and the 'sine' and 'cosine' like solutions of Hill’s equations.

Deviations from the ideal settings generate additional terms in the Hills equation leading to inhomogeneous differential equations:

\[ x'' + K(s)x = G(s) \]  \hspace{1cm} (119)

with \( G(s + L) = G(s) \). For dipole and quadrupole perturbations the inhomogeneous part can be written as

\[ G = \begin{cases} -\Delta k_0 & \text{dipole field} \\ -\Delta k_1 \cdot x & \text{quadrupole field} \end{cases} \]  \hspace{1cm} (120)

The minus sign for a dipole perturbation comes from our choice of a rotating coordinate system (see Eq. (8)). Note here that an additional dipole with the same field strength as the nominal bending magnets in the storage rings results in a perturbation that is proportional to \( 1/\rho \) while the focusing term in Eq. (30) is proportional to \( 1/\rho^2 \). The additional power in \( \rho \) comes from the changing orientation of the coordinate system which follows the design radius of curvature around the storage ring. Treating a dipole magnet as a perturbation implies that the non-perturbed system features no change in the orientation of the coordinate system and one obtains a force term that is only proportional to \( 1/\rho \).

The inhomogeneous part of the equations of motion for an arbitrary magnetic field perturbation can be written as

\[ G(s) = -\frac{F_{L,x}}{p \cdot v} = -\frac{q \cdot B_y(s)}{p} \]  \hspace{1cm} (121) \hspace{1cm} (122)

where the minus sign comes from our choice of a rotating coordinate system.

In vector form we can write the perturbed equations of motion as a system of linear first order inhomogeneous differential equations

\[ \frac{d\vec{z}}{ds} + \begin{pmatrix} 0 & -1 \\ K(s) & 0 \end{pmatrix} \cdot \vec{z} = \begin{pmatrix} 0 \\ G_x(s) \end{pmatrix}. \]  \hspace{1cm} (123)

2.1 Dipole perturbations

There are three main sources for dipole field perturbations in a storage ring:

- errors in the dipole field settings
- energy errors of the particles in the storage ring
- feed down errors from quadrupole fields

Perturbations of the dipole field generate a change in the closed orbit of the storage ring. These orbit perturbations can severely limit the machine performance and the successful operation of a storage ring requires a precise measurement and correction of the orbit perturbations. For example, closed orbit errors can result in

- an effective reduction of the mechanical aperture and potential beam losses
- a coupling of the horizontal and vertical particle motion
- a change of the beam energy (this will be illustrated below)
- larger oscillation amplitudes inside the magnets that expose the particles to a reduced field quality and non-linear magnetic field errors
– change of the beam size due to dispersion
– and a change of the beam separation inside a collider.

Before we quantify the impact of dipole field perturbations onto the closed orbit and address the important topic of orbit correction algorithms inside a storage ring we summarize first the main sources for dipole perturbations inside a storage ring and illustrate the effect on the closed orbit with various examples from the LEP storage ring at CERN.

2.1.1 Errors in the dipole field settings

Errors in the dipole field settings generate a perturbation term of the form

$$\Delta k_0 = q \cdot \frac{\Delta B}{p}. \quad (124)$$

The dipole field setting is limited by the accuracy to which one knows the magnet transfer function that translates a given electric current in the magnet into a magnetic field value and the accuracy to which one can adjust the electric current in the power converters of the storage ring. In an ideal magnet the transfer function should be a linear function defined by Amperes law:

$$\oint H = N \cdot I, \quad (125)$$

where $N$ is the number of coil windings and $I$ the electric current in the coil. However, the transfer function in a real magnet differs from the linear dependence due to hysteresis, persistent current and saturation effects in the magnets. A precise knowledge of the transfer function at low and high field settings requires a thorough measurement of the magnetic field as a function of the magnet current prior to the magnet installation in the tunnel. Measurements should typically provide a relative transfer accuracy of the order of

$$\Delta B \approx 10 \cdot 10^{-4} \cdot B_0, \quad (126)$$

where $B_0$ is the design value of the magnetic field. The accuracy of the power converter adjustments is normally much better than the accuracy of the magnet transfer function. For example, for the LHC the power converter accuracy reaches values of

$$\Delta I \approx 10^{-6} \cdot I_0. \quad (127)$$

However, there can be unexpected effects in the real storage ring that spoil the accuracy of the calibration measurements in the laboratory. One curious example for such an unexpected effect was the perturbation of the electric magnet current in LEP due to the near by passing of the French TGV train line [6]. The TGV train picks up its electric current via a suspended power line above the rail tracks and the return current flows through the train track back to ground. Some fraction of the return current can pass through the LEP magnet system wherever the train tracks run by in close vicinity of the LEP tunnel. Figure 5 show a schematic illustration of this effect. The top picture in Fig. 6 shows the measured return current in the TGV rail tracks as a function of time. The picture in the middle shows the measured voltage on the LEP beam pipe during the same time interval and the bottom picture the measured change of the dipole field in the tunnel (measured with the help of NMR probes) [7]. The three pictures in Fig. 6 nicely illustrate the correlation of the time varying dipole field in the LEP tunnel with the measured voltage on the TGV rail tracks.

2.1.2 Energy errors of the particles in the storage ring

The particle deflection inside a dipole field depends on the particle energy (see Eqs. (5) and (6)) and any deviation from the design beam energy will ultimately result in an error of the deflection angle and the
closed orbit. Expanding the expression for the normalized dipole strength in Eq. (6) around the design beam momentum one obtains to first order in the momentum error

\[ k_0 = k_{0,0} + \Delta k_0 \]
\[ = \left( 1 - \frac{\Delta p}{p_0} \right) \cdot k_0, \quad (128) \]

and

\[ \Delta k_0 = -\frac{\Delta p}{p_0} \cdot k_0 \quad (130) \]

where \( k_{0,0} \) denotes the normalized dipole strength for an on-momentum particle. Substituting Eq. (130) into Eqs. (3) and (4) one obtains for the deflection angle error

\[ \Delta \chi \approx -l \cdot \frac{\Delta p}{p_0} \cdot k_{0,0}. \quad (131) \]

The resulting change of the closed orbit in the storage ring is referred to as the dispersion orbit. Even if the central beam energy can be quite accurately adjusted via beam based measurements the particles inside the beam will always have a non-vanishing energy spread. The precise value of this energy spread depends on the RF system and the size of the storage ring but is usually of the order of

\[ \Delta p \approx 10^{-3} \cdot p_0, \quad (132) \]

where \( p_0 \) is the central beam energy. Due to dispersion in the storage all particles inside the beam circulate therefore in the storage ring on slightly different closed orbits (dispersion orbits).

### 2.1.3 Feed down errors from quadrupole fields

The magnetic field of a quadrupole magnet can be parameterized as

\[ B_x = +g \cdot y \quad (133) \]
\[ B_y = +g \cdot x. \quad (134) \]
Fig. 6: The measured voltage on the TGV rail tracks (top) and the resulting modulation of the voltage on the LEP beam pipe (middle) and the dipole bending field (bottom) as a function of time.

Assuming a horizontal misalignment of the quadrupole magnet the field seen by the passing beam can be written as

\[
B_x = +g \cdot y \\
B_y = +g \cdot (x + \tilde{x}),
\]

(135)  

(136)

where \(x\) and \(y\) are the coordinates within the moving coordinate system and \(\tilde{x}\) the horizontal quadrupole offset with respect to the ideal design orbit. Equation (136) shows that the quadrupole displacement adds an additional vertical dipole field component. The additional dipole field component generates an additional deflection angle of the beam and changes ultimately the closed orbit in the storage ring. The quadrupole alignment errors are very important contributions to the closed orbit inside a storage ring and we will illustrate this effect with a few examples from the LEP storage ring at CERN.

Placing magnets inside a real storage ring can only be done with a finite accuracy. Most machines aim at an accuracy of \(\Delta z, x < 1\) mm. However, the final precision depends on the size of the storage ring and the stability of the tunnel floor. Even if the magnets could be initially placed with a very high accuracy the tunnel will move with time and the magnet alignment eventually degrade. Fig. 7 shows the
Fig. 7: Vertical quadrupole positions in the LEP tunnel for different time periods: top before alignment in December 1992 on a circular presentation. The next lower picture shows the same data with the $s$ coordinate mapped onto the horizontal axis. The third picture shows the quadrupole positions after alignment in 1993 and the last two pictures the positions one year later before and after a second alignment campaign.
vertical quadrupole positions in the LEP tunnel for different time periods [8]. The top picture shows the magnet positions before alignment in December 1992 on a circular presentation. The next lower picture shows the same data with the s coordinate mapped onto the horizontal axis. The third picture shows the quadrupole positions after alignment in April 1993 and the last two pictures the positions one year later before and after a second alignment campaign. Looking at the large negative offset between IR7 and IR8 one clearly recognizes how the magnet positions moved after the first alignment campaign and how the negative magnet offset builds up again.

Figure 8 shows another example for the motion of the magnetic elements in a storage ring. It shows the transverse position of the quadrupole magnets in the super-conducting proton storage ring HERA as a function of time. The HERA storage ring is located underground in a big city (Hamburg, Germany) and Fig. 8 illustrates how the urban activity perturbs the magnet position in the underground tunnel. The top picture shows the maximum peak-to-peak variations and the bottom picture the RMS values for the quadrupole magnet motion [9]. One clearly recognizes that the quadrupole motion is rather small over the weekend, builds up in the morning of each weekday, peaks around noon and calms down again during the night. Figure 8 even shows that the quadrupole motion, and thus the activity in the city, peaks on Monday and reduces slightly during the course of the week indicating that people return to work with new energy after the weekend and become increasingly tired over the week.

2.1.4 The beam energy variations as a function of the closed orbit

The beam energy in a synchrotron is determined by the RF frequency and the total machine circumference. This can be illustrated by looking at the resonance condition between the beam revolution frequency and the RF frequency inside the cavities. A necessary condition for the RF capture of the particles is that the RF frequency is an integer multiple of the revolution frequency. This 'resonance' condition can be written as

\[ f_{RF} = n \cdot f_{rev}. \]  

(137)

Assuming a constant RF frequency the above equation implies that the revolution frequency can not change for all particles that are captured inside an RF bucket. Assuming further for the sake of simplicity a uniform constant dipole field all along the storage ring the revolution frequency can be obtained from Eq. (3). Substituting \( p = m_0 \cdot \gamma \cdot v \), where \( \gamma \) is the relativistic 'gamma' factor

\[ \gamma = \frac{1}{\sqrt{1 - v^2/c^2}}. \]  

(138)

and \( L = 2\pi \rho \) where \( L \) is the total machine circumference, one obtains from Eq. (3) the cyclotron equation:

\[ f_{rev} = \frac{v}{L}. \]  

(139)

\[ f_{rev} = \frac{1}{2\pi} \cdot \frac{qB_0}{m_0 \gamma}. \]  

(140)

The only parameters that can still change in the above expression are the machine circumference \( L \) and the relativistic gamma factor. The RF resonance condition in Eq. (137) therefore implies that the beam energy (\( \gamma \)) changes if the machine circumference changes due to closed orbit errors.

For example, the tidal forces were a peculiar source for a periodic deformation of a storage ring tunnel and the closed orbit in the LEP machine. The common motion of the two-body system of the earth and the moon around their center of gravity generates a periodic deformation of the earth sphere. On the side facing the moon the earth is deformed due to the gravitational pull. The effect is illustrated in Fig. 9. Observing that the earth rotates every 24 hours once around its own axis the above effects generate two modulations of the earth surface with a periodicity of 12 and 24 hours.
Fig. 8: The measured motion of the quadrupole magnets in the super-conducting proton storage ring HERA in Hamburg Germany. The top picture shows the peak to peak variations and the bottom picture the RMS variation of the quadrupole magnets in the underground storage ring tunnel as a function of time. The horizontal axis covers the period of one week.

For large storage rings this modulation can result in a modulation of the tunnel floor and thus to a modulation of the quadrupole alignment and the closed orbit. Figure 10 shows the beam energy variation in the LEP machine as a function of time. The presented data covers a time interval of 24 hours and one clearly recognizes the two modulations with a periodicity of 12 and 24 hours. The solid curve shows the expected modulation due to the two-body motion of the moon and the earth (tidal effect) and the dots indicate measurement data from energy calibration measurements in LEP [10]. The high precision experiments in LEP required a precise knowledge of the LEP beam energy to a precision of $\Delta E \leq 0.003$. The energy modulation shown in Fig. 10 is approximately one order of magnitude larger than this target and a precise analysis of the LEP experimental data requires a correction of the beam energy that takes into account the time of the day and the phase of the moon.

Another source for periodic tunnel movements in LEP was the water level in Lake Geneva. The LEP tunnel is located near Lake Geneva and a change in the water level ultimately changes the position of part of the LEP tunnel and thus the closed orbit and the beam energy. The solid line in Fig. 11 shows the lake level as a function of operation days. The LEP machine started operation in spring where the lake water level is high after the snow in mountains has melted and filled the lake. Over the summer the water level decreases again and obtains an equilibrium towards the end of the year. The tunnel movement implies a change of the quadrupole positions in the tunnel and therefore a change of the closed orbit in...
Fig. 9: The two-body system of the earth and the moon rotating around their center of gravity generating two deformations of the earth sphere: one on the side facing the moon due to the gravitational pull and one on the opposite side of the moon due to the centrifugal force.

the LEP machine. The square data points in Fig. 11 show measurements of the horizontal orbit in one of the LEP arcs and the circular data points measurements of the beam energy. One clearly recognizes the correlation of the LEP beam energy with the water level in the Lake Geneva. The observed effect was again larger than the accuracy requested by the LEP experiments and a precise analysis of the LEP experimental data required not only a precise knowledge of the phase of the moon but also of the water level in Lake Geneva.

2.1.5 Integer resonance

The examples of the previous section underline the need to quantify and correct the orbit response to a given dipole perturbation. Before we calculate the orbit response with the help of the inhomogeneous Hill’s Eqs.(119) and (123) we will briefly illustrate the impact of the machine tune onto the closed orbit response.

The machine tune specifies the number of transverse betatron oscillations during one revolution in the storage ring, Eq. (39). Let us assume an ideal storage ring with a circular design closed orbit. Let us further assume that the particle motion lies initially on the design orbit and that the machine has one single dipole perturbation in one location of the ring. Once the particle passes through the dipole perturbation it experiences a deflection and the particle starts a betatron oscillation around the design closed orbit. This case is illustrated in the top part of Fig. 12. The deflection angle is given by Eqs. (3) and (4):

$$\Delta \chi = l \cdot \Delta k_0$$

(141)

and the betatron oscillation is given by the sine-like solution in Eq. (58) with the initial conditions

$$x(s_0) = 0 \text{ and } x'(s_0) = \Delta \chi,$$

(142)

where \(s_0\) refers to the position of the perturbation and where the phase term \(\phi_0\) in the sine-like solution refers to the phase variable at the location of the perturbation. Together with the initial conditions (142) and using Eq. (62) the trajectory response after the perturbation can be written as

$$x(s) = \sqrt{\beta(s) \beta(s_0)} \cdot \Delta \chi \cdot \sin (\phi(s) - \phi(s_0))$$

(143)

$$x'(s) = \Delta \chi \sqrt{\beta(s_0)} [\cos (\phi(s) - \phi(s_0) + \alpha(s) \sin (\phi(s) - \phi(s_0))] / \sqrt{\beta(s)}.$$

(144)
Fig. 10: The beam energy variation in the LEP machine as a function of time. The solid curve shows the expected modulation due to the two-body motion of the moon and the earth (tidal effect) and the dots indicate measurement data from energy calibration measurements.

Fig. 11: Correlation of the LEP orbit and the LEP beam energy with the Lac Leman water level.
For an integer tune we have

$$\phi(s + L) = \phi(s) + n \cdot 2\pi \quad (145)$$

where $L$ is the circumference of the storage ring and we can write the trajectory response after one full revolution as

$$\begin{align*}
x(s_0 + L) &= 0 \\
x'(s_0 + L) &= \Delta \chi.
\end{align*} \quad (146) \quad (147)$$

In other words, the particle comes back to the perturbation with the same betatron oscillation amplitude and the same trajectory slope with which it left the perturbation on the previous passing. This case is illustrated in the bottom part of Fig. 12. The contributions from consecutive passings therefore add up and the particle motion is unstable.

For a half-integer tune we have

$$\phi(s + L) = \phi(s) + n \cdot \pi \quad (148)$$

and we can write the trajectory response after one full revolution as

$$\begin{align*}
x(s_0 + L) &= -x(s_0) \\
x'(s_0 + L) &= -x'(s_0) = -\Delta \chi.
\end{align*} \quad (149) \quad (150)$$

In other words, the particle comes back to the perturbation with the opposite betatron oscillation amplitude and trajectory slope with which it left the perturbation on the previous passing. This case is illustrated in the bottom part of Fig. 13. The contributions from two consecutive passings therefore cancel each other. The particle motion lies again on the ideal design orbit after the second passage through the perturbation ($\rightarrow$ zero orbit). At the third passage the picture repeats itself. The long term particle motion is therefore stable. Since the closed orbit is defined as the average trajectory over many turns the net closed orbit response to the perturbation is just half of the trajectory response during the first passage:

$$CO(s) = \frac{1}{2} \sqrt{\beta(s)\beta(s_0)} \cdot \Delta \chi \cdot \sin (|\phi(s) - \phi(s_0)|) \quad (151)$$
Fig. 13: The trajectory response to a single dipole perturbation in an ideal circular storage ring for a half-integer tune value. The top picture shows the trajectory response after the first and the bottom picture after the second passage through the perturbation. The deflections of two consecutive passages through the perturbation cancel each other.

\[ CO(s) = \frac{1}{2} \sqrt{\beta(s) \beta(s_0)} \cdot l \cdot k_0 \cdot \sin (|\phi(s) - \phi(s_0)|). \]  

(152)

The absolute value in the argument of the sine function indicates that the slope of the closed orbit changes its sign at the location of the dipole perturbation. The derivative of the closed orbit has therefore a discontinuity at the location of the perturbation. Such a 'kink' in the closed orbit response is illustrated in Fig. 13 and is a clear signature for dipole perturbations in closed orbit measurements. However, in a real storage ring it is not always straightforward to recognize this pattern. In a real storage ring the closed orbit is only sampled at dedicated locations which are equipped with Beam Position Monitors (BPMs) (see Section 4.1 for more details). The closed orbit readings are then displayed in the control room in regular intervals which do not necessarily reflect the actual phase advance between the readings and an accurate interpretation of the closed orbit data requires some care. For example, Fig. 14 shows a simulated closed orbit response in the CERN SPS storage ring with one single dipole perturbation as it would appear in the SPS control room [14]. The phase advance between the individual readings in the central part is approximately 90° and four consecutive reading correspond to one full betatron oscillation. One clearly recognizes the 'kink' in this pattern at the central reading at monitor 57. Figure 15 shows a measured closed orbit for the CERN LEP storage ring before orbit correction. One could be tempted to conclude from Fig. 15 that the initial closed orbit features a regular oscillation pattern and that the closed orbit perturbation is caused by one single dipole error. However, the LEP orbit monitors are not all spaced by regular 90° intervals and the actual phase advance between consecutive monitors can vary significantly along the machine. A proper analysis of the orbit will show in this case that the closed orbit perturbation is caused by a large number of small randomly distributed dipole kicks.

Before we discuss the exact calculation of the closed orbit response for arbitrary tune values in Section 3 we will briefly estimate the closed orbit response for other than integer and half-integer tune values. To this end we introduce the smooth approximation of the betatron oscillation. Assuming a constant beta function along the storage ring one can write

\[ \alpha(s) = 0 \]  

(153)

\[ \phi(s) = \int_{s_0}^{s} \frac{1}{\beta(\tau)} d\tau + \phi_0 \]

\[ = \frac{s - s_0}{\langle \beta \rangle} + \phi_0 \]  

(154)
where $L$ is the circumference and $\langle \beta \rangle$ the average beta function in the storage ring. We can then write the trajectory response after the dipole perturbation as

$$
x(s) = A \cdot \sin \left( 2\pi Q \cdot \frac{s}{L} + \phi_0 \right)
$$

$$
x'(s) = A \cdot \frac{2\pi Q}{L} \cos \left( 2\pi Q \cdot \frac{s}{L} + \phi_0 \right).$$

(156)

The periodic boundary condition of the closed orbit solution can now be written as

$$CO(s_0 + L) = CO(s_0).$$

(157)

For the canonical momentum we note the condition

$$\lim_{\epsilon \to 0} CO'(s_0 + L - \epsilon) - CO'(s_0 + \epsilon) = \Delta \chi$$

(158)

In the smooth approximation we have further $\alpha(s) = 0$ and obtain the following boundary conditions for the canonical momentum

$$\lim_{\epsilon \to 0} CO'(s_0 + L - \epsilon) = -CO'(s_0 + \epsilon).$$

(159)

Note that while condition (158) is always satisfied by the closed orbit response the condition (159) is not necessarily satisfied for $\alpha(s_0) \neq 0$. From Eqs. (157) and (159) we can calculate the initial phase in Eq. (156) that satisfies the closed orbit boundary condition

$$\phi_0 = \frac{\pi}{2} - Q.$$  

(160)

Inserting the closed orbit phase into Eq. (158) we obtain

$$\frac{2\pi Q}{L} \cdot A \cdot \cos \left( \frac{\pi}{2} - 2\pi Q \right) = \frac{\Delta \chi}{2}.$$  

(161)

Equation (161) can be solved for the peak oscillation amplitude in Eq. (156)

$$A = \frac{1}{2 \sin (2\pi Q)} \cdot \langle \beta \rangle \cdot \Delta \chi.$$  

(162)

where we have expressed the average phase advance in the storage ring in terms of the average beta function using expressions (154) and (155). Equation (162) illustrates again the closed orbit instability for integer tune values ($\rightarrow \sin (2\pi Q) = 0$).

### 2.2 Quadrupole perturbations

There are four main sources for quadrupole field perturbations in a storage ring:

- errors in the quadrupole field settings
- field errors in magnets
- energy errors of the particles in the storage ring
- feed down errors from sextupole fields
Fig. 14: Simulated vertical closed orbit in the SPS with one single dipole perturbation. The phase advance between the individual reading in the central part is approximately $90^\circ$ and four consecutive reading correspond to one full betatron oscillation. One clearly recognizes the 'kink' in this pattern at the central reading at monitor 57.

Fig. 15: Vertical closed orbit in LEP before orbit correction.
Errors in the quadrupole field settings generate a perturbation term of the form

\[ \Delta k_1 = q \cdot \Delta g_p. \]  \hspace{1cm} (163)

The additional quadrupole field components generate additional focusing terms and ultimately change the beta functions and the tune in the storage ring. These perturbations can severely limit the machine performance and the successful operation of a storage ring requires a precise measurement and correction of the perturbations.

For example,

- tune errors can move the particle motion onto dangerous resonances (see for example the integer resonance in Fig. 12)
- errors in the beta-functions and the resulting change in the oscillation amplitudes can increase the beam size and reduce the effective mechanical aperture
- larger oscillation amplitudes inside the magnets can expose the particles to a reduced field quality and additional non-linear magnetic field errors
- roll errors of the quadrupole magnets can generate a coupling of the horizontal and vertical particle motion.

### 2.2.1 Errors in the quadrupole field settings

As for the case of dipole field settings the accuracy of the quadrupole field setting is limited by the knowledge of the magnet transfer function and the accuracy of the electric power converter settings in the storage ring.

### 2.2.2 Field errors in magnets

Field errors in magnets can add additional quadrupole field components in the storage ring. For example, the main dipole magnets in a storage ring can feature a quadrupole field error component that is proportional to the dipole field and that changes the net quadrupole field seen by the circulating particles in the storage ring.

### 2.2.3 Energy errors of the particles in the storage ring

The particle deflection inside a quadrupole field depends on the particle position inside the quadrupole and its energy, see Eq. (25). Any deviation from the design beam energy will ultimately result in an error of the quadrupole gradient. Expanding the expression for the normalized quadrupole strength in Eq. (25) around the design beam momentum one obtains to first order in the momentum error

\[ k_1 = k_{1,0} + \Delta k_1 \]

\[ = (1 - \frac{\Delta p}{p_0}) \cdot k_1, \]  \hspace{1cm} (164)

and

\[ \Delta k_1 = -\frac{\Delta p}{p_0} \cdot k_{1,0}, \]  \hspace{1cm} (165)

where \( k_{1,0} \) denotes the normalized quadrupole strength for an on-momentum particle. The change in the quadrupole focusing results in a change of the total machine tune. The energy dependent term of the machine tune is called the chromaticity and the part of the chromaticity that is caused by the energy dependent change of the quadrupole focusing is called the natural chromaticity.
 Feed down errors from sextupole fields

The magnetic field of a sextupole magnet can be parameterized as

\[ B_x = g' \cdot x \cdot y \]  \hspace{1cm} (167)  
\[ B_y = \frac{1}{2} \cdot g' \cdot (x^2 - y^2) \]  \hspace{1cm} (168)  

and Fig. 16 shows a schematic illustration of the sextupole field lines with the rotating coordinate system as the reference system. It is worth mentioning here that, contrary to the quadrupole field parameterization, Eqs. (168) and (167) have the same signs for the 'magnet' and 'accelerator' coordinate systems and there is no ambiguity when referring to a positive sextupole gradient. In analogy to the normalized dipole and quadrupole strength we define the normalized sextupole strength

\[ k_2 = \frac{e \cdot g'}{p} \]  \hspace{1cm} (169)  

which can be conveniently written as

\[ k_2 = 0.2998 \cdot \frac{g'[T/m^2]}{p[GeV]} \]  \hspace{1cm} (170)  

The normalized sextupole strength has the units \( [\lambda] = 1/m^3 \). Assuming a horizontal misalignment of the sextupole magnet the field seen by the passing beam can be written as

\[ B_x = g' \cdot x \cdot y \]  \hspace{1cm} (171)  
\[ B_y = \frac{1}{2} \cdot g' \cdot ([x + \bar{x}]^2 - y^2) \]  \hspace{1cm} (172)  

where \( x \) and \( y \) are the coordinates within the moving coordinate system and \( \bar{x} \) the horizontal sextupole offset with respect to the ideal design orbit. Expanding the square on the right hand side of Eq. (172) yields

\[ B_x = g' \cdot x \cdot y + \]  \hspace{1cm} (173)  

Fig. 16: Schematic sketch of a sextupole cross section: the red arrows indicate the magnetic field lines.
Equation (174) shows that the sextupole displacement adds an additional quadrupole term with an effective gradient of
\[ g' \cdot \tilde{x} \cdot y \]
\[ B_x = \frac{1}{2} \cdot g' \cdot (x^2 - y^2) + \]
\[ g' \cdot \tilde{x} \cdot x + \frac{1}{2} \cdot g' \cdot \tilde{x}^2 \]  
(174)

Equation (174) shows that the sextupole displacement adds an additional quadrupole term with an effective gradient of \( g = g' \cdot \tilde{x} \) and an effective normalized quadrupole strength of
\[ k_1 = k_2 \cdot \tilde{x}. \]  
(175)

### 2.2.5 Half-integer resonance

Before we quantify the impact of quadrupole field perturbations onto the solutions of Hills equations we illustrate a new type of resonance that is driven by quadrupole field errors. Let us assume an ideal storage ring with a circular design closed orbit and that the storage ring has a single, localized quadrupole gradient error in one location of the ring. Let us further assume that the particle motion has an initial orbit oscillation around the closed orbit of the storage ring. Once the particle passes with an offset through the quadrupole magnet is experiences an additional deflection due to the quadrupole gradient error that has not yet been included in the discussion of the 'integer resonance'. The top part in Fig. 13 illustrates the additional betatron oscillation that starts after the passage through the quadrupole perturbation.

The deflection angle is given by Eqs. (136), (3) and (4):
\[ \Delta \chi = l \cdot \tilde{x} \cdot \Delta \lambda \]  
(176)

where \( \tilde{x} \) is the particle offset inside the perturbed quadrupole magnet. Let us assume in the following that the initial orbit oscillation is given by the cosine-like solution in Eq. (59). The additional betatron oscillation due to Eq. (176) is given by the sine-like solution in Eq. (58) and the initial conditions after the first passage through the perturbed quadrupole magnet can be written as
\[ x(s_0) = \tilde{x} \]  
\[ x'(s_0) = \Delta \chi, \]  
(177)

where \( s_0 \) refers to the position of the perturbation. Together with the initial conditions (178) the trajectory response after the perturbation can be written as
\[ x(s) = \tilde{x} \cdot \sqrt{\frac{\beta(s)}{\beta_0}} \cdot \left( \cos(\Delta \phi) + \alpha_0 \sin(\Delta \phi) + \sqrt{\beta(s)} \cdot \Delta \chi \cdot \sin(\Delta \phi) \right) \]  
(178)
\[ x'(s) = \left( [\alpha(s) - \alpha_0] \cdot \cos(\Delta \phi) - [1 - \alpha(s) \alpha_0] \cdot \sin(\Delta \phi) \right) / \sqrt{\beta(s) \beta_0} + \]  
\[ \Delta \chi \sqrt{\beta_0} \left[ \cos(\Delta \phi) + \alpha(s) \sin(\Delta \phi) \right] / \sqrt{\beta(s)} \]  
(179)

For a half-integer tune we have
\[ \Delta \phi(s + L) = \Delta \phi(s) + n \cdot 2\pi + \pi \]  
(180)

where \( L \) is the circumference of the storage ring and we can write the trajectory response after one full revolution as
\[ x(s_0 + L) = -\tilde{x} \]  
\[ x'(s_0 + L) = -\Delta \chi. \]  
(181)

(182)

In other words, the particle comes back to the perturbation with the opposite betatron oscillation amplitude and the opposite trajectory slope that it left the perturbation with on the previous passing. Since the deflection angle depends on the particle offset inside the quadrupole magnet it also changes its sign at the second passing and points again in the same direction as the trajectory slope. This case is illustrated in the bottom part of Fig. 17. The contributions from consecutive passings therefore add up and the particle motion is unstable.
Fig. 17: The trajectory response to a single quadrupole perturbation in an ideal circular storage ring for a half-integer tune value. The top picture shows the trajectory response after the first and the bottom picture after the second passage through the perturbation. The deflections of two consecutive passages through the perturbation add up and the particle motion is unstable.

3 Solution to the inhomogeneous equations of motion

After having illustrated the effect of the 'integer' and 'half integer' resonances we now proceed in deriving quantitative expressions for the effect of the perturbations. We will follow two different procedures for the dipole and quadrupole field perturbations. From the discussion in Section 2.1.5 we know that dipole perturbations change the closed orbit in the storage ring and we will use the inhomogeneous differential equations as a starting point for our calculations. From the discussions in Section 2.2.5 we know that quadrupole perturbations change the beta-function values and the tune in the storage ring and we will use transfer maps as the starting point for our calculations.

3.1 Closed orbit response to dipole errors

The starting point for our calculation is the inhomogeneous Hill’s equation in vector form (123). The space of all solutions to Hill’s equations is given by the linear combination of the 'sine' and 'cosine' like solutions (62). Using the matrix $\Phi$ with the 'sine' and 'cosine' like solutions as columns one can write the general solution to the homogeneous Hill’s equations as

$$\vec{\psi} = \Phi \cdot \vec{u},$$

(183)

where $\vec{u}$ is a vector of coefficients that defines the initial conditions of the solution at position $s_0$. We set the constant phase contribution $\phi_0$ in Eqs. (58) and (59) to zero since any constant phase can be expressed in terms of a linear combination of the 'sine' and 'cosine' like solutions

$$\phi_0 = 0.$$  

(184)

The space of all solutions to the inhomogeneous Hill’s equations is given by the solution space of the homogeneous equation plus one particular solution to the inhomogeneous equations. In other words, we only need to find one particular solution to the inhomogeneous equations in order to solve the system of linear first order differential equations for any boundary condition! A particular solution to the inhomogeneous differential equations can be obtained via a ‘variation of the constant’ vector $\vec{u}$ in Eq. (183). In other words, we write the particular solution of the inhomogeneous Hill’s Equation again
of as linear combination of the 'sine' and 'cosine' like solutions (see Eq. (183)) but allow the coefficient vector \( \vec{u} \) to be itself a function of the \( s \) coordinate. Substituting this 'Ansatz' into the inhomogeneous Hills Equation (123) we obtain

\[
\Phi \cdot \frac{d\vec{u}}{ds} = \begin{pmatrix} 0 \\ -\Delta k_0 \end{pmatrix}.
\]

Solving Eq. (185) for \( \vec{u}(s) \) one obtains

\[
\vec{u}(s) = \int_{s_0}^{s} \Phi(s_0, t)^{-1} \cdot \begin{pmatrix} 0 \\ -\Delta k_0 \end{pmatrix} dt
\]

and the general closed orbit response can be written in vector notation as

\[
\vec{CO}(s) = \begin{pmatrix} CO(s) \\ CO'(s) \end{pmatrix}
\]

\[
\vec{CO}(s) = c_1 \cdot \vec{C} + c_2 \cdot \vec{S} + \Phi(s_0, s) \cdot \int_{s_0}^{s} \Phi^{-1}(s_0, t) \cdot \begin{pmatrix} 0 \\ -\Delta k_0 \end{pmatrix} dt
\]

with \([c_1] = m\) and \([c_2] = 1\).

The inverse matrix \( \Phi^{-1} \) can be calculated from Eq. (61) via standard matrix inversion procedures for \( 2 \times 2 \) matrices

\[
U = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \rightarrow U^{-1} = \frac{1}{\det(U)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}
\]

with

\[
\det(U) = a \cdot d - b \cdot c.
\]

Applying Eqs. (189) and (188) to \( \Phi \) we get

\[
\det(\Phi) = 1
\]

and

\[
\Phi^{-1}(t) = \begin{pmatrix} S'(t) & -S(t) \\ -C'(t) & C(t) \end{pmatrix},
\]

where we have suppressed \( s_0 \) in the arguments of the map and the 'sine' and 'cosine' like solutions in order to simplify the notation. Substituting Eq. (191) into Eq. (187) we get:

\[
CO(s) = c_1 \cdot C(s) + c_2 \cdot S(s) + XO(s)
\]

\[
CO'(s) = c_1 \cdot C'(s) + c_2 \cdot S'(s) + XO'(s)
\]

where we have introduced the notation

\[
XO(s) = XO(s, s_0) = S(s) \int_{s_0}^{s} (-\Delta k_0(t)) \cdot C(t) dt - C(s) \int_{s_0}^{s} (-\Delta k_0(t)) \cdot S(t) dt
\]

\[
XO'(s) = XO(s, s_0) = S'(s) \int_{s_0}^{s} (-\Delta k_0(t)) \cdot C(t) dt - C'(s) \int_{s_0}^{s} (-\Delta k_0(t)) \cdot S(t) dt.
\]

Now we only need to determine the coefficients \( c_1 \) and \( c_2 \) so that Eqs. (192) and (193) satisfy the periodic boundary conditions of the closed orbit:

\[
CO(s + L) = CO(s)
\]

\[
CO'(s + L) = CO'(s).
\]
For $\phi_0 = 0$ we have for the 'sine' and 'cosine' like solutions

\[ C(s_0) = 1; \quad C'(s_0) = 0; \quad S(s_0) = 0; \quad S'(s_0) = 1 \]

and the boundary conditions at $s = s_0$ can be rewritten in matrix form as

\[
M \cdot \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = - \begin{pmatrix} XO(s_0 + L) \\ XO'(s_0 + L) \end{pmatrix},
\]

with

\[
M = \Phi(s_0 + L) - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}
\]

\[
= \begin{pmatrix} C(s_0 + L) - 1 & S(s_0 + L) \\ C'(s_0 + L) & S'(s_0 + L) - 1 \end{pmatrix}.
\]

The solution for the coefficients are given by

\[
\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = -M^{-1} \cdot \begin{pmatrix} XO(s_0 + L) \\ XO'(s_0 + L) \end{pmatrix},
\]

where the inverse matrix $M^{-1}$ is given by

\[
M^{-1} = \frac{1}{\det(M)} \begin{pmatrix} S'(s_0 + L) - 1 & -S(s_0 + L) \\ -C'(s_0 + L) & C(s_0 + L) - 1 \end{pmatrix}.
\]

### 3.1.1 Determinant for the coefficient matrix

First we calculate the determinant in Eq. (203). The determinant of $M$ is given by

\[
\det(M) = (C(s_0 + L) - 1) \cdot [S'(s_0 + L) - 1] - S(s_0 + L) \cdot C'(s_0 + L)
\]

and can be written in a compact form recognizing that $\Phi(s_0 + L)$ is the linear one-turn map at the position $s_0$ and that the trace of the one-turn map is given by $2 \cdot \cos(2\pi Q)$ [1]:

\[
\det(M(s_0 + L)) = \det(\Phi) - \text{Tr}[\Phi(s_0 + L)] + 1
\]

\[
= 2 - 2 \cdot \cos(2\pi Q)
\]

\[
= 4 \cdot \sin^2(\pi Q)
\]

where we applied the product rule for trigonometric functions

\[
\sin \alpha \cdot \sin \beta = \frac{1}{2} \cdot [\cos (\alpha - \beta) - \cos (\alpha + \beta)].
\]

### 3.1.2 Solving for the coefficient $c_1$

Solving Eq. (202) for the coefficients $c_1$ and $c_2$ yields:

\[
c_1 = \frac{-1}{4 \sin^2(\pi Q)} \cdot \left[ S' \cdot XO - S \cdot XO' - XO' \right]
\]

\[
c_2 = \frac{-1}{4 \sin^2(\pi Q)} \cdot \left[ -C' \cdot XO + C \cdot XO' - XO' \right],
\]
where all \( s \) dependent functions are to be evaluated at \( s = s_0 + L \). Substituting Eqs. (194) and (195) into Eq. (207) yields for the coefficient \( c_1 \):

\[
c_1 = \left[ S' \cdot C \cdot \int_{s_0}^{s_0+L} (-\Delta k_0(t))S(t)dt - \int_{s_0}^{s_0+L} (-\Delta k_0(t))C(t)dt \right]_{b_1}^{a_1} + \left[ S \cdot C' \cdot \int_{s_0}^{s_0+L} (-\Delta k_0(t))S(t)dt + S' \cdot \int_{s_0}^{s_0+L} (-\Delta k_0(t))C(t)dt \right]_{b_2}^{a_2} + C(s_0 + L) \cdot \int_{s_0}^{s_0+L} (-\Delta k_0(t))S(t)dt + S(s_0 + L) \cdot \int_{s_0}^{s_0+L} (-\Delta k_0(t))C(t)dt \right] / (4 \cdot \sin^2(\pi Q))
\]

The terms \( b_1 \) and \( b_2 \) in Eq. (209) cancel and the terms \( a_1 \) and \( a_2 \) can be simplified using \( S' C - SC' = \det[\Phi] = 1 \). The coefficient \( c_1 \) can be now written as:

\[
c_1 = \left[ \int_{s_0}^{s_0+L} (-\Delta k_0(t))S(t)dt \right]_{s_0}^{s_0+L} - C(s_0 + L) \cdot \int_{s_0}^{s_0+L} (-\Delta k_0(t))S(t)dt + S(s_0 + L) \cdot \int_{s_0}^{s_0+L} (-\Delta k_0(t))C(t)dt \right] / (4 \cdot \sin^2(\pi Q))
\]

Substituting for the cosine and sine like solutions

\[
C(t) = \frac{\sqrt{\beta(t)}}{\sqrt{\beta_0}} \cdot [\cos (\Delta \phi(t)) + \alpha_0 \cdot \sin (\Delta \phi(t))] \quad (211)
\]

\[
S(t) = \sqrt{\beta(t)\beta_0} \cdot \sin (\Delta \phi(t)) \quad (212)
\]

and at \( s = s_0 + L \)

\[
C(s_0 + L) = \cos (2\pi Q) + \alpha_0 \cdot \sin (2\pi Q) \quad (213)
\]

\[
S(s_0 + L) = \beta_0 \cdot \sin (2\pi Q) \quad (214)
\]

one obtains for the coefficient \( c_1 \)

\[
c_1 = \frac{\sqrt{\beta_0}}{4 \cdot \sin^2(\pi Q)} \cdot \left[ \int_{s_0}^{s_0+L} (-\Delta k_0(t))\sqrt{\beta(t)} \cdot \sin (\Delta \phi(t))dt - \cos (2\pi Q) \cdot \int_{s_0}^{s_0+L} (-\Delta k_0(t))\sqrt{\beta(t)} \cdot \sin (\Delta \phi(t))dt - \alpha_0 \cdot \sin (2\pi Q) \cdot \int_{s_0}^{s_0+L} (-\Delta k_0(t))\sqrt{\beta(t)} \cdot \sin (\Delta \phi(t))dt + \sin (2\pi Q) \cdot \int_{s_0}^{s_0+L} (-\Delta k_0(t))\sqrt{\beta(t)} \cdot \cos (\Delta \phi(t))dt + \right]
\]
\[
\sin (2\pi Q) \cdot \int_{s_0}^{s_0+L} \alpha_0 \cdot \sin (\Delta \phi(t)) dt
\] (215)

The terms \(a_1\) and \(a_2\) cancel and Eq. (215) simplifies to

\[
c_1 = \left[ (1 - \cos (2\pi Q)) \cdot \int_{s_0}^{s_0+L} (-\Delta k_0(t)) \sqrt{\beta(t)} \cdot \sin (\Delta \phi(t)) dt + \right.
\]
\[
\sin (2\pi Q) \cdot \int_{s_0}^{s_0+L} (-\Delta k_0(t)) \sqrt{\beta(t)} \cdot \cos (\Delta \phi(t)) dt \right] \cdot \frac{\sqrt{\beta_0}}{4 \cdot \sin^2 (\pi Q)} \] (216)

Applying the product rule for trigonometric functions

\[
\sin \alpha \cdot \cos \beta = \frac{1}{2} \cdot [\sin (\alpha - \beta) + \sin (\alpha + \beta)] \] (217)

yields finally

\[
c_1 = \frac{\sqrt{\beta_0}}{4 \cdot \sin^2 (\pi Q)} \cdot \int_{s_0}^{s_0+L} (-\Delta k_0(t)) \sqrt{\beta(t)} \cdot [\sin (\Delta \phi(t)) + \sin (2\pi Q - \Delta \phi)] dt
\]
\[
= \frac{\sqrt{\beta_0}}{4 \cdot \sin^2 (\pi Q)} \cdot 2 \cdot \sin (\pi Q) \cdot \int_{s_0}^{s_0+L} (-\Delta k_0(t)) \sqrt{\beta(t)} \cdot \cos (\Delta \phi(t) - \pi Q) dt
\]
\[
= \frac{\sqrt{\beta_0}}{2 \cdot \sin (\pi Q)} \cdot \int_{s_0}^{s_0+L} (-\Delta k_0(t)) \sqrt{\beta(t)} \cdot \cos (\Delta \phi(t) - \pi Q) dt
\] (218)

Now we are almost finished with our calculation of the closed orbit response. We only have to relate the coefficient \(c_1\) to the closed orbit response and unveil its \(s\) dependence. Using again the boundary conditions (198) at \(s = s_0\) in Eq. (187) we can write the closed orbit solution at \(s = s_0\) as

\[
CO(s = s_0) = c_1(s_0). \] (219)

The reference point \(s_0\) for the calculation of the coefficients \(c_1\) and \(c_2\) is arbitrary. In fact we could have chosen any value for the calculations in Eqs. (197) to (218). Replacing in Eq. (218) \(s_0\) by \(s\) we finally get for the closed orbit response at any position in the storage ring

\[
CO(s) = \frac{\sqrt{\beta(s)}}{2 \cdot \sin (\pi Q)} \cdot \int_{s}^{s+L} (-\Delta k_0(t)) \sqrt{\beta(t)} \cdot \cos (\phi(t) - \phi(s) - \pi Q) dt
\] (220)

Assuming only one isolated dipole error at \(s = s_0\) we obtain for the closed orbit response before \((s = s_0 - \Delta s)\) and after \((s = s_0 + \Delta s)\) the perturbation

\[
CO(s_0 - \Delta s) = \frac{\sqrt{\beta(s)}}{2 \cdot \sin (\pi Q)} \cdot l \cdot (-\Delta k_0(t)) \sqrt{\beta(s_0)} \cdot \cos ([\phi(s_0) - \phi(s_0 - \Delta s)] - \pi Q) \] (221)
\[
CO(s_0 + \Delta s) = \frac{\sqrt{\beta(s)}}{2 \cdot \sin (\pi Q)} \cdot l \cdot (-\Delta k_0(t)) \sqrt{\beta(s_0)} \cdot \cos ([\phi(s_0 + L) - \phi(s_0 + \Delta s)] - \pi Q)
\]
\[
= \frac{\sqrt{\beta(s)}}{2 \cdot \sin (\pi Q)} \cdot l \cdot (-\Delta k_0(t)) \sqrt{\beta(s_0)} \cdot \cos (\pi Q + [\phi(s_0) - \phi(s_0 + \Delta s)])
\]
\[
= \frac{\sqrt{\beta(s)}}{2 \cdot \sin (\pi Q)} \cdot l \cdot (-\Delta k_0(t)) \sqrt{\beta(s_0)} \cdot \cos ([\phi(s_0 + \Delta s) - \phi(s_0)] - \pi Q) \] (222)
where $l$ denotes the length of the dipole field perturbation and where we have used ($\phi(s + L) = \phi(s) + 2\pi Q$) and ($\cos(x) = \cos(-x)$) in the last two transformation. Both cases can be combined into one expression by writing

$$CO(s) = \frac{\sqrt{\beta(s)}}{2 \cdot \sin(\pi Q)} \cdot l \cdot (-\Delta k_0(t)) \sqrt{\beta(s_0)} \cdot \cos(|\phi(s_0) - \phi(s)| - \pi Q). \quad (223)$$

Assuming a half integer tune in Eq. (223) we recuperate the result in Eq. (151).

The total closed orbit response for more than one perturbation is just the linear super imposition of the contributions from all perturbations along the storage ring and we can write the total closed orbit response for distributed dipole perturbations as

$$CO(s) = \frac{\sqrt{\beta(s)}}{2 \cdot \sin(\pi Q)} \cdot \int (-\Delta k_0(t)) \sqrt{\beta(t)} \cdot \cos(|\phi(t) - \phi(s)| - \pi Q) dt. \quad (224)$$

### 3.2 Dispersion orbit

Next we will use Eq. (224) to calculate the dispersion function in a storage ring. The dispersion function $D(s)$ in a storage ring specifies the change in the closed orbit as a function of the particle momentum:

$$CO(s, \Delta p) = \frac{\Delta p}{p_0} \cdot D(s). \quad (225)$$

Inserting the expression of the dipole field error for an off momentum particle, Eq. (130), into Eq. (224) and using $1/\rho(t) = k_0(t)$ one obtains for the dispersion function

$$D_x(s) = \frac{\sqrt{\beta(s)}}{2 \cdot \sin(\pi Q)} \cdot \int \frac{1}{\rho(t)} \sqrt{\beta(t)} \cdot \cos(|\phi(t) - \phi(s)| - \pi Q) dt. \quad (226)$$

Equation (226) confirms our observation in Section 1.7.4 that quadrupole magnets do not directly contribute to the dispersion function since the radius of curvature is infinite in an ideal quadrupole magnet. However, a quadrupole error changes the phase advance in the storage ring and thus, generates a perturbation of the dispersion function. We will come back to this point in Section 3.5 when we treat quadrupole perturbations.

The integral expressions (224) and (226) provide a convenient tool for an analytical estimate of the closed orbit and dispersion perturbation due to dipole field errors. However, while the integral expressions are convenient for 'back of the envelope' calculations they are not well suited for an implementation in numerical tools. Most numerical tools for the twiss function and closed orbit calculation therefore use a matrix approach as discussed in Section 1.8. Starting from the 'one turn' map for the perturbed system one can calculate the closed orbit as the fixed point of the 'one turn' map and the dispersion function as the Eigenvector components for the Eigenvalue 1 of the extended 'one-turn' map.

### 3.3 Tune changes due to quadrupole field errors

We have seen in Eq. (38) that a change in the focusing strength affects the beta function and via Eq. (39) the phase advance along the storage ring. Both quantities are accessible via the transfer matrix $\Phi$ that propagates the particle coordinates from one $s$ value to another. The first quantity we are going to calculate is the change of the total tune in the storage ring in case of quadrupole gradient errors. The 'one turn' map in Eq. (63) offers a direct access to the total tune in the storage ring:

$$\text{Tr}(\Phi(s_0 + L)) = 2 \cdot \cos(2\pi Q) \quad (227)$$

In order to estimate the tune change due to a quadrupole gradient error we need to identify how a quadrupole perturbation changes the matrix elements of the 'one turn' map.
The 'one turn' map of the unperturbed storage ring is given by Eq. (63)

\[ M_0(s, s + L) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \cos (2\pi Q_0) + \begin{pmatrix} \alpha_0(s) & \beta_0(s) \\ -\gamma_0(s) & -\alpha_0(s) \end{pmatrix} \cdot \sin (2\pi Q_0), \]  

(228)

where we have introduced a subscript '0' in order to indicate the unperturbed twiss parameters. Let us assume a quadrupole error occurs only at one location in the storage ring ('s_0') and only over a short element of length 'ds'. The unperturbed transformation matrix of such a thin quadrupole element is given by Eq. (104):

\[ m_0 = \begin{pmatrix} 1 & 0 \\ k_1 ds & 1 \end{pmatrix}. \]  

(229)

The 'one turn' map of the perturbed storage ring can now be written as

\[ M = m \cdot m_0^{-1} \cdot M_0, \]  

(230)

where \( m_0^{-1} \) is the inverse matrix of the unperturbed quadrupole element, given by Eqs. (104) and (188)), and \( m \) the transfer matrix of the perturbed quadrupole element:

\[ m = \begin{pmatrix} 1 & 0 \\ -(k_1 + \Delta k_1) ds & 1 \end{pmatrix}. \]  

(231)

Matrix multiplication yields for the perturbed 'one turn' map

\[ M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \]  

\[ = \begin{pmatrix} \cos (2\pi Q_0) + \alpha_0 \sin (2\pi Q_0) & \beta_0 \sin (2\pi Q_0) \\ -\Delta k_1 ds \cos (2\pi Q_0) - \gamma_0 \sin (2\pi Q_0) & \cos (2\pi Q_0) - (\beta_0 \Delta k_1 ds - \alpha_0) \sin (2\pi Q_0) \end{pmatrix}, \]  

(232)

where all optic functions are evaluated at the location of the perturbation (s = s_0). The tune of the perturbed storage ring is given by the trace of (232):

\[ \cos 2\pi Q = \frac{1}{2} \text{Tr}(M) = \cos (2\pi Q_0) + \frac{1}{2} \beta_0 \Delta k_1 ds \sin (2\pi Q_0). \]  

(233)

For small tune errors \( \Delta Q \) we can write

\[ \cos (2\pi Q) = \cos (2\pi Q_0 + \Delta Q) \approx \cos (2\pi Q_0) \cdot \cos (2\pi \Delta Q) - \sin (2\pi Q_0) \cdot \sin (2\pi \Delta Q) \approx \cos (2\pi Q_0) - \sin (2\pi Q_0) \cdot 2\pi \Delta Q \]

and we get for the tune error

\[ \Delta Q = \frac{1}{4\pi} \cdot \beta_0(s_0) \cdot \Delta k_1(s_0) ds. \]  

(234)

If the perturbation is distributed along the storage ring one obtains the total tune change by integrating Eq. (234) along the s coordinate of the storage ring:

\[ \Delta Q = \frac{1}{4\pi} \int \beta_0(s) \cdot \Delta k_1(s) ds. \]  

(235)

The tune shift is proportional to the quadrupole perturbation and the local beta function at the perturbed quadrupole. This implies that quadrupole elements with large beta functions, for example the quadrupole magnets next to the interaction point of a 'low beta' insertion, are particularly sensitive to gradient errors.
and special care has to be taken to the powering of these elements. A positive aspect of this feature is that one can measure the local beta function inside a quadrupole magnet by changing the gradient of the quadrupole magnet by a known amount. If the change in the normalized quadrupole strength is known one can use Eq. (235) to calculate the local beta function from the measured tune change in the storage ring.

Assuming a storage ring with an equal beta functions in the quadrupole magnets one can deduce from Eq. (235) that random gradient errors in the quadrupole magnets will not change the overall tune in the machine. In a real machine the assumption of equal beta functions is not strictly speaking valid but the conclusion that random quadrupole errors in a regular lattice structure have a much smaller effect on the total machine tune than systematic gradient errors is still true.

3.4 Chromaticity

We will use Eq. (235) to calculate the natural chromaticity in a storage ring. The natural chromaticity specifies the linear momentum dependence of the tune in a storage ring without sextupole magnet contributions:

\[ \frac{\Delta p}{p_0} = Q_0 + \xi_{\text{nat}} \cdot \frac{\Delta p}{p_0} \]  

(236)

A change in the particle momentum implies a change in the normalized quadrupole gradient (25):

\[ k_1 = k_{1,0} + \Delta k_1 \]

\[ = \frac{\Delta k_1}{p_0 + \Delta p} \]

\[ = \frac{\Delta k_1}{p_0} \cdot \frac{1}{1 + \Delta p/p_0} \]

\[ = k_{1,0} - k_{1,0} \cdot \frac{\Delta p}{p_0} + O[(\Delta p/p_0)^2]. \]  

(237)

Inserting the expression for \( \Delta k_1 \) from Eq. (237) into Eq. (235) we obtain for the natural chromaticity

\[ \xi_{\text{nat}} = -\frac{1}{4\pi} \oint \beta_0(s) \cdot k_{1,0}(s) ds. \]  

(238)

In Sections 1.3 and 1.5 we have seen that a storage ring without quadrupole fields has only a weak geometrical focusing and thus, large beta functions and a small 'tune'. Approximating the quadrupole strength and the beta function along the storage ring their average values (smooth approximation) we can use Eq. (39) to obtain the approximation

\[ \xi_{\text{nat}} \approx -\frac{1}{4\pi} \int \frac{1}{\langle \beta \rangle} ds. \]  

(239)

Comparing Eq. (239) with Eq. (39) we obtain

\[ \xi_{\text{nat}} \approx -Q_0. \]  

(240)

The estimate (240) is only a rough approximation but provides a useful estimate for the order of magnitude of the natural chromaticity in a storage ring and shows that the natural chromaticity is always negative. A positive chromaticity can only be generated by sextupole magnets.

In Section 2.2.4 we have seen that a closed orbit displacement inside a sextupole magnet generates an effective quadrupole field error. Using the dispersion orbit for the horizontal beam offset in Eq. (175) we get for the quadrupole feed-down

\[ \Delta k_1(s) = k_2(s) \cdot D_x(s) \cdot \frac{\delta p}{p_0}. \]  

(241)
Sextupole magnets in regions with dispersion therefore generate an additional momentum dependence of the global tune. The sextupole chromaticity specifies the linear momentum dependence of the tune in a storage ring from sextupole magnet contributions:

$$Q \left( \frac{\delta p}{p_0} \right) = Q_0 + \xi_{\text{sext}} \cdot \frac{\delta p}{p_0} \quad (242)$$

Inserting Eq. (241) into Eq. (235) we get for the sextupole chromaticity in a storage ring

$$\xi_{\text{sext}} \approx \frac{1}{4\pi} \oint \beta_0(s) \cdot D_x(s) \cdot k_2(s) ds. \quad (243)$$

The sextupole chromaticity is therefore proportional to the product of the sextupole strength and the local dispersion- and beta functions inside the sextupole magnet. Placing sextupole magnets into regions with dispersion therefore allows to control the chromaticity in the storage ring via an appropriate adjustment of the sextupole strength and thus to either compensate for the natural chromaticity or to generate positive chromaticity values in the storage ring. The later is particularly interesting for the stabilization of collective instabilities (head-tail instability).

### 3.5 β-beat due to quadrupole field errors

Quadrupole gradient errors not only change the tune in a storage ring but also the beta function around the machine. Looking at the perturbed ‘one turn’ map for a single quadrupole error (232) and comparing it with the general ‘one turn’ map expression (228) one observes that the beta function at the location of the perturbation is given by the $M_{12}$ matrix element and the total tune of the machine

$$\beta(s_0) = \frac{M_{12}}{\sin (2\pi Q)} = \frac{\beta_0(s_0) \cdot \sin (2\pi Q_0)}{\sin (2\pi Q)}. \quad (244)$$

Since the perturbed tune on the right-hand side of Eq. (244) depends on the quadrupole gradient error, see Eq. (234), the local beta function also changes due to the quadrupole gradient error. We can therefore proceed similarly as for the tune change calculation by evaluating the perturbed ‘one turn’ map. The perturbed beta function at the location of the quadrupole error can be obtained by inserting Eq. (234) into Eq. (244). However, the while tune is a global storage ring parameter the beta function varies along the storage ring. In order to evaluate the effect of a single quadrupole gradient error on the beta function we not only need to look at the expression (244) at one location in the ring but need to find an expression for the perturbed ‘one turn’ map at any location in the storage ring.

In Section 1.6 we saw that the ‘sine’ and ‘cosine’ like solutions can propagate the particle coordinates in a storage ring from one location to another:

$$\begin{pmatrix} x(s) \\ x'(s) \end{pmatrix} = \Phi(s_0, s) \cdot \begin{pmatrix} x(s_0) \\ x'(s_0) \end{pmatrix}, \quad (245)$$

with

$$\Phi(s_0, s) = \begin{pmatrix} C(s) & S(s) \\ C'(s) & S'(s) \end{pmatrix}. \quad (246)$$

We can use the above property to generate a ‘one turn’ map for any location in the storage ring:

$$M(s, s + L) = \Phi(s_0, s) \cdot M(s_0, s_0 + L) \cdot \Phi^{-1}(s_0, s). \quad (247)$$
Since we assumed only one localized quadrupole gradient error at location \(s_0\) the transport matrices \(\Phi(s_0, s)\) and \(\Phi^{-1}(s_0, s)\) extend only over the unperturbed part of the storage ring and we can use the unperturbed 'sine' and 'cosine' functions from Eqs. (58) and (59) for the following analysis:

\[
\Phi(s_0, s) = \begin{pmatrix}
\frac{\sqrt{\pi} \cos(\Delta \phi) + \alpha_0 \sin(\Delta \phi)}{\sqrt{\beta_0}} & \frac{\sqrt{\beta_0} \sin(\Delta \phi)}{\sqrt{\pi} \beta_0 ds} \\
\frac{(\alpha_0 - \cos(\Delta \phi) - (1 + \alpha_0) \sin(\Delta \phi)}{\sqrt{\beta_0}} + \frac{\sqrt{\beta_0} \cos(\Delta \phi) - \alpha \sin(\Delta \phi)}{\sqrt{\pi} \beta_0 ds}
\end{pmatrix},
\]  

(248)

Using again Eq. (188) we can write the inverse matrix in Eq. (247) as

\[
\Phi^{-1}(s_0, s) = \begin{pmatrix}
S'(s) & -S(s) \\
-C'(s) & C(s)
\end{pmatrix}.
\]  

(249)

Inserting the expressions (61) and (249) into Eq. (247) we obtain after matrix multiplication:

\[
\begin{pmatrix}
M_{11}S' - M_{12}C' + M_{21}S - M_{22}C \\
M_{11}S' - M_{12}C' + M_{21}S - M_{22}C
\end{pmatrix} = \begin{pmatrix}
-\alpha \beta_0 \sin(\Delta \phi) \\
-\alpha \beta_0 \sin(\Delta \phi)
\end{pmatrix}.
\]  

(250)

Comparison with the generic expression of the 'one turn' map now yields

\[
\beta(s) \cdot \sin(2\pi Q) = [\alpha \beta_0 \sin(\Delta \phi)].
\]  

(251)

Using

\[
\sin^2(\Delta \phi) = \frac{1}{2} \cdot (1 - \cos(2\Delta \phi))
\]  

(252)

\[
\cos(\Delta \phi) \cdot \sin(\Delta \phi) = \frac{1}{2} \cdot \sin(2\Delta \phi)
\]  

(253)

and inserting

\[
S' = \beta_0 \cdot \left[\sin(\Delta \phi) + \alpha_0(s_0) \cdot \sin^2(\Delta \phi)\right]
\]  

(254)

\[
S^2 = \beta_0 \cdot \left[\beta_0(s_0) \cdot \sin^2(\Delta \phi)\right]
\]  

(255)

\[
C^2 = \frac{\beta_0(s_0)}{\beta_0(s_0)} \cdot \left[\cos(\Delta \phi) + 2\alpha_0(s_0) \cdot \cos(\Delta \phi) \cdot \sin(\Delta \phi) + \alpha^2(s_0) \cdot \sin(\Delta \phi)\right]
\]  

(256)

and the matrix elements from Eq. (232) into expression (251) we get after some algebra

\[
\beta(s) \cdot \sin(2\pi Q) = \beta_0(s_0) \cdot \Delta k_1 ds \cdot \frac{\cos(2\pi Q_0)}{\sin(2\pi Q_0)} - \frac{1}{4} \beta_0^2(s_0) \cdot (\Delta k_1 ds)^2.
\]  

(257)

Now we only need to find an expression for \(\sin(2\pi Q)\) and solve Eq. (257) for \(\beta(s)\). Using Eq. (233) and \(\sin(x) = \sqrt{1 - \cos^2(x)}\) we get

\[
\sin(2\pi Q) = \sin(2\pi Q_0) \cdot \sqrt{1 - \frac{\beta_0(s_0) \cdot \Delta k_1 ds \cdot \cos(2\pi Q_0)}{\sin(2\pi Q_0)}} - \frac{1}{4} \beta_0^2(s_0) \cdot (\Delta k_1 ds)^2.
\]  

(258)

Keeping only terms up to first order in \((\Delta k_1 ds)^2\) we thus obtain

\[
\frac{1}{\sin(2\pi Q)} = \frac{1}{\sin(2\pi Q_0)} \cdot \left[1 - \frac{1}{2} \cdot \beta_0(s_0) \cdot \Delta k_1 ds \cdot \frac{\cos(2\pi Q_0)}{\sin(2\pi Q_0)} + O((\Delta k_1 ds)^2)\right].
\]  

(259)
Inserting the above expression into Eq. (257) and solving for the relative beta function error \((\beta(s) - \beta_0(s))/\beta_0(s)\) we get
\[
\frac{\Delta \beta(s)}{\beta_0(s)} = -\frac{1}{2 \sin (2\pi Q_0)} \cdot \beta_0(s_0) \cdot \cos (2 \cdot |\phi(s) - \phi(s_0)| - 2\pi Q_0) \cdot \Delta k_1 ds,
\]
(260)
where we have replaced \(\Delta \phi\) by \(\phi(s) - \phi(s_0)\). Eq. (260) gives to first order in the quadrupole gradient error the relative beta function error along the storage ring for a single, localized quadrupole perturbation at location \(s_0\). Since we are limiting our analysis of the beta function perturbation only to first terms linear in the quadrupole gradient error we can add the contributions of various perturbations along the storage ring. Integrating Eq. (260) over \(ds\) along the storage ring circumference we get:
\[
\frac{\Delta \beta(s)}{\beta_0(s)} = -\frac{1}{2 \sin (2\pi Q_0)} \int_s^{s+L} \beta_0(t) \cdot \Delta k_1(t) \cdot \cos (2 \cdot |\phi(s) - \phi(t)| - 2\pi Q_0) \cdot dt.
\]
(261)
which can be rewritten (using the same argumentation as for Eq. (224))
\[
\frac{\Delta \beta(s)}{\beta_0(s)} = -\frac{1}{2 \sin (2\pi Q_0)} \int \beta_0(s_0) \cdot \Delta k_1(t) \cdot \cos (2 \cdot |\phi(s) - \phi(t)| - 2\pi Q_0) \cdot dt.
\]
(262)
Equation (260) shows that the beta function perturbation due to a single quadrupole error propagates at twice the betatron frequency around the storage ring. The super imposition of a fundamental oscillation frequency with a higher harmonic is also called a beating and the beta function perturbation is also referred to as the 'beta-beat'. It is interesting to note that, due to the phase dependent term of the cosine function in Eq. (262), a systematic quadrupole gradient error distributed along the storage ring results in a much small beta function error compared to one single localized error. This is quite interesting in case one intends to change the global tune in a storage ring. We have seen in Eq. (235) that the total tune change in the machine is proportional to the integrated product of the beta function and the change in the quadrupole gradient. Distributing the change in the quadrupole gradient over many magnets along the storage ring rather than using one single quadrupole magnet yields the same tune change at the cost of a much smaller beta function error. Another possibility for a global tune knob in the storage ring is to use two quadrupole magnets at locations with the same beta function value spaced by a phase advance of \(\Delta \phi - \pi/2\). While the total tune change does not depend on the phase between the two quadrupole magnets the 'beta-beat' oscillates at twice the betatron frequency and the 'beta-beat' of two quadrupole magnets spaced by \(\Delta \phi - \pi/2\) cancels.

Another interesting aspect worthwhile noting is that a quadrupole error does not only change the global tune in the storage ring, see Eq. (235), but also the local phase advance. The change in the local phase advance along the storage ring can be calculated via Equations (37)
\[
\phi(s) = \phi_0(s) + \Delta \phi(s)
\]
(263)
\[
= \int_{s_0}^{s} \frac{1}{\beta(\tau)} d\tau + \phi_0
\]
(264)
\[
= \int_{s_0}^{s} \frac{1}{\beta_0(\tau) + \Delta \beta(\tau)} d\tau + \phi_0
\]
(265)
\[
= \int_{s_0}^{s} \frac{1}{\beta_0(\tau) + \Delta \beta(\tau)} d\tau + \phi_0
\]
(266)
\[
= \int_{s_0}^{s} \frac{1}{\beta_0(\tau) + \Delta \beta(\tau)} d\tau + \phi_0
\]
(267)
\[
\approx \int_{s_0}^{s} \frac{1}{\beta_0(\tau)} d\tau + \phi_0 - \int_{s_0}^{s} \frac{\Delta \beta(\tau)}{\beta_0^2(\tau)} d\tau.
\]
(268)
In Section 1.7.4 we stated that a quadrupole magnet does not directly contribute to the dispersion function but that the change in phase advance due to quadrupole errors can generate a perturbation of the
Fig. 18: Schematic layout of the Beam Position Monitor (BPM) and dipole corrector magnet arrangement in a FODO cell. Each focusing quadrupole magnet (QF) is equipped with a horizontal BPM (HBPM) and horizontal dipole corrector magnet (DX) and each defocusing quadrupole magnet (QD) is equipped with a vertical BPM (VBPM) and vertical dipole corrector magnet (DV).

dispersion function. Inserting expression (268) into Eq. (226) we can, in principle, estimate the change in the dispersion function due to quadrupole gradient errors. However, rather than performing a quantitative analysis here we only want to make a general observation. Looking at Eq. (226) we observe that any local perturbation of the dispersion function oscillates with the betatron phase advance along the storage ring. Two localized perturbation of the dispersion function spaced by a phase advance of $\Delta \phi = \pi$ therefore cancel each other along the storage ring. Coming back to our tune knobs in the storage ring we can minimize the perturbation of the dispersion function by using two quadrupole magnets with equal beta function values spaced by a phase advance of $\Delta \phi = \pi$. Earlier in this section we already observed that we can minimize the 'beta-beat' in the machine by using two quadrupole magnets spaced by a phase advance of $\Delta \phi = \pi/2$. A very elegant implementation of a tune knob therefore uses four quadrupole magnets with equal beta functions: two pairs of quadrupoles with a spacing of $\Delta \phi = \pi/2$ spaced by a phase advance of $\Delta \phi = \pi$.

4 Orbit correction

Perturbations to the dipole fields will lead to a change of the closed orbit in the storage ring. The following section addresses the correction of such closed orbit errors in a storage ring by means of additional correction dipole magnets.

4.1 Beam position monitors and orbit corrector elements

One prerequisite for correcting the closed orbit in a storage ring is a measurement of the actual closed orbit in the storage ring. In Section 2 we have seen that quadrupole alignment errors are the most common source for closed orbit distortions. It is therefore desirable to have a Beam Position Monitor (BPM) and a dipole corrector magnet next to each quadrupole magnet. Such a layout allows a local correction of the closed orbit error due to quadrupole misalignments. Since the closed orbit perturbation is proportional to the beta function at the misaligned quadrupole magnet the alignment error of a focusing quadrupole magnet mainly affects the horizontal and the alignment error of a defocusing quadrupole mainly the vertical closed orbit. Most storage rings therefore economize the BPM and dipole corrector installation by featuring only a horizontal BPM and dipole corrector element next to the each focusing and only a vertical BPM and dipole corrector element next to each defocusing quadrupole magnet. Figure 18 shows a schematic picture for such an arrangement in a FODO lattice cell.

4.2 Local orbit bumps

A single dipole field perturbation at the location $s_1$ changes the canonical momentum of the particles

$$\Delta x' = \chi$$

(269)
and excites a trajectory oscillation downstream of the dipole perturbation

\[ x(s) = \sqrt{\beta(s_1)\beta(s)} \cdot \chi(s_1) \cdot \sin (\phi(s) - \phi(s_1)) \]  
\[ x'(s) = \frac{\sqrt{\beta(s_1)}}{\sqrt{\beta(s)}} \cdot \chi(s_1) \cdot \cos (\phi(s) - \phi(s_1)) - \frac{\sqrt{\beta(s_1)}}{\sqrt{\beta(s)}} \cdot \alpha(s) \cdot \sin (\phi(s) - \phi(s_1)) \],

which is nothing else as the ‘sine-like’ solution of the Hills equation. As long as we limit the trajectory oscillation to less than one revolution after the perturbation we do not have to worry about multiple deflections by the dipole field perturbation during consecutive passages. The above trajectory response is therefore valid only for

\[ s_1 \leq s < s_1 + L. \]  

The idea of a closed orbit bump is to apply within one revolution after the first dipole deflection a further deflection via additional dipole magnets that exactly cancel the trajectory response of the first deflection. This approach terminates the trajectory oscillation within one turn and the periodic boundary conditions of the closed orbit response are automatically satisfied by the net trajectory response. The orbit deflection due to the dipole deflections extends only over a limited area of the storage ring and the above procedure allows to either generate local orbit deflections (bumps) or to compensate locally an existing orbit error. In the following we will briefly discuss two options for a closed orbit bump generation:

- the generation of a closed orbit bump with two dipole corrector magnets (→ π-bump)
- the generation of a closed orbit bump with three or four dipole corrector magnets

### 4.2.1 π-Bumps

180° downstream from the initial deflection the trajectory response (270) passes again through zero but with a canonical momentum that has the opposite sign of the initial deflection angle. We denote the location 180° downstream from the initial deflection by \( s_2 \) and can write

\[ x'(s_2) = \chi(s_1) \cdot \frac{\sqrt{\beta(s_1)}}{\sqrt{\beta(s_2)}}. \]  

Provided the storage ring features a second dipole corrector magnet at the location \( s_2 \) we can compensate the trajectory oscillation if the deflection angles of the second dipole magnet satisfies the condition

\[ \chi(s_2) = \frac{\sqrt{\beta(s_1)}}{\sqrt{\beta(s_2)}} \cdot \chi(s_1). \]  

Such a compensation of the trajectory oscillation is called the closure of a closed orbit bump. Figure 19 illustrates this closure schematically for a 90° FODO lattice. The above procedure can, in principle, be used to correct the closed orbit along the storage ring. However, the generation of π-bumps requires a lattice phase advance that is an integer fraction of 180° (→ 90° or 60°) and the closed orbit correction with this method is therefore very sensitive to optics imperfections and phase errors. Furthermore the orbit correction via orbit bumps requires a large number of corrector magnets and is thus quite sensitive to BPM errors. We will see later that in most cases an efficient (and thus more robust) orbit correction can be obtained using only a small subset of the corrector magnets in the storage ring. The generation of π-bumps is therefore limited only to special cases where the phase advance between the dipole corrector magnets is precisely known.
Fig. 19: Schematic picture for the closure of a \(\pi\)-bump. Two orbit corrector magnet spaced by a phase advance of \(\Delta\phi = \pi\) excite two trajectory oscillations. The two trajectory oscillations cancel each other at the exit of the second corrector magnet provided the weighted deflection angle of both corrector magnets is equal. The closed orbit deflection appears in this case only over the section between the two corrector elements.

4.2.2 Three and four corrector bumps

A correction that is more robust against optics and phase errors can be achieved by using more than two dipole corrector magnets. Using, for example, three dipole corrector magnets provides three independent parameters for adjusting the closed orbit at one location and to terminate the trajectory oscillations at the exit of the third corrector dipole. The method can therefore be applied for any phase advance between the corrector magnets. One condition for the cancellation of the trajectory oscillation at the exit of the third corrector dipole is that the trajectory response passes through zero at the third dipole magnet. Denoting the location of the three corrector magnets with \(\phi(3)\) and \(s(3)\) the condition can now be adjusted to generate the desired closed orbit bump amplitude at a given position between the three corrector magnets. Figure 20 illustrates schematically the closure of a three corrector bump for a regular FODO lattice with a phase advance different from 60\(^\circ\) and 90\(^\circ\).

The above condition can be combined with Eq. (276) to generate the following condition for the closure of the three corrector bump:

\[
\chi(s_3) = \left( \frac{\sin(\phi(s_3) - \phi(s_1))}{\tan(\phi(s_3) - \phi(s_2))} - \cos(\phi(s_3) - \phi(s_1)) \right) \cdot \frac{\sqrt{\beta(s_1)\beta(s_3)}}{\sqrt{\beta(s_2)}} \cdot \chi(s_1). \tag{278}
\]

Note that the \(\alpha(s)\) dependence in Eq. (277) canceled after the combination with condition (276). The free parameter \(\chi(s_1)\) can now be adjusted to generate the desired closed orbit bump amplitude at a given position between the three corrector magnets. Figure 20 illustrates schematically the closure of a three corrector bump for a regular FODO lattice with a phase advance different from 60\(^\circ\) and 90\(^\circ\).

The above procedure can, in principle, be used to correct the closed orbit along the storage ring. However, as for the case of the \(\pi\)-bumps it requires a large number of corrector magnets and is thus
sensitive to BPM errors. The three corrector bump is therefore mainly used for adjusting the closed orbit offset at dedicated locations in the storage ring. For example, this can be quite useful for steering the beam around localized aperture limitations inside the storage ring or for steering the closed orbit in the injection and extraction regions of the storage ring.

In some cases it is not only important to adjust the closed orbit but also the slope of the closed orbit in a certain location of the storage ring. Such an operation requires one more parameter (2 for setting $x$ and $x'$ to zero at the last corrector magnet and 2 for the adjustment of $x$ and $x'$ at the desired position within the bump) and can be achieved with four corrector dipole magnets. Following the same line of argumentation as in the previous cases one can derive the following condition for the **four corrector bump** closure. Denoting the location of the four corrector magnets with $s_1$, $s_2$, $s_3$ and $s_4$ and the target closed orbit amplitude and slope at the position $s_b$ with $x_b$ and $x'_b$ and assuming that the position $s_b$ lies in between the position of the second and third corrector magnet ($s_2 \leq s_b \leq s_3$) one can write the condition for closure as [11]:

$$
\chi(s_1) = \frac{1}{\sqrt{\beta(s_1)\beta(s_b)}} \cdot \frac{\cos(\phi(s_b) - \phi(s_2)) - \alpha(s_b) \cdot \sin(\phi(s_b) - \phi(s_2))}{\sin(\phi(s_2) - \phi(s_1))} \cdot x_b
$$

$$
\chi(s_2) = \frac{1}{\sqrt{\beta(s_2)\beta(s_b)}} \cdot \frac{\cos(\phi(s_b) - \phi(s_1)) - \alpha(s_b) \cdot \sin(\phi(s_b) - \phi(s_1))}{\sin(\phi(s_2) - \phi(s_1))} \cdot x_b
\quad + \frac{\sqrt{\beta(s_b)}}{\sqrt{\beta(s_2)}} \cdot \frac{\sin(\phi(s_b) - \phi(s_1))}{\sin(\phi(s_2) - \phi(s_1))} \cdot x'_b
$$

$$
\chi(s_3) = \frac{1}{\sqrt{\beta(s_3)\beta(s_b)}} \cdot \frac{\cos(\phi(s_b) - \phi(s_4)) - \alpha(s_b) \cdot \sin(\phi(s_b) - \phi(s_4))}{\sin(\phi(s_4) - \phi(s_3))} \cdot x_b
\quad - \frac{\sqrt{\beta(s_b)}}{\sqrt{\beta(s_3)}} \cdot \frac{\sin(\phi(s_b) - \phi(s_4))}{\sin(\phi(s_4) - \phi(s_3))} \cdot x'_b
$$

$$
\chi(s_4) = \frac{1}{\sqrt{\beta(s_4)\beta(s_b)}} \cdot \frac{\cos(\phi(s_b) - \phi(s_4)) - \alpha(s_b) \cdot \sin(\phi(s_b) - \phi(s_4))}{\sin(\phi(s_4) - \phi(s_3))} \cdot x_b
\quad + \frac{\sqrt{\beta(s_b)}}{\sqrt{\beta(s_4)}} \cdot \frac{\sin(\phi(s_b) - \phi(s_4))}{\sin(\phi(s_4) - \phi(s_3))} \cdot x'_b.
$$

The adjustment of the closed orbit slope at a given location implies the consideration of the additional ‘twiss’ parameter $\alpha(s_b)$ at the location of the orbit adjustment. In the previous cases of orbit bumps the
\( \alpha \) parameter canceled and did not enter in the final condition for the bump closure. Examples for ‘four corrector bumps’ include the closed orbit adjustments at the injection and extraction kicker and septa magnets and the generation of dedicated crossing angle bumps at the interaction points of a collider storage ring with a large number of bunches. Figure 21 illustrates schematically the closure of a four corrector bump for a regular FODO lattice with a phase advance different from 60° and 90°.

### 4.3 Harmonic filtering

Next we want to illustrate that one only requires a small number of parameters for the correction of a random orbit distortion. For this purpose we come back to the ‘smooth approximation’ of the betatron oscillation with constant beta functions and a constant focusing strength along the storage ring. The equation of motion reduce to the equation of a Harmonic oscillator in this case

\[
x'' + \left( \frac{2\pi Q}{L} \right)^2 \cdot x = 0
\]

and we can write the solutions for the betatron oscillations as

\[
x(s) = A \cdot e^{i \frac{2\pi Q}{L} \cdot s}.
\]

In case of localized dipole perturbations along the ring the equation of motion can be written as an inhomogeneous second order differential equation:

\[
x'' + \left( \frac{2\pi Q}{L} \right)^2 \cdot x = \sum_{i=1}^{m} \chi_i \cdot \delta(s - s_i)
\]

where we have introduced the \( \delta \) function which attains the value 1 when the argument is zero and vanishes everywhere else. The periodicity of the closed orbit and the perturbations along the storage ring imply that we can expand the closed orbit solution and the dipole perturbations into a Fourier series with the machine circumference as the fundamental period:

\[
x(s) = \sum_{n} d_n \cdot e^{\frac{2\pi Q}{L} \cdot n \cdot s}
\]

\[
F(s) = \sum_{n} f_n \cdot e^{\frac{2\pi Q}{L} \cdot n \cdot s}.
\]
Fig. 22: Schematic picture for the Fourier coefficient response of the closed orbit to an excitation of a random distribution of dipole deflections along the storage ring. The Fourier coefficients peak for index close to the integer tune of the storage ring. The presented example features a tune of $Q = 64.28$.

Inserting the Ansatz (287) into the equation of motion we can solve for the Fourier coefficients of the closed orbit:

\[ d_n = \frac{f_n}{\left(\frac{2\pi Q}{L}\right)^2 - \left(\frac{2\pi n}{L}\right)^2}. \]  

From Eq. (289) we conclude that the Fourier series of the closed orbit response is dominated by the coefficients who’s index is close to the integer machine tune:

\[ n \approx Q. \]  

Figure 22 shows the Fourier coefficients of the closed orbit response for a random dipole perturbation along a storage ring with an tune of $Q = 64.28$ and illustrates that the closed orbit response to a random perturbation is dominated by a small number of coefficients. This in turn implies that an correction should be possible using only a small number of parameters. For example, if the orbit corrector magnets in the storage ring are powered to circuits that drive individual harmonics of the orbit response (such circuits are also called harmonic bumps) one only needs to use a small number of such knobs for a global orbit correction. A more detailed description of the Harmonic filtering can be found in [13].

4.4 Least square fit

The closed orbit responds linearly to the dipole corrector excitation. One could therefore attempt to generate a response matrix $A$ for the closed orbit as a function of the corrector excitation and to use the inverse matrix for identifying the required corrector powering that minimizes the closed orbit in the storage ring. Defining a vector

\[ \vec{COR} = (a_1, a_2, \ldots, a_m) \]  

that contains the strength of all corrector magnets in the storage ring and a vector

\[ BPM = (b_1, b_2, \ldots, b_n) \]  

that contains all BPM readings in the storage ring one can write the equation for the orbit correction as

\[ BPM - A \cdot \vec{COR} = 0 \]  

\[ (293) \]
where $A$ is the $n \times m$ response matrix to the corrector powering. One could be tempted to write the solution to the above problem as

$$\vec{COR} = A^{-1} \cdot \vec{BPM},$$

(294)

where $A^{-1}$ is the inverse of the response matrix $A$. Unfortunately the response matrix is usually not invertible. In most cases it is not even a square matrix one can therefore only try to minimize the norm

$$||B\vec{PM} - A \cdot \vec{COR}||_p,$$

(295)

where the $p$-norm of vector of length $n$ is defined as

$$||\vec{X}||_p = \left( \sum_{i=1}^{n} x_i^p \right)^{1/p}$$

(296)

For $p = 2$ we obtain the square norm and a Least Square Fit aims at calculating a matrix $B$ such that

$$||B\vec{PM} - A \cdot \vec{B} \cdot B\vec{PM}||_2,$$

(297)

attains a minimum. The vector $\vec{B} \cdot B\vec{PM}$

(298)

defines the required corrector settings for a given closed orbit reading. The quadratic norm can be minimized by a pseudo inversion using [12]

$$B = (A^T \cdot A)^{-1} \cdot A^T,$$

(299)

where $A^T$ is the transpose of the response matrix $A$. The required corrector powering can then be written as

$$\vec{COR} = (A^T \cdot A)^{-1} \cdot B\vec{PM},$$

(300)

The main advantage of the Least Square algorithm is that it is straight forward and that matrix inversion algorithms are easily available and fast enough to cope even with large storage rings (large number of BPMs and orbit corrector elements). The main disadvantages of the algorithm is that it always uses all available orbit corrector magnets and thus is sensitive to wrong measurements due to BPM errors. Furthermore the procedure can generate local bumps in case some orbit corrector elements are not independent. This case can occur, for example, if some BPMs in the storage ring are not working and have to be removed from the orbit readings. In this case two columns (or rows) become linearly dependent and the matrix $A$ is called rank deficient. In a real machine the phase advance between two orbit corrector magnets is rarely exactly equal to $n \cdot \pi$ but might have a value that is very close to it. In this case the matrix $A$ is not exactly rank deficient but ill-conditioned.

### 4.5 Singular value decomposition

One procedure for analysing if the matrix $A$ is ill-conditioned is to decompose it into a product of two orthogonal and one ‘diagonal’ matrix:

$$A = O_1 \cdot D \cdot O_2,$$

(301)

where $O_1$ and $O_2$ are two orthogonal matrices and $D$ a diagonal matrix

$$D = \begin{pmatrix}
\sigma_{11} & 0 & \cdots & 0 \\
0 & \sigma_{22} & \cdots & 0 \\
0 & 0 & \sigma_{33} & \cdots \\
\vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & \sigma_{kk}
\end{pmatrix}$$

(302)
The matrix $D$ has $k$ non-vanishing diagonal elements where $k$ satisfies

$$k \leq \min(n,m).$$ (303)

The singular value decomposition (SVD) defines an algorithm for calculating a above matrices for a given matrix $A$ [12]. The diagonal elements $\sigma_i$ provide a measure for how ‘ill-conditioned’ the matrix $A$ is. If one or more of the diagonal elements $\sigma_i$ are zero the matrix $A$ is 'rank-deficient'. If one or more of the diagonal elements $\sigma_i$ are close to zero the matrix $A$ is 'ill-conditioned'.

The problem of small eigenvalues can be illustrated if we construct the pseudo inverse matrix in Eq. (299) with the help of the diagonal and orthogonal matrices in Eq. (301). If we define a matrix

$$\hat{D} = \begin{pmatrix}
1/\sigma_{11} & 0 & \ldots & 0 \\
0 & 1/\sigma_{22} & 0 & \ldots & 0 \\
0 & 0 & 1/\sigma_{33} & 0 & \ldots & 0 \\
& \vdots & \ddots & \vdots \\
0 & \ldots & 0 & 1/\sigma_{kk} \\
0 & \ldots & \ldots & \ldots \\
0 & \ldots & \ldots & \ldots \\
0 & \ldots & \ldots & \ldots \\
0 & \ldots & \ldots & \ldots \\
0 & \ldots & \ldots & \ldots \\
\end{pmatrix}$$ (304)

it satisfies

$$D \cdot \hat{D} = \begin{pmatrix}
1 & 0 & \ldots & 0 \\
0 & 1 & 0 & \ldots & 0 \\
0 & 0 & 1 & 0 & \ldots & 0 \\
& \vdots & \ddots & \vdots \\
0 & \ldots & 0 & 1 \\
\end{pmatrix}$$ (305)

and we can define the correction matrix $B$

$$B = O_1^T \cdot \hat{D} \cdot O_1^T$$ (306)

that minimizes the quadratic norm (297) and defines the the required corrector settings for a given closed orbit reading via:

$$C\overline{R} = A \cdot B \cdot BF.M.$$ (307)

The matrices $O_1^T$ and $O_2^T$ are the transposed matrices of the orthogonal matrices $O_1$ and $O_2$ in Eq. (301).

From Equation refinverse-svd it becomes obvious that small diagonal values in the matrix $D$ imply a strong amplification of orbit reading errors during the calculation of the required corrector strength. In the case of zero diagonal values the procedure will generate infinite corrector values. The ‘Least Square Fit’ algorithm is therefore not stable if the matrix $D$ has small eigenvalues. In order to cope with such a situation one can use the SVD algorithm to identify the redundant corrector magnets and to remove them from the orbit correction algorithm. However, apart from the removed redundant corrector magnets the orbit correction via SVD (effectively a ‘Least Square Fit’ with a conditioned matrix) still uses all remaining corrector magnets while we have seen in Section 4.3 that an efficient orbit correction requires only the use of a small subset of the corrector magnets.

4.6 Most effective corrector algorithm

Another method for the closed orbit correction is to minimize the norm (297) by an empirical adjustment of a small subset of orbit corrector elements. A brute force approach of selecting randomly a subset of
corrector magnets and evaluating the correction efficiency for all possible subsets is quite time consuming and therefore not a practical approach. A better approach is to look first for the most effective corrector in the machine. After an adjustment one keeps the corrector strength of this first element fixed and looks for the next most effective corrector magnet. This approach converges much faster but contains the risk of missing the best correction and of generating local orbit bumps in the machine if there is a set of redundant corrector elements. This risk can be avoided by using the above selective approach and performing a cross correlation between the closed orbit residues and the remaining corrector elements. This approach is implemented in the MICADO algorithm that is widely used in the operation of storage rings [12].

Figure 23 shows an example of a measured closed orbit in the CERN LEP storage ring before and after correction [14] with the MICADO algorithm.

References

[14] Courtesy of Jörg Wenninger CERN.
Fig. 23: Vertical closed orbit in LEP before (top) and after (bottom) orbit correction.
Basic principles of RF superconductivity and superconducting cavities

P. Schmüser
Institut für Experimentalphysik, Universität Hamburg, Germany

Abstract
The basics of superconductivity are outlined with special emphasis on the features which are relevant for the application of superconductors in radio frequency cavities for particle acceleration. For a cylindrical resonator ("pill box cavity") the electromagnetic field in the cavity and important parameters such as resonance frequency, quality factor and shunt impedance are calculated analytically. The design and performance of practical cavities is shortly addressed.

1 Introduction
1.1 Advantages and limitations of superconductor technology in accelerators
The vanishing electrical resistance of superconducting coils as well as their ability to provide magnetic fields far beyond those of saturated iron is the main motivation for using superconducting (sc) magnets in all new large proton, antiproton and heavy ion accelerators. Superconductivity does not only open the way to much higher particle energies but at the same time leads to a substantial reduction of operating costs. In the normal-conducting Super Proton Synchrotron SPS at CERN a power of 52 MW is needed to operate the machine at an energy of 315 GeV while at HERA a cryogenic plant with 6 MW electrical power consumption is sufficient to provide the cooling of the superconducting magnets with a stored proton beam of 920 GeV. Hadron energies in the TeV regime are practically inaccessible with standard magnet technology. Another important application of superconducting materials is in the large experiments at hadron or lepton colliders where superconducting detector magnets are far superior to normal magnets.

In the case of accelerating cavities the advantage of superconductors is not at all that obvious. In fact, three of the proposed linear electron positron colliders are based on copper acceleration structures: the ‘Next Linear Collider’ NLC [3] at Stanford, the ‘Japanese Linear Collider’ JLC [4] at Tsukuba, and the ‘Compact Linear Collider’ CLIC [5] at CERN, while only the international TESLA project [6, 7] uses sc niobium cavities. The traditional arguments against superconductor technology in linear colliders have been the low accelerating fields achieved in sc cavities and the high cost of cryogenic equipment. Superconducting cavities face a strong physical limitation: the microwave magnetic field must stay below the critical field of the superconductor. For the best superconductor for cavities, niobium, this corresponds to a maximum accelerating field of about 45 MV/m while normal-conducting cavities operating at high frequency (above 5 GHz) should in principle be able to reach 100 MV/m or more. In practice, however, sc cavities were often found to be limited at much lower fields of some 5 MV/m and hence were totally non-competitive for a linear collider. Great progress was achieved with the 340 five-cell cavities of the Continuous Electron Beam Accelerator Facility CEBAF [8] at Jefferson Laboratory in Virginia, USA. These 1.5 GHz niobium cavities were developed at Cornell University and produced by industry. They exceeded the design gradient of 5 MV/m and achieved 8.4 MV/m after installation in the accelerator (in several specially prepared cavities even 15-20 MV/m were reached). Building upon the CEBAF experience the intensive R&D of the TESLA collaboration has succeeded in raising the accelerating field in multicell cavities to more than 25 MV/m. There is a realistic chance to reach even 35 MV/m, and to reduce substantially the cost for the cryogenic installation.

* Adapted from a review article [1].
1 The parameters of high energy lepton and hadron colliders are summarized in [2].
While superconducting magnets operated with direct current are free of energy dissipation, this is not the case in microwave cavities. The non-superconducting electrons (see sect. 2) experience forced oscillations in the time-varying magnetic field and dissipate power in the material. Although the resulting heat deposition is many orders of magnitude smaller than in copper cavities it constitutes a significant heat load on the refrigeration system. As a rule of thumb, 1 W of heat deposited at 2 K requires almost 1 kW of primary ac power in the refrigerator. Nevertheless, there is now a worldwide consensus that the overall efficiency for converting primary electric power into beam power is about a factor two higher for a superconducting than for a normal-conducting linear collider with optimized parameters in either case [9]. Another definite advantage of a superconducting collider is the low resonance frequency of the cavities that can be chosen (1.3 GHz in TESLA). The longitudinal (transverse) wake fields generated by the ultrashort electron bunches upon passing the cavities scale with the second (third) power of the frequency and are hence much smaller in TESLA than in NLC \((f = 11 \text{ GHz})\). The wake fields may have a negative impact on the beam emittance (the area occupied in phase space) and on the luminosity of the collider.

1.2 Characteristic properties of superconducting cavities

The fundamental advantage of superconducting niobium cavities is the extremely low surface resistance of a few nano-ohms at 2 Kelvin as compared to several milli-ohms in copper cavities. The quality factor \(Q_0\) (\(2\pi\) times the ratio of stored energy to energy loss per cycle) is inversely proportional to the surface resistance and may exceed \(10^{10}\). Only a tiny fraction of the incident radio frequency (rf) power is dissipated in the cavity walls, the lion’s share is transferred to the beam. The physical limitation of a sc resonator is given by the requirement that the rf magnetic field at the inner surface has to stay below the critical field of the superconductor (about 190 mT for niobium), corresponding to an accelerating field of \(E_{\text{acc}} = 45 \text{ MV/m}\). In principle the quality factor should stay constant when approaching this fundamental superconductor limit but in practice the curve \(Q_0 = Q_0(E_{\text{acc}})\) ends at considerably lower values, often accompanied with a strong decrease of \(Q_0\) towards the highest gradient reached in the cavity. The main reasons for the performance degradation are excessive heating caused by impurities on the inner surface or by field emission of electrons. The cavity becomes partially normal-conducting, associated with strongly enhanced power dissipation. Because of the exponential increase of surface resistance with temperature this may result in a run-away effect and eventually a quench of the entire cavity.

Field emission of electrons from sharp tips is the most severe limitation in high-gradient superconducting cavities. Small particles on the cavity surface act as field emitters. By applying the clean room techniques developed in semiconductor industry it has been possible to raise the threshold for field emission in multicell cavities from about 10 MV/m to more than 20 MV/m in the past few years. The preparation of a smooth and almost mirror-like surface by electrolytic polishing is another important improvement. A detailed description of sc cavities is found in [10].

2 Basics of superconductivity

The unusual features of superconducting magnets and cavities are closely linked to the physical properties of the superconductor itself. For this reason a basic understanding of superconductivity is indispensable for the design, construction and operation of superconducting accelerator components. Only the traditional ‘low-temperature’ superconductors are treated since up to date the use of ‘high-temperature’ ceramic superconductors in these devices is rather limited [11, 12]. For more comprehensive presentations I refer to the excellent text books by W. Buckel [13] and by D.R. Tilley and J. Tilley [14].
2.1 Overview
Superconductivity — the infinitely high conductivity below a ‘critical temperature’ $T_c$ — is observed in a large variety of materials but, remarkably, not in some of the best normal conductors like copper, silver and gold, except at very high pressures. This is illustrated in Fig. 1 where the resistivity of copper, tin and the ‘high-temperature’ superconductor YBa$_2$Cu$_3$O$_7$ is sketched as a function of temperature. Table 1 lists some important superconductors together with their critical temperatures at vanishing magnetic field.

![Fig. 1: The low-temperature resistivity of copper, tin and YBa$_2$Cu$_3$O$_7$.](image)

Table 1: Critical temperature $T_c$ in K of selected superconducting materials for vanishing magnetic field.

<table>
<thead>
<tr>
<th>Material</th>
<th>$T_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al</td>
<td>1.14</td>
</tr>
<tr>
<td>Hg</td>
<td>4.15</td>
</tr>
<tr>
<td>Sn</td>
<td>3.72</td>
</tr>
<tr>
<td>Pb</td>
<td>7.9</td>
</tr>
<tr>
<td>Nb</td>
<td>9.2</td>
</tr>
<tr>
<td>Ti</td>
<td>0.4</td>
</tr>
<tr>
<td>NbTi</td>
<td>9.4</td>
</tr>
<tr>
<td>Nb$_3$Sn</td>
<td>18</td>
</tr>
</tbody>
</table>

There is an intimate relation between superconductivity and magnetic fields. W. Meissner and R. Ochsenfeld discovered in 1933 that a superconducting element like lead expels a weak magnetic field from its interior when cooled below $T_c$, while in stronger fields superconductivity breaks down and the material goes to the normal state. The spontaneous exclusion of magnetic fields upon crossing $T_c$ cannot be explained in terms of the Maxwell equations of classical electrodynamics and indeed turned out to be of quantum-theoretical origin. In 1935 H. and F. London proposed an equation which offered a phenomenological explanation of the field exclusion. The London equation relates the supercurrent density $J_s$ to the magnetic field:

$$ \nabla \times \vec{J}_s = -\frac{n_s e^2}{m_e} \vec{B} $$  \hspace{1cm} (1)

where $n_s$ is the density of the super-electrons. In combination with the Maxwell equation $\nabla \times \vec{B} = \mu_0 \vec{J}_s$ we get the following equation for the magnetic field in a superconductor

$$ \nabla^2 \vec{B} = \frac{\mu_0 n_s e^2}{m_e} \vec{B} = 0 $$ \hspace{1cm} (2)

For a simple geometry, namely the boundary between a superconducting half space and vacuum, and with a magnetic field parallel to the surface, Eq. (2) reads

$$ \frac{d^2 B_y}{dx^2} - \frac{1}{\lambda_L^2} B_y = 0 \quad \text{with} \quad \lambda_L = \sqrt{\frac{m_e}{\mu_0 n_s e^2}}. $$ \hspace{1cm} (3)

Here we have introduced a very important superconductor parameter, the London penetration depth $\lambda_L$. The solution of the differential equation is

$$ B_y(x) = B_0 \exp(-x/\lambda_L). $$ \hspace{1cm} (4)
So the magnetic field does not abruptly drop to zero at the superconductor surface but penetrates into the material with exponential attenuation (Fig. 2). For typical material parameters the penetration depth is quite small, namely 20–50 nm. In the bulk of a thick superconductor the magnetic field vanishes which is just the Meissner-Ochsenfeld effect.

The justification of the London equation remained obscure until the advent of the microscopic theory of superconductivity by Bardeen, Cooper and Schrieffer in 1957. The BCS theory is based on the assumption that the supercurrent is not carried by single electrons but rather by pairs of electrons of opposite momenta and spins, the so-called Cooper pairs. The London penetration depth remains invariant under the replacements \( n_s \rightarrow n_c = n_s/2 \), \( e \rightarrow 2e \) and \( m_e \rightarrow m_c = 2m_e \). The BCS theory revolutionized our understanding of superconductivity. All Cooper pairs occupy a single quantum state, the BCS ground state, whose energy is separated from the single-electron states by a temperature dependent energy gap \( E_g = 2\Delta(T) \). The critical temperature is related to the energy gap at \( T = 0 \) by

\[
1.76 k_B T_c = \Delta(0) .
\]  

(5)

Here \( k_B = 1.38 \cdot 10^{-23} \text{ J/K} \) is the Boltzmann constant. The magnetic flux through a superconducting ring is found to be quantized, the smallest unit being the elementary flux quantum

\[
\Phi_0 = \frac{h}{2e} = 2.07 \cdot 10^{-15} \text{ Vs} .
\]  

(6)

These and many other predictions of the BCS theory, like the temperature dependence of the energy gap and the existence of quantum interference phenomena, have been confirmed by experiment and often found practical application.

A discovery of enormous practical consequences was the finding that there exist two types of superconductors with rather different response to magnetic fields. The elements lead, mercury, tin, aluminium and others are called ‘type I’ superconductors. They do not admit a magnetic field in the bulk material and are in the superconducting state provided the applied field stays below a critical field \( H_c(B_c = \mu_0 H_c) \) is usually less than 0.1 Tesla). All superconducting alloys like lead-indium, niobium-titanium, niobium-tin and also the element niobium belong to the large class of ‘type II’ superconductors. They are characterized by two critical fields, \( H_{c1} \) and \( H_{c2} \). Below \( H_{c1} \) these substances are in the Meissner phase with complete field expulsion while in the range \( H_{c1} < H < H_{c2} \) they enter the mixed phase in which the magnetic field pierces the bulk material in the form of flux tubes. Many of these materials remain superconductive up to much higher fields (10 Tesla or more).

### 2.2 Energy balance in a magnetic field

A material like lead makes a phase transition from the normal to the superconducting state when it is cooled below \( T_c \) and when the magnetic field is less than \( H_c(T) \). This is a phase transition comparable to
the transition from water to ice below 0°C. Phase transitions take place when the new state is energetically favoured. The relevant thermodynamic energy is here the so-called Gibbs free energy $G$. Free energies have been measured for a variety of materials. For temperatures $T < T_c$ they are found to be lower in the superconducting than in the normal state while $G_{\text{sup}}$ approaches $G_{\text{norm}}$ in the limit $T \to T_c$, see Fig. 3a. What is now the impact of a magnetic field on the energy balance? A magnetic field has an energy density $\mu_0/2 \cdot H^2$, and according to the Meissner-Ochsenfeld effect the magnetic energy must be pushed out of the material when it enters the superconducting state. Hence the free energy per unit volume in the superconducting state increases quadratically with the applied field:

$$G_{\text{sup}}(H) = G_{\text{sup}}(0) + \frac{\mu_0}{2} H^2.$$  

The normal-state energy remains unaffected. The material stays superconductive as long as $G_{\text{sup}}(H) < G_{\text{norm}}$. Equation (7) implies the existence of a maximum tolerable field, the ‘critical field’, above which superconductivity breaks down. It is defined by the condition that the free energies in the superconducting and in the normal state be equal

$$G_{\text{sup}}(H_c) = G_{\text{norm}} \Rightarrow \frac{\mu_0}{2} H_c^2 = G_{\text{norm}} - G_{\text{sup}}(0).$$  

Equation (8) is also meaningful for type II superconductors and defines in this case the thermodynamic critical field which lies between $H_{c1}$ and $H_{c2}$. The quantity $\mu_0/2 \cdot H_c^2 = G_{\text{norm}} - G_{\text{sup}}(0)$ can be interpreted as the Cooper-pair condensation energy per unit volume.

### 2.3 Coherence length and distinction between type I and type II superconductors

In very thin sheets of superconductor (thickness $< \lambda_c$) the magnetic field does not drop to zero at the centre. Consequently less magnetic energy needs to be expelled which implies that the critical field of a thin sheet may be much larger than the $H_c$ of a thick slab. From this point of view it might appear energetically favourable for a thick slab to subdivide itself into an alternating sequence of thin normal and superconducting slices. The magnetic energy is indeed lowered that way but there is another energy to be taken into consideration, namely the energy required to create the normal-superconductor interfaces. At the boundary between the normal and the superconducting phase the density $n_c$ of the super-current carriers (the Cooper pairs) does not jump abruptly from zero to its value in the bulk but rises smoothly over a finite length $\xi$, called coherence length, see Fig. 2.
The relative size of the London penetration depth $\lambda_L$ and the coherence length $\xi$ decides whether a material is a type I or a type II superconductor. Creation of a boundary means a loss of Cooper-pair condensation energy in a thickness $\xi$ but a gain of magnetic energy in a thickness $\lambda_L$. There is a net energy gain if $\lambda_L > \xi$. So a subdivision of the superconductor into an alternating sequence of thin normal and superconducting slices is energetically favourable if the London penetration depth exceeds the coherence length.

A more refined treatment is provided by the Ginzburg-Landau theory (see e.g. [14]). Here one introduces the Ginzburg-Landau parameter

$$\kappa = \frac{\lambda_L}{\xi}. \quad (9)$$

The criterion for type I or II superconductivity is found to be

- type I: $\kappa < \frac{1}{\sqrt{2}}$
- type II: $\kappa > \frac{1}{\sqrt{2}}$.

Table 2 lists the penetration depths and coherence lengths of some important superconducting elements. Niobium is a type II conductor but close to the border to type I, while indium, lead and tin are clearly in the type I class. The coherence length $\xi$ is proportional to the mean free path of the conduction electrons in the metal. In alloys the mean free path is generally much shorter than in pure metals hence alloys are always type II conductors.

In reality a type II superconductor is not subdivided into thin slices but the field penetrates the sample in the form of flux tubes which arrange themselves in a triangular pattern which can be made visible by evaporating iron atoms onto a superconductor surface sticking out of the liquid helium. The fluxoid pattern shown in Fig. 4a proves beyond any doubt that niobium is indeed a type II superconductor. Each flux tube or fluxoid contains one elementary flux quantum $\Phi_0$ which is surrounded by a Cooper-pair vortex current. The centre of a fluxoid is normal-conducting and covers an area of roughly $\pi \xi^2$.

When we apply an external field $H$, fluxoids keep moving into the specimen until their average magnetic flux density is identical to $B = \mu_0 H$. The fluxoid spacing in the triangular lattice $d = \sqrt{2\Phi_0/(\sqrt{3}B)}$ amounts to 20 nm at 6 Tesla. The upper critical field is reached when the current vortices of the fluxoids start touching each other at which point superconductivity breaks down. In the Ginzburg-Landau theory the upper critical field is given by

$$B_{c2} = \sqrt{2} \kappa B_c = \frac{\Phi_0}{2\pi \xi^2}. \quad (10)$$

For niobium-titanium with an upper critical field $B_{c2} = 14$ T this formula yields $\xi = 5$ nm. The coherence length is larger than the typical width of a grain boundary in NbTi which means that the supercurrent can freely move from grain to grain. In high-$T_c$ superconductors the coherence length is often shorter than the grain boundary width, and then current flow from one grain to the next is strongly impeded.
2.4 Flux flow resistance and flux pinning

For application in accelerator magnets, a superconducting wire must be able to carry a large current in the presence of a field of 5 – 10 Tesla. Type I superconductors are definitely ruled out because their critical field is far too low (below 0.1 Tesla). Type II conductors appear promising at first sight: they feature large upper critical fields, and high currents are permitted to flow in the bulk material. However, there is the problem of flux flow resistance. A current flowing through an ideal type II superconductor, which is exposed to a magnetic field, exerts a Lorentz force on the flux lines and causes them to move through the specimen, see Fig. 4b. This is a viscous motion and leads to heat generation. So although the current itself flows without dissipation, the sample acts as if it had an Ohmic resistance. The statement is even formally correct. The moving fluxoids represent a moving magnetic field which, according to theory of special relativity, is equivalent to an electric field \( \vec{E}_{\text{equiv}} = \vec{B} \times \vec{v}/c^2 \). It is easy to see that \( \vec{E}_{\text{equiv}} \) and \( \vec{J} \) point in the same direction just like in a normal resistor. To obtain useful wires for magnet coils, flux flow has to be prevented by capturing the fluxoids at pinning centres. These are defects or impurities in the regular crystal lattice. The most important pinning centres in niobium-titanium are normal-conducting titanium precipitates in the so-called \( \alpha \) phase whose size is in the range of the fluxoid spacing (≈ 10 nm at 6 Tesla). Figure 5 shows a microscopic picture of a conductor with very high current density (3700 A/mm\(^2\) at 5 T and 4.2 K).

Fig. 4: (a) Fluxoid pattern in niobium (courtesy U. Essmann). The distance between adjacent flux tubes is 0.2 \( \mu \text{m} \). (b) Fluxoid motion in a current-carrying type II superconductor.

Fig. 5: Micrograph of NbTi. The \( \alpha \)-titanium precipitates appear as lighter strips. The area covered is 840 nm wide and 525 nm high. Courtesy P.J. Lee and D.C. Larbalestier.

A type II superconductor with strong pinning is called a hard superconductor. Hard superconductors are very well suited for high-field magnets, they permit dissipationless current flow in high magnetic fields. There is a penalty, however: these conductors exhibit a strong magnetic hysteresis which is the origin of the very annoying 'persistent-current' multipoles in superconducting accelerator magnets.
2.5 Magnetization of a hard superconductor

A type I superconductor shows a reversible response to a varying external magnetic field $H$. The magnetization is given by the straight line $M(H) = -H$ for $0 < H < H_c$ and then drops to zero. An ideal type II conductor without any flux pinning should also react reversibly. A hard superconductor, on the other hand, is only reversible in the Meissner phase because then no magnetic field enters the bulk, so no flux pinning can happen. If the field is raised beyond $H_{c1}$ magnetic flux enters the sample and is captured at pinning centres. When the field is reduced again these flux lines remain bound and the specimen keeps a frozen-in magnetization even for vanishing external field. One has to invert the field polarity to achieve $M = 0$ but the initial state ($H = 0$ and no captured flux in the bulk material) can only be recovered by warming up the specimen to destroy superconductivity and release all pinned flux quanta, and by cooling down again.

A typical hysteresis curve is shown in Fig. 6. There is a close resemblance with the hysteresis in iron except for the sign: the magnetization in a superconductor is opposed to the magnetizing field because the physical mechanism is diamagnetism. The magnetic hysteresis is associated with energy dissipation. When a hard superconductor is exposed to a time-varying field and undergoes a cycle like the loop in Fig. 6, the energy loss is given by the integral

$$Q_{\text{hyst}} = \oint \mu_0 M(H) dH .$$

(11)

It is equal to the area enclosed by the loop. This energy must be provided by the power supply of the field-generating magnet and is transformed into heat in the superconductor when magnetic flux quanta are moved in and out of the specimen.

2.6 Critical current density

For a hard superconductor, not only temperature $T$ and magnetic field $H$ have to be specified but also current density $J$. The material can be conveniently characterized by its critical surface in a $(T, H, J)$ coordinate system. For the most important conductor used in magnets, niobium-titanium, this surface is

---

$^2$This statement applies only for long cylindrical or elliptical samples oriented parallel to the field.
depicted in Fig. 7. Superconductivity prevails everywhere below the surface and normal conductivity above it. A hard superconductor is not exactly free of any resistance. The critical current density (at a given temperature and field) is usually defined by the criterion that the resistivity be \( \rho = \rho_c = 10^{-14} \Omega \text{m} \). In the vicinity of this point the resistivity is a very steep function of current density. It can be parametrised with a power law

\[
\rho(J) = \rho_c \left( \frac{J}{J_c} \right)^n.
\]

The exponent \( n \) is a quality index which may be as large as 50 for a good multifilamentary NbTi conductor.

Fig. 7: Sketch of the critical surface of NbTi. Also indicated are the regions where pure niobium and pure titanium are superconducting. The critical surface has been truncated in the regime of very low temperatures and fields where only sparse data are available.

2.7 Superconductors in microwave fields

Superconductivity in microwave fields is not treated adequately in standard text books. For this reason I present in this section a simplified explanation of the important concepts. A similar treatment can be found in [10]. Superconductors are free from energy dissipation in direct-current (dc) applications, but this is no longer true for alternating currents (ac) and particularly not in microwave fields. The reason is that the high-frequency magnetic field penetrates a thin surface layer and induces oscillations of the electrons which are not bound in Cooper pairs. The power dissipation caused by the motion of the unpaired electrons can be characterized by a surface resistance. In copper cavities the surface resistance is given by (see sect. 3.2.3)

\[
R_{\text{surf}} = \frac{1}{\delta \sigma}
\]

where \( \delta \) is the skin depth and \( \sigma \) the conductivity of the metal.

The response of a superconductor to an ac field can be understood in the framework of the two-fluid model. An ac current in a superconductor is carried by Cooper pairs (the superfluid component) as well as by unpaired electrons (the normal component). Let us study the response to a periodic electric field. The normal current obeys Ohm’s law

\[
J_n = \sigma_n E_0 \exp(-i\omega t)
\]

\(^3\)A similar model is used to explain the peculiar properties of liquid helium below 2.17 K.
while the Cooper pairs receive an acceleration \( m_c \dot{v}_c = -2e E_0 \exp(-i\omega t) \), so the supercurrent density becomes

\[
J_s = i n_c \frac{2e^2}{m_e \omega} E_0 \exp(-i\omega t) .
\]

If we write for the total current density

\[
J = J_n + J_s = \sigma E_0 \exp(-i\omega t)
\]

we get a complex conductivity:

\[
\sigma = \sigma_n + i \sigma_s \quad \text{with} \quad \sigma_s = \frac{2 n_c e^2}{m_e \omega} = \frac{1}{\mu_0 \lambda_L^2 \omega} .
\]

We know already that the rf magnetic field penetrates a superconductor much less than a normal conductor, namely only to a depth \( \lambda_L \). The surface resistance is the real part of the complex surface impedance

\[
R_{surf} = \Re \left( \frac{1}{\lambda_L (\sigma_n + i \sigma_s)} \right) = \frac{1}{\lambda_L} \cdot \frac{\sigma_n}{\sigma_n^2 + \sigma_s^2} .
\]

Since \( \sigma_n^2 \ll \sigma_s^2 \) at microwave frequencies one can disregard \( \sigma_n^2 \) in the denominator and obtains \( R_{surf} \propto \sigma_n / (\lambda_L \sigma_s^2) \). So we arrive at the surprising result that the microwave surface resistance is proportional to the normal-state conductivity.

The conductivity of a normal metal is given by the classic Drude expression

\[
\sigma_n = n_n e^2 \ell / m_e v_F
\]

where \( n_n \) is the density of the unpaired electrons, \( \ell \) their mean free path and \( v_F \) the Fermi velocity. The normal electrons are created by thermal breakup of Cooper pairs. There is an energy gap \( E_g = 2\Delta(T) \) between the BCS ground state and the free electron states. By analogy with the conductivity of an intrinsic (undoped) semiconductor we get \( n_n \propto \exp(-E_g/(2k_B T)) \) and hence

\[
\sigma_n \propto \ell \exp(-\Delta(T)/(k_B T)) .
\]

Using \( 1/\sigma_s = \mu_0 \lambda_L^2 \omega \) and \( \Delta(T) \approx \Delta(0) = 1.76 k_B T_c \) we finally obtain for the BCS surface resistance

\[
R_{BCS} \propto \lambda_L^3 \omega^2 \ell \exp(-1.76 T_c/T) .
\]

This formula displays two important aspects of microwave superconductivity: the surface resistance depends exponentially on temperature, and it is proportional to the square of the rf frequency.

### 3 Design principles and properties of superconducting cavities

#### 3.1 Choice of superconductor

In principle the critical temperature of the superconductor should be as high as possible. However, copper cavities coated with a high-\( T_c \) superconductor layer have shown unsatisfactory performance [11], therefore the helium-cooled low-\( T_c \) superconductors are applied. In contrast to magnets were hard superconductors with large upper critical field (10–20 T) are needed, the superconductor in microwave applications is not limited by the upper critical field but rather by the thermodynamic critical field (or possibly the ‘superheating field’) which is well below 0.5 T for all known superconducting elements and alloys. Moreover, strong flux pinning appears undesirable as it is coupled with hysteretic losses. Hence a ‘soft’ superconductor must be used, and pure niobium is the best candidate although its critical temperature is only 9.2 K and the thermodynamic critical field about 190 mT. Niobium-tin (Nb₃Sn) may
appear more favorable since it has a higher critical temperature of 18 K and a superheating field of 400 mT; however, the gradients achieved in Nb$_3$Sn coated copper cavities were below 15 MV/m, probably due to grain boundary effects in the Nb$_3$Sn layer [16]. For these reasons pure niobium has been chosen in all large scale installations of sc cavities. There remain two choices for the cavity layout: the cavity is made from copper and the inner surface is coated with a thin layer of Nb or, alternatively, the cavity is made from solid Nb. The former approach has been taken with great success with the 350 MHz cavities of the Large Electron Positron ring LEP at CERN. In the TESLA linear collider, however, gradients of more than 25 MV/m are needed, and these are presently only accessible with cavities made from solid niobium. A high thermal conductivity is needed to guide the heat generated at the inner cavity surface through the wall to the liquid helium coolant. For this reason, the material must be of extreme purity with contaminations in the ppm range.

The BCS surface resistance is given by eq. (21). The exponential temperature dependence is experimentally verified, see Fig. 8. For niobium the surface resistance at 1.3 GHz amounts to about 800 nΩ at 4.2 K and drops to 15 nΩ at 2 K. Operation at 2 K is essential for achieving high accelerating gradients in combination with very high quality factors. Superfluid helium is an excellent coolant owing to its high heat conductivity. In addition to the BCS term there is a residual resistance caused by impurities, frozen-in magnetic flux or lattice distortions.

$$R_{\text{surf}} = R_{\text{BCS}} + R_{\text{res}}.$$  \hspace{1cm} (22)

$R_{\text{res}}$ is temperature independent and amounts to a few nΩ for a clean niobium surface but may readily increase if the surface is contaminated.

Formula (21) applies if the mean free path $\ell$ of the unpaired electrons is much larger than the coherence length $\xi$. In niobium this condition is usually not fulfilled and one has to replace $\lambda_L$ in the above equation by [17]

$$\Lambda = \lambda_L \sqrt{1 + \xi/\ell}. \hspace{1cm} (23)$$

Combining equations (21) and (23) we arrive at the surprising statement that the surface resistance does not assume its minimum value when the superconductor is of very high purity ($\ell \gg \xi$) but rather of...
moderate purity with a mean free path \( \ell \approx \xi \), see Fig. 8. The measured BCS resistance in the sputter-coated LEP cavities is in fact a factor of two lower than in bulk niobium cavities [19]. The sputtered niobium layer has a low \( RRR \) (see sect. 3.6) and an electron mean free path \( \ell \approx \xi \).

### 3.2 Pill box cavity

The simplest model of an accelerating cavity is a hollow cylinder which is often called pill box. When the beam pipes are neglected the field pattern inside the resonator and all relevant cavity parameters can be calculated analytically.

#### 3.2.1 Field pattern

For particle acceleration we need a longitudinal electric field on the axis, hence we look for TM (transverse magnetic) eigenmodes of the cylindrical resonator. The field lines are sketched in Fig. 9. We use cylindrical coordinates \((r, \theta, z)\) where \(z\) denotes the beam direction (cavity axis), \(r = \sqrt{x^2 + y^2}\) the distance from the axis and \(\theta\) the azimuthal angle. We search for an eigenmode with cylindrical symmetry (independence of \(\theta\)) and with longitudinal electric and azimuthal magnetic field. The wave equation for the electric field reads

\[
\frac{\partial^2 E_z}{\partial r^2} + \frac{1}{r} \frac{\partial E_z}{\partial r} = \frac{1}{c^2} \frac{\partial^2 E_z}{\partial t^2} .
\]

For a harmonic time dependence \(E_z(r) \cos(\omega t)\) and with the new variable \(u = r \omega / c\) one obtains

\[
\frac{\partial^2 E_z}{\partial u^2} + \frac{1}{u} \frac{\partial E_z}{\partial u} + E_z(u) = 0 .
\]

This is the Bessel equation of zero order with the solution \(J_0(u)\). Hence the radial dependence of the electric field is

\[
E_z(r) = E_0 J_0 \left( \frac{\omega r}{c} \right) .
\]

For a perfectly conducting cylinder of radius \(R_c\) the longitudinal electric field must vanish at \(r = R_c\), so \(J_0(\omega R_c / c) = 0\). The first zero of \(J_0(u)\) is at \(u = 2.405\). This defines the frequency of the lowest eigenmode (we call it the fundamental mode in the following):

\[
f_0 = \frac{2.405 c}{2 \pi R_c} , \quad \omega_0 = \frac{2.405 c}{R_c} .
\]

In a cylindrical cavity the frequency does not depend on the length \(L_c\). The magnetic field can be computed from the equation

\[
\frac{\partial E_z}{\partial r} = \mu_0 \frac{\partial H_\theta}{\partial t} .
\]

Hence we obtain for the fundamental TM mode

\[
E_z(r, t) = E_0 J_0 \left( \frac{\omega_0 r}{c} \right) \cos(\omega_0 t) ,
\]
Electric and magnetic field are 90° out of phase. The azimuthal magnetic field vanishes on the axis and assumes its maximum close to the cavity wall.

### 3.2.2 Stored energy

The electromagnetic field energy is computed by integrating the energy density \( \left( \varepsilon_0/2 \right) E^2 \) (at time \( t = 0 \)) over the volume of the cavity. This yields

\[
U = \frac{\varepsilon_0}{2} 2\pi L_c E_0^2 \int_0^{R_c} J_0^2 \left( \frac{\omega_0 r}{c} \right) r dr
\]

where \( a = 2.405 \) is the first zero of \( J_0 \). Using the relation \( \int_0^a J_0^2(u) du = 0.5(aJ_1(a))^2 \) we get for the energy stored in the cavity

\[
U = \frac{\varepsilon_0}{2} E_0^2 (J_1(2.405))^2 \pi R_c^2 L_c .
\]

### 3.2.3 Power dissipation in the cavity

We consider first a cavity made from copper. The rf electric field causes basically no losses since its tangential component vanishes at the cavity wall while the azimuthal magnetic field penetrates into the wall with exponential attenuation and induces currents within the skin depth\(^4\). These alternating currents give rise to Ohmic heat generation. The skin depth is given by

\[
\delta = \sqrt{\frac{2}{\mu_0 \omega \sigma}}
\]

where \( \sigma \) is the conductivity of the metal. For copper at room temperature and a frequency of 1 GHz the skin depth is \( \delta = 2\mu \) m. Consider now a small surface element. From Ampere’s law \( \oint \vec{H} \cdot d\vec{s} = I \) follows that the current density in the skin depth is related to the azimuthal magnetic field by \( j = H_\theta/\delta \). Then the dissipated power per unit area is\(^5\)

\[
\frac{dP_{\text{diss}}}{dA} = \frac{1}{2\sigma \delta} H_\theta^2 = \frac{1}{2} R_{\text{surf}} H_\theta^2 .
\]

Here we have introduced a very important quantity for rf cavities, the surface resistance:

\[
R_{\text{surf}} = \frac{1}{\sigma \delta} .
\]

In a superconducting cavity \( R_{\text{surf}} \) is given by equations (21) to (22). The power density has to be integrated over the whole inner surface of the cavity. This is straightforward for the cylindrical mantle where \( H_{\theta} = \frac{E_0}{\mu_0 c} J_1 (\omega_0 R/c) \) is constant. To compute the power dissipation in the circular end plates one has to evaluate the integral \( \int_0^a (J_1(u))^2 u du = a^2 (J_1(a))^2 /2 \). Again \( a = 2.405 \) is the first zero of \( J_0 \). The total dissipated power in the cavity walls is then

\[
P_{\text{diss}} = R_{\text{surf}} \cdot \frac{E_0^2}{2 \mu_0^2 c^2} (J_1(2.405))^2 2\pi R_c L_c \left( 1 + R_c/L_c \right) .
\]

\(^4\)For a thorough discussion of the skin effect see J.D. Jackson, Classical Electrodynamics, chapt. 8.

\(^5\)In equation (33) the quantity \( H_\theta \) denotes the amplitude of the magnetic field without the periodic time factor \( \sin(\omega_0 t) \).
3.2.4 Quality factor

The quality factor is an important parameter of a resonating cavity. It is defined as $2\pi$ times the number of cycles needed to dissipate the stored energy, or, alternatively, as the ratio of resonance frequency $f_0$ to the full width at half height $\Delta f$ of the resonance curve

$$Q_0 = 2\pi \frac{U_{f_0}}{P_{\text{diss}}} = \frac{f_0}{\Delta f}.$$  

(36)

Using the formulas (31) and (35) we get the important equation

$$Q_0 = \frac{G}{R_{\text{surf}}} \quad \text{with} \quad G = \frac{2.405 \mu_0 c}{2(1 + R_c/L_c)}$$  

(37)

which states that the quality factor of a cavity is obtained by dividing the so-called ‘geometry constant’ $G$ by the surface resistance. $G$ depends only on the shape of the cavity and not on the material. A typical value is 300 $\Omega$. We want to point out that the quality factor $Q_0$ defined here is the intrinsic or ‘unloaded’ quality factor of a cavity. If the cavity is connected to an external load resistor by means of a coupler another quality factor ($Q_{\text{ext}}$) has to be introduced to account for the energy extraction through the coupler.

3.2.5 Accelerating field, peak electric and magnetic fields

A relativistic particle needs a time $L_c/c$ to travel through the cavity. During this time the longitudinal electric field changes. The accelerating field is defined as the average field seen by the particle

$$E_{\text{acc}} = \frac{1}{L_c} \int_{-L_c/2}^{L_c/2} E_0 \cos(\omega_0 z/c) \, dz, \quad V_{\text{acc}} = E_{\text{acc}} L_c.$$  

(38)

Choosing a cell length of one half the rf wavelength, $L_c = c/(2f_0)$, we get $E_{\text{acc}} = 0.64 E_0$ for a pill box cavity.

The peak electric field at the cavity wall is $E_0$. The peak magnetic follows from eq. (29). We get

$$E_{\text{peak}}/E_{\text{acc}} = 1.57, \quad B_{\text{peak}}/E_{\text{acc}} = 2.7 \text{ mT/(MV/m)}.$$  

(39)

If one adds beam pipes to the cavity these number increase by 20 - 30%.

3.3 Shunt impedance

To understand how the rf power coming from the klystron is transferred through the cavity to the particle beam it is convenient to represent the cavity by an equivalent parallel LCR circuit. The parallel Ohmic resistor is called the shunt impedance $R_{\text{shunt}}$ although this quantity has only a real part. The relation between the peak voltage in the equivalent circuit and the accelerating field in the cavity is

$$V_0 = V_{\text{acc}} = E_{\text{acc}} L_c.$$ 

The power dissipated in the LCR circuit is

$$P_{\text{diss}} = \frac{V_0^2}{2R_{\text{shunt}}}.$$ 

Identifying this with the dissipated power in the cavity, eq. (35), we get the following expression for the shunt impedance of a pillbox cavity$^6$

$$R_{\text{shunt}} = \frac{2L_c^2 \mu_0 c^2}{\pi^3(J_1(2.405))^2 R_c (R_c + L_c)} \cdot \frac{1}{R_{\text{surf}}}.$$  

(40)

$^6$ $R_{\text{shunt}}$ is often defined by $P_{\text{diss}} = V_0^2 / R_{\text{shunt}}$, then the $(R/Q)$ parameter is a factor of 2 larger.
The surface resistance of a superconducting cavity is extremely small, about \(15 \text{n}\Omega\) at 2 K; consequently, the shunt impedance is extremely large, in the order of \(5 \cdot 10^{12} \Omega\). Note that “on resonance” \((\omega = \omega_0 = 1/\sqrt{LC})\) the parallel LCR circuit behaves like a purely Ohmic resistor whose value is equal to the shunt impedance.

The ratio of shunt impedance to quality factor is an important cavity parameter

\[
(R/Q) \equiv \frac{R_{\text{shunt}}}{Q_0} = \frac{4Lc\mu_0 c}{\pi^3(J_1(2.405))^2} 2.405 R_c
\]

The \((R/Q)\) parameter is independent of the material, it depends only on the shape of the cavity. A typical value for a 1-cell cavity is \((R/Q) = 100 \Omega\).

### 3.4 Shape of practical cavities

The first sc cavities were built in the late 1960’s with the conventional pill-box shape. They showed unexpected performance limitations: at field levels of a few MV/m a phenomenon called multiple impacting (or multipacting for short) was observed. The effect is as follows: stray electrons which are released from the wall (for instance by cosmic rays) gain energy in one half-period of the electromagnetic field and return to their origin in the next half period were they impinge with a few 100 eV onto the wall and release secondary electrons which repeat the same procedure. This way an avalanche of electrons is created which absorbs energy from the rf field, heats the superconductor and eventually leads to a breakdown of superconductivity. It was found out many years later that this problem is avoided in cavities having the shape of a rotational ellipsoid. When electrons are emitted near the iris of an elliptical cavity and accelerated by the rf field, they return to a point away from their origin in the next half period, and the same applies for the possible next generations of electrons. Thereby the daughter electrons move more and more into the equator region where the rf electric field is small and the multiplication process dies out. For a thorough discussion I refer to [10].

In electron-positron storage rings quite often single-cell cavities are used. These are particularly well suited for the large beam currents of up to 1 A in the high luminosity 'B meson factories'. At larger energies like in LEP (104 GeV per beam) multicell cavities are more efficient to compensate for the huge synchrotron radiation losses (3 GeV per revolution in LEP). In a linear collider almost the full length of the machine must be filled with accelerating structures and then long multicell cavities are mandatory. There are, however, several effects which limit the number of cells \(N_c\) per resonator. With increasing \(N_c\) it becomes more and more difficult to tune the resonator for equal field amplitude in every cell. Secondly, in a very long multicell cavity 'trapped modes' may be excited by the short particle bunches. These are coupled oscillations at high frequency which are confined to the inner cells and have such a low amplitude in the beam pipe sections that they cannot be extracted by a higher-order mode coupler. Trapped modes have a negative influence on the following bunches and must be avoided. The number \(N_c = 9\) chosen for TESLA appears a reasonable upper limit. The TESLA cavity [22] is shown in Fig. 10.

Superconducting cavities are always operated in standing-wave mode\(^7\). The fundamental TM\(_{010}\) mode is chosen with a longitudinal electric field on the axis. In a cavity with \(N_c\) cells the fundamental mode splits into \(N_c\) coupled modes. The \(\pi\) mode with \(180^\circ\) phase difference between adjacent cells transfers the highest possible energy to the particles. The cell length \(L_c\) is determined by the condition that the electric field has to be inverted in the time a relativistic particle needs to travel from one cell to the

\(^7\)In normal-conducting linacs like SLAC the travelling wave mode may be chosen. Basically the electrons 'ride' on the crests of the rf wave which propagates with the speed of light. In a superconducting linac a travelling wave is not attenuated by wall losses, and in order to preserve the basic advantage of superconductivity - almost no rf power is wasted - one would have to extract the rf wave after some length and feed it back through a superconducting wave guide to the input coupler. The required precision in rf phase would be extremely demanding and would make such a system far more complicated than a standing-wave linac.
next, so $L_c = c/(2 f_0)$. For nonrelativistic protons or ions the cell length is $L_c = v/(2 f_0)$. The iris radius influences the cell-to-cell coupling parameter $k_{cell}$ which is in the order of 1 - 2%. The frequencies of the coupled modes are given by the formula

$$f_m = \frac{f_0 \sqrt{1 + 2 k_{cell} \cos(m\pi/N_c)}}{1 + 2 k_{cell} \cos(m\pi/N_c)}, \quad 1 \leq m \leq N_c . \quad (42)$$

### 3.5 Choice of frequency

The losses in a microwave cavity are proportional to the product of conductor area and surface resistance. For a given length of a multicell resonator, the area scales with $1/f$ while the surface resistance scales with $f^2$ for $R_{BCS} \gg R_{res}$ (see eq. (21)) and becomes independent of $f$ for $R_{BCS} \ll R_{res}$. At $T = 2$ K the BCS term dominates above 3 GHz and here the losses grow linearly with frequency, whereas below 300 MHz the residual resistance dominates and the losses are proportional to $1/f$. To minimize power dissipation in the cavity wall one should therefore select $f$ in the range 300 MHz to 3 GHz. Cavities in the 350 to 500 MHz regime are commonly used in electron-positron storage rings. Their large size is advantageous to suppress wake field effects and losses from higher order modes. However, for a linac of several 10 km length the niobium and cryostat costs would be prohibitive for these bulky cavities, hence a higher frequency has to be chosen. Considering material costs $f = 3$ GHz might appear the optimum but there are compelling arguments for choosing about half this frequency.

- The wake fields generated by the short electron bunches depend on radius as $1/r^2$ for longitudinal and as $1/r^3$ for transverse wakes. Since the iris radius of a cavity is inversely proportional to its eigenfrequency, the wake field losses scale with the second resp. third power of the frequency. Beam emittance growth and beam-induced cryogenic losses are therefore much higher at 3 GHz.
- The $f^2$ dependence of the BCS resistance makes a 3 GHz cavity thermally unstable at gradients above 30 MV/m, hence choosing this frequency would preclude a possible upgrade of the TESLA collider to 35 MV/m [10].
3.6 Heat conduction in niobium and heat transfer to the liquid helium

![Graph showing heat conductivity and temperature relationship](image)

**Fig. 11:** Measured heat conductivity in niobium samples with \( RRR = 270 \) and \( RRR = 500 \) as a function of temperature [20].

The heat produced at the inner cavity surface has to be guided through the cavity wall to the superfluid helium bath. At 2 - 4 K, impurities have a strong impact on the thermal conductivity of metals. Niobium of very high purity is needed (contamination in the ppm range). The heat conductivity drops by about an order of magnitude when lowering the temperature from 4 to 2 K, as shown in Fig. 11. The residual resistivity ratio \( RRR = R(300\,K)/R(10K) \) is a good measure for the purity of the material: large \( RRR \) means high electrical and thermal conductivity at low temperature.

The beneficial effect of a high thermal conductivity on the cavity performance is demonstrated in Fig. 12. Here the quality factor \( Q_0 \) is plotted as a function of the accelerating field \( E_{acc} \) for a nine-cell TESLA cavity before and after a 1400\(^\circ\)C heat treatment. The \( RRR \) of the niobium increases during this treatment from 380 to 760, and obviously the cavity can be excited to higher fields afterwards. What one can also observe in this figure is a drop of the quality factor towards high fields. The main reason is in this case the onset of electron field emission (see below). These electrons are accelerated in the rf field and impinge on the cavity walls where they deposit energy and may even generate X rays by bremsstrahlung.

![Graph showing excitation curve](image)

**Fig. 12:** The excitation curve \( Q_0 = Q_0(E_{acc}) \) of a nine-cell TESLA cavity before and after the heat treatment at 1400\(^\circ\)C.
Fig. 13: Excitation curves of electropolished cavities at a helium temperature of 2 K. Left: 1-cell cavities, right: 9-cell cavity.

Low frequency cavities (350-500 MHz) have a small BCS surface resistance at 4.2 K and are effectively cooled by normal liquid helium. The heat flux should not exceed a few kW/m² to obtain nucleate boiling with a close contact between liquid and metal. At higher heat fluxes one enters the film boiling regime where a vapour film covers the surface. Here the cavity may easily warm up beyond $T_c$ at areas of excessive heating. The $f^2$ dependence of the BCS resistance implies that for cavities of higher frequency superfluid helium at 1.8 - 2 K is more appropriate. At the metal-helium interface a temperature jump is observed which is attributed to phonon mismatch. The so-called Kapitza resistance amounts to about $1.5 \cdot 10^{-4}$ m²K/W [21] for a clean niobium surface in contact with superfluid helium.

### 3.7 Maximum accelerating field

Superconductivity breaks down when the microwave magnetic field at the cavity surface exceeds a critical value. The situation is clearcut for a type I superconductor such as lead where the superconducting state prevails up to $B_c = 80$ mT. For the type II conductor niobium it is not so obvious which of the critical fields should be used. The values at 2 K are approximately

- $B_{c1} = 160$ mT
- $B_c = 190$ mT
- $B_{c2} = 300$ mT

A very safe value would be $B_{c1}$ because then no magnetic flux enters the bulk niobium. The corresponding gradient of 38 MV/m in TESLA-type cavities has been definitely exceeded in single-cell cavities, more than 40 MV/m have been repeatedly reached. There is some evidence that the maximum tolerable rf magnetic field is close to the thermodynamic critical field of 190 mT but the issue is not finally settled yet.

The most promising road towards the ultimate gradient is a cavity surface preparation by electrolytic polishing. In figure 13 we show the performance of several electropolished cavities. These results prove that a centre-of-mass energy of 800 GeV can indeed be reached in the TESLA electron-positron collider.

### 3.8 Thermal instability and field emission

The fundamental advantage of superconducting cavities is their extremely low surface resistance of about 10 nΩ at 2 K, leading to rf losses which are 5 to 6 orders of magnitude lower than in copper cavities. The drawback is that even tiny surface contaminations are potentially harmful as they decrease the quality
factor and may even lead to a thermal breakdown (quench) of the superconductor due to local overheating.

Temperature mapping at the outer cavity wall usually reveals that the heating is not uniform over the whole surface but that certain spots exhibit larger temperature rises, often beyond the critical temperature of the superconductor. Hence the cavity becomes partially normal-conducting, associated with strongly enhanced power dissipation. Because of the exponential increase of surface resistance with temperature this may result in a run-away effect and eventually a quench of the entire cavity. Analytical models and numerical codes are available to describe this effect. The tolerable defect size depends on the purity of the material. As a typical number, the diameter of a normal-conducting spot must be less than 50 \( \mu m \) to avoid a thermal instability at 25 MV/m.

Field emission of electrons from sharp tips has been a notorious limitation of high-gradient sc cavities. The typical indication is that the quality factor drops exponentially above a certain threshold field, and X rays are observed. The field emission current density is given by the Fowler-Nordheim equation [23], adapted for rf fields:

\[
j_{FE} \propto \frac{E_{loc}^{2.5}}{\Phi} \exp(-C\frac{\Phi^{3/2}}{E_{loc}}) \tag{43}\]

Here \( \Phi \) is the work function of the metal, \( C \) a constant and \( E_{loc} \) the local electric field. At sharp tips on the surface the local field may be several 100 times larger than the accelerating field. Perfect cleaning by rinsing with high-pressure ultrapure water is the most effective remedy against field emission. Using the clean room techniques developed in semiconductor industry it has been possible to raise the threshold for field emission in multicell cavities from about 10 MV/m to more than 20 MV/m in the past few years. Thermal instabilities and field emission are discussed at much greater detail in [10].

References

NON-LINEARITIES IN LIGHT SOURCES

A. Streun
Paul Scherrer Institut, Villigen, Switzerland

Abstract
Correction of a light source’s large chromaticity, while maintaining sufficient dynamic aperture and momentum acceptance, requires careful optimization of the sextupole configuration in first and second order of sextupole strength. In this paper we try to explain chromaticity and its correction in a most intuitive way. Accompanied by the stepwise improvement of a test lattice, the scheme is then expanded from plain correction with two sextupole families to installation of ‘harmonic’ sextupole for first and second order optimization. The approach is very visual and pragmatic at the expense of the mathematical formalism which may be found in the references.

1 INTRODUCTION
Sextupole magnets to correct chromaticity are the dominating nonlinear elements in a light-source lattice, because light sources are built for lowest emittance and thus require very strong horizontal focusing. Strong quadrupoles cause large chromaticities that require strong sextupoles for compensation in order to avoid single-bunch head–tail instability and several multi-bunch instabilities.

The parabolic field variation in a sextupole makes it an essentially nonlinear device which causes chaotic or unbounded particle motion beyond some maximum stable amplitude. The phase-space area enclosing stable particle oscillations is called dynamic acceptance, its projection onto physical space, dynamic aperture.

Light sources require a large horizontal dynamic aperture for injection. In particular, a top-up operation depends on clean and efficient injection. Light sources also require a large momentum acceptance for sufficient beam lifetime: The Touschek scattering contribution to beam lifetime is usually large in light sources because, on account of small beam emittance, the particle density in the bunch is high, thus raising the probability for intra-beam scattering events with large momentum changes. Vertical dynamic aperture is less relevant because the physical apertures will be rather low anyway due to the presence of low-gap insertion devices. Obviously, this also prohibits any scheme for vertical injection. Typical requirements are about $10 \ldots 30$ mm·mrad for horizontal acceptance, about $1 \ldots 5$ mm·mrad for vertical acceptance and about $\pm 3 \ldots 4\%$ for relative momentum acceptance.

Thus light-source design has to face the conflict between a requirement for large dynamic acceptance and the unavoidable presence of strong nonlinear devices attacking this acceptance. It is no exaggeration to consider this problem the most challenging task in light-source lattice design.

Apart from the sextupoles, other nonlinearities have to be taken into account, like higher multipoles in ring magnets due to saturation or geometric imperfections, and in insertion devices due to finite pole width and other imperfections. However, these nonlinearities can be suppressed by proper design of components, and tolerance limits will be provided based on tracking calculations. But the sextupoles, to make this clear, are nonlinear by design and require appropriate treatment. This is the subject of this paper. We proceed in four steps: Section 2 explains how quadrupoles generate chromaticity, Section 3 shows how chromaticity is corrected with two families of sextupoles, and how the dynamic acceptance is destroyed. Section 4 explains how introducing several more sextupole families helps to suppress the adverse effects while maintaining chromaticity corrections and how to perform them. Section 5 considers second-order sextupole effects, also important for low-emittance light sources, and how further dynamic acceptance optimization can be achieved.
2 CHROMATICITY

As sketched here, a quadrupole shows chromatic aberrations, i.e., a variation of focal length with momentum. Since its strength is given by

$$b_2 = \frac{1}{(B \rho)} \frac{d B_y}{d x}$$

with the ‘magnetic rigidity’

$$(B \rho) := \frac{B}{e} = 3.3356 \text{Tm} \cdot E [\text{GeV}]$$

a function of momentum, the focusing strength varies as

$$b_2(\delta) = \frac{b_2}{(1 + \delta)} \approx b_2 (1 - \delta) \quad \delta := \frac{\Delta p}{p} \ll 1.$$ 

---

**Fig. 1:** A double-cell (two double-bend achromats out of 32) of the original ESRF lattice with dispersion-free straight sections as a typical test case for light-source lattices. Uncorrected chromaticity results in strong variation of beta functions with momentum.

As a consequence, the off-momentum optical functions in a strong focusing lattice, as shown in Fig. 1 from the European Synchrotron Radiation Facility (ESRF), vary significantly from the on-momentum functions even for small momentum deviations. Chromaticity is defined as the variation of tune with momentum and derived in the most simple way as a gradient distortion to the one-turn
matrix [1]:
\[
\begin{pmatrix}
1 & 0 \\
\pm b_2 \delta & 1
\end{pmatrix} \cdot \begin{pmatrix}
\frac{\cos 2\pi Q}{\sin 2\pi Q} & \beta \sin 2\pi Q \\
-\frac{\sin 2\pi Q}{\beta} & \cos 2\pi Q
\end{pmatrix} = \begin{pmatrix}
\frac{\cos 2\pi \tilde{Q}}{\sin 2\pi \tilde{Q}} & \beta \sin 2\pi \tilde{Q} \\
-\frac{\sin 2\pi \tilde{Q}}{\beta} & \cos 2\pi \tilde{Q}
\end{pmatrix} := \tilde{M}.
\]

The new tune is given by the half-trace of the new matrix:
\[
\frac{1}{2} \text{Tr}(\tilde{M}) = \cos 2\pi \tilde{Q} = \cos 2\pi (Q + \Delta Q) = \cos 2\pi Q \pm \frac{1}{2} b_2 \delta \sin 2\pi Q \delta s.
\]

Considering small deviations in tune, the cosine is expanded and we obtain the horizontal and vertical chromaticity $\xi_{x/y}$ by integration over all gradient distortions over the ring:
\[
\Delta Q \ll 1 \rightarrow \Delta Q = \mp \frac{1}{4\pi} b_2 \delta \beta \delta s \quad \xi_{x/y} = \frac{\Delta Q}{\delta} = \mp \frac{1}{4\pi} \oint_C b_2(s) \beta_{x/y}(s) ds.
\]

Both chromaticities are naturally negative, where in a light source in particular the horizontal chromaticity will become large in absolute value: values can be of the order of $\xi_x \approx -100$. Negative chromaticity must be avoided for suppression of the head–tail instability, an unstable oscillation between leading and trailing electrons inside a bunch with onset at very low currents [2]. Also coupled-bunch oscillations are driven by negative chromaticity. On the other hand, a large absolute value of (negative or positive) chromaticity would result in a wide tune spread of the beam halo leading to particle losses at low-order resonances and thus low momentum acceptance. For that reason, the chromaticities have to be zero or at moderate positive values.

![Horizontal Tune](image1)

**Fig. 2:** Variation of beta functions and tunes with momentum without chromaticity correction for the lattice shown in Fig. 1. Note how the large horizontal chromaticity leads to a wide tune spread extending over several integers.

Figure 2 shows the large variation of tunes with momentum due to natural chromaticity. Considering that particles get lost at least at integer and half-integer resonances, the momentum acceptance for the example shown would amount to less than 0.5% even for very low currents. Reasonable currents could not be stored anyway due to head–tail instability.
The magnetic field in a sextupole varies as
\[ B_y(x) = \frac{1}{2} B'' x^2 . \]
Thus in a small range around some \( \hat{x} \) it can be considered like a quadrupole with gradient
\[ B'_y(\hat{x}) = B'' \cdot \hat{x} \]
and thus adds or subtracts focusing strength depending on \( \hat{x} \). Dispersion \( D \) as generated by a dipole magnet sorts particles by momentum.

Thus, if
\[ \hat{x} = D \delta \]
and the sextupole strength is well chosen, the chromatic aberration of a quadrupole can be compensated.

The kick on a particle in a quadrupole or sextupole is given by
\[
\text{quadrupole:} \quad \begin{align*}
\Delta x' &= -b_2 L x \\
\Delta y' &= b_2 L y
\end{align*}
\]
\[
\text{sextupole:} \quad \begin{align*}
\Delta x' &= -b_3 L (x^2 - y^2) \\
\Delta y' &= b_3 L (2 xy)
\end{align*}
\]
with \( b_2 = \frac{1}{(B \rho) \frac{d^2 B_y}{dx^2}} \); with \( b_3 = \frac{1}{2} \frac{1}{(B \rho) \frac{d^2 B_y}{dx^2}} \).

Chromatic aberrations modify the magnet strength as
\[ b_n(\delta) = b_n/(1 + \delta) \approx b_n(1 - \delta) . \]

Adding dispersion to the beam is a transformation
\[ x \to D \delta + x \quad y \to y . \]
Inserting these transformed coordinates into the kick equations and keeping up to second orders in products of \( x, y \) and \( \delta \) gives
\[
\text{quadrupole:} \quad \begin{align*}
\Delta x' &= -b_2 L x + [b_2 L] \delta x \\
\Delta y' &= b_2 L y - [b_2 L] \delta y \\
\end{align*}
\]
\[
\text{sextupole:} \quad \begin{align*}
\Delta x' &= -(2 b_3 LD) \delta x - b_3 L (x^2 - y^2) - b_3 L D^2 \delta^2 , \\
\Delta y' &= +2 b_3 L D \delta y + 2 b_3 L xy .
\end{align*}
\]
Obviously for \( 2 b_3 LD = \frac{1}{2} b_2 L \) the sextupole corrects the quadrupole’s chromaticity. However, there are other terms coming with the sextupoles which cause the problems with dynamic apertures.

This simple example has to be generalized to quadrupolar and sextupolar fields distributed around the ring. Chromaticities are then given as integral expressions:
\[
\xi_{x/y} = \pm \frac{1}{4 \pi} \oint_C \left[ 2 b_3(s) D(s) - b_2(s) \right] \beta_{x/y}(s) \, ds . \tag{1}
\]
Considering discrete, thin magnets, i.e., neglecting any change of optical functions over the magnet length, this equation can be expressed as a sum:

$$\xi_{x/y} = \frac{1}{4\pi} \left( \pm \sum_{\text{sext}} 2(b_3L_n) \beta_{(x/y)n} D_n \mp \sum_{\text{quad}} (b_2L_n) \beta_{(x/y)n} \right).$$  \qquad (2)

Defining two families of sextupoles $\text{SF, SD}$ for correction of horizontal and vertical chromaticity, this equation can be written as a $2 \times 2$ linear system for the 2-vector of sextupole strengths:

$$\frac{1}{2\pi} \left( + \sum_{n \in \text{SF}} \beta_{xn} D_n + \sum_{n \in \text{SD}} \beta_{yn} D_n \right) \left( \begin{array}{c} (b_3L)_{\text{SF}} \\ (b_3L)_{\text{SD}} \end{array} \right)_{1 \times 2} = \frac{1}{4\pi} \left( + \sum_{\text{Quad}} (b_2L) \beta_{xn} \right)_{1 \times 2} \left( \begin{array}{c} \Delta \xi_x \\ \Delta \xi_y \end{array} \right) \implies \langle b_3L\rangle_{\text{SF}}, \langle b_3L\rangle_{\text{SD}}. \quad (3)$$

Here $[\Delta \xi_{x/y}]$ indicates an optional non-zero value of chromaticity to be obtained from the sextupoles. From the linear system, a first recommendation as to how to place the sextupoles becomes visible: All sextupoles should be at locations of large dispersion. In order to decouple the two families, i.e., to avoid that they act against each other, or, mathematically speaking, that the linear system degenerates, the SF family members should be located at locations of large $\beta_x$ and low $\beta_y$, and the SD family members at locations of large $\beta_y$ and low $\beta_x$. Figure 1 shows the placement of the sextupoles. Finding a good place for SF is usually easy, since $\beta_x = x^2/\epsilon$ and dispersion $D = x/\delta$ are both subject to horizontal focusing and thus behave in a similar fashion. In Fig. 1 the central DBA quadrupole required for reflecting the dispersion function was split into two quadrupoles and the SF sextupole inserted between. As SD sextupole locations with optimum conditions are hard to find, in Fig. 1 two SDs are inserted at medium dispersion and $\beta_y$ slightly larger than $\beta_x$.

![Figure 3: Variation of beta functions and tunes with momentum after chromaticity correction with two families of sextupoles: Linear chromaticities have disappeared as is visible by the horizontal tangent to the tune vs. $dp/p$ curves, but second-order chromaticity remains.](image)

Figure 3 shows the result of successful chromaticity correction. However, tracking particles in phase space reveals how the dynamic aperture breaks down as shown in the Poincaré plots of Fig. 4. The horizontal dynamic aperture is too small to inject a beam into the machine, and the vertical to horizontal coupling is significant.
Fig. 4: Poincaré plots of particle motion in horizontal (red) and vertical (green) phase space before (left) and after (right) plain chromaticity correction with two families of sextupoles: Without sextupoles the motion is linear and the dynamic acceptances are unlimited. With sextupoles, nonlinear effects lead to a breakdown of dynamic acceptances. The bright blue circles indicate the physical aperture from an unrealistic quadratic beampipe of 160 mm width. (Real beampipes are typically 60...90 mm wide and 20...40 mm high.)

4 FIRST-ORDER Sextupole OPTIMIZATION

The lattice with plain chromaticity emerging from the previous section would not be able to operate because the dynamic aperture is too small to inject a beam, and the beam lifetime would be very short. But chromaticity correction is mandatory in order to store significant current in the machine. So, the method of chromaticity correction has to be improved in order to suppress the adverse sextupole effects.

In the following we outline a ‘standard procedure’ of sextupole optimization which is derived and described in detail in Refs. [3–5].

In order to understand what sextupoles actually do to the beam and how they destroy it, and in order to find a cure subsequently, the single-particle Hamiltonian has to be studied. In a lattice made from dipoles, quadrupoles, and sextupoles it is given by

\[ H(s) = \frac{p_x^2 + p_y^2}{2(1 + \delta)} - \text{Dipoles} + \text{Quadrupoles} + \text{Sextupoles} \]

The goal is to find a sextupole distribution such that \( H_2 + H_3 \) becomes achromatic while staying linear. For that purpose, the linear betatron oscillation for a flat (i.e., \( D = D_x, D_y = 0 \)) lattice, given by

\[ x(s) = \sqrt{2J_x \beta_x(s) \cos \phi(s)} + D(s) \delta \]

\[ y(s) = \sqrt{2J_x \beta_x(s) \cos \phi(s)} \]

is introduced into Eq. (4), the powers of the trigonometric functions are turned into linear functions of multiple arguments, and the different modes, i.e., terms with the same arguments are collected. As a result, the quadrupole/sextupole Hamiltonian can be represented by a sum over different frequencies:

\[ \int_{\text{cell}} [H_2(s) + H_3(s)] \, ds = \sum h_{jk\text{imp}}, \text{ with} \]

\[ h_{jk\text{imp}} \propto \sum N_{\text{ext}}(b_3 L)_n \beta_{x_{kn}} \beta_{y_{kn}} D_n^p e^{i((j-k)\phi_{xn} + (l-m)\phi_{yn})} - \sum N_{\text{quad}}(b_2 L)_n \beta_{x_{kn}} \beta_{y_{kn}} e^{i((j-k)\phi_{xn} + (l-m)\phi_{yn})} \]

\[ p \neq 0 \]
This complex expression becomes more intuitive when abbreviating

\[ h = \sum_n V_n e^{i\phi_n} [ + \ldots \text{quads for } p \neq 0 \ldots] . \]

The Hamiltonian modes are sums of complex vectors, where each vector corresponds to a sextupole, its length given essentially by the sextupole’s integrated strength (multiplied with the optical functions at its location which is the ‘lever arm’ to act on the beam), and its complex phase by the [modal multiple] of the betatron phases at its location.

All together, nine Hamiltonian modes are found: Two of them have zero betatron phases, they are given by

\[ h_{11001} = +J_x \delta \left( \sum_n (2b_3 L_n) n \beta_{xn} D_n - \sum_n (b_2 L_n) n \beta_{yn} \right) \]
\[ h_{00111} = -J_y \delta \left( \sum_n (2b_3 L_n) n \beta_{yn} D_n - \sum_n (b_2 L_n) n \beta_{yn} \right) . \]

Comparison to Eq. (2) reveals, that these are just the chromaticities.

Seven modes (and their complex conjugates) have phase arguments, thus they do not add up over many turns but they will show a resonant structure: Long-term behaviour is revealed by considering \( N \) repetitions of the lattice structure, i.e., many cells and many turns, and extrapolating to infinity:

\[
|h_{jklmp}^\infty| = \left| \sum_{n=0}^\infty h_{jklmp} e^{i(\pi n Q_{\text{cell}}x + (l-m) \pi n Q_{\text{cell}}y)} \right| = \frac{|h_{jklmp}|}{2 \sin \pi ((j-k) Q_{\text{cell}}x + (l-m) Q_{\text{cell}}y)} . \quad (7)
\]

In particular we find modes driving the following resonances:

- \( h_{21000} = h_{12000} \rightarrow Q_x \)
- \( h_{30000} = h_{03000} \rightarrow 3 Q_x \)
- \( h_{10110} = h_{01110} \rightarrow Q_x \)
- \( h_{10200} = h_{01020} \rightarrow Q_x + 2 Q_y \)
- \( h_{10020} = h_{01020} \rightarrow Q_x - 2 Q_y \)
- \( h_{20001} = h_{02001} \rightarrow 2 Q_x \)
- \( h_{00201} = h_{00021} \rightarrow 2 Q_y \).

These resonances correspond to forbidden lines in the tune diagram, i.e., working points close to these lines must be avoided, otherwise particles will be resonantly extracted from the beam and get lost. The last two are chromatic half-integer resonance drive terms and cause chromatic variation of the beta functions.

As long as the periodicity of the machine is large and non-linearities are not too strong, there may be one or a few single resonances dominating the dynamics. In this case, the calculation may be
continued further by harmonic expansion of the Hamiltonian modes in order to obtain single-resonance drive terms [6]. The traditional approach to suppress a single or a small group of terms (while exciting others) by appropriate distribution of sextupoles has proven successful in the past for several machines. However, a very advanced light source may have rather low periodicity, thus the tune space is densely covered with sextupolar and other resonances, and many resonances will contribute to the dynamics. In this more general case, it is better to suppress the $h_{jklmp}$ modes right from the beginning because they are the source of all the resonances.

This is done by adding more sextupoles, even in dispersion-free regions, which do not contribute to chromaticity correction, but are solely installed for minimization of the resonance driving modes. Four families of these—traditionally called ‘harmonic’ sextupoles—are visible in the dispersion-free matching sections in Fig. 1. Since each sextupole corresponds to a complex vector, diagrams like those shown in Fig. 5 can help to suppress the resonance driving modes while keeping the chromaticities compensated.

Even after this simple manual optimization using four additional sextupole families, the gain in dynamic acceptance is impressive when comparing the result of phase space tracking with Fig. 4 (right), which was obtained without harmonic sextupoles. Note that the harmonic sextupoles are no small correctors but of comparable strength to the chromaticity sextupoles.

Another improvement is visible in the frequency spectra of a test particle before and after installing harmonic sextupoles as shown in Fig. 6: amplitudes of peaks related to sextupolar resonances are significantly reduced in height.

This procedure can be set up in a more systematic fashion. There are nine modes to be adjusted, two of them real, seven complex, which gives a total of 16 quantities to be minimized.
However, most machines have symmetry points usually used as reference for tracking etc. Seen from such a symmetry point, every element has a mirror image of opposite betatron phase. Thus the complex parts cancel, and we are left with only nine equations, which, in principle, could be solved by means of nine sextupole families.

These equations [see Eqs. (6)] are linear in sextupole strength. If there are \( M \) sextupole families, we get a \( 9 \times M \) linear system for the vector of sextupole strengths:

\[
\left\{ \sum_{n \in \{S_m\}} \beta_n D_n \left( \ldots e^{i\{\ldots \phi_n \}} \ldots \right) \right\} \cdot \left\{ (b_1 L)_m \right\} = \left\{ \sum_{\text{Quad}} (b_2 L) \ldots \right\} . \tag{8}
\]

This is just a generalization of the \( 2 \times 2 \) system for plain chromaticity correction from Eq. (3). For \( M = 9 \) the system is quadratic and can be solved exactly, for \( M \neq 9 \) the method of Singular Value Decomposition (SVD) [7] is able to obtain a solution too, which for \( M < 9 \) returns a least-square minimized approximation to the right-hand side vector, while for \( M > 9 \) selects the solution associated with minimum sextupole strength. Singular value decomposition also provides insight into the good behaviour of the linear system by returning a vector of weighting factors. Degeneration of the system is indicated by the appearance of very small weighting factors.

In fact, and particularly for light sources, the system tends to degenerate; this has been investigated in detail in Ref. [4]. The horizontal betatron phase-advance per cell in a typical light source lattice is \( \Delta \phi_{x,\text{cell}} \approx 140 \ldots 160^\circ \) in order to obtain low emittance (see Section 4.4 and Fig. 7 in Ref. [8]). This is quite close to \( 180^\circ \), thus \( e^{i2\phi_x} \approx 1 \). As a consequence the \( 2Q_x \) resonances driving term \( h_{20001} \) becomes proportional to the chromaticity \( \xi_x \propto h_{11001} \). This means that no sextupole pattern exists to suppress this term and \( \xi_x \).

Resonances of \( 2Q_x \) cause a momentum-dependent beta-beat \( \partial \beta_x / \partial \delta \), which is the source of second-order chromaticity \( \partial^2 Q_x / \partial \delta^2 \). In fact, light source lattices are often limited in momentum acceptance due to large second-order chromaticity driving off-momentum particles to resonances. Even after installation of harmonic sextupoles, second-order chromaticities remain large, as visible by the parabolic shape of the tune versus momentum curves in Fig. 7.
Linear chromaticities = 0

Fig. 7: Dynamic acceptances after first-order sextupole optimization. The dotted yellow line indicates the $3Q_x = 106$ resonance.

In order to avoid or cure this degeneration problem, either particular betatron phase relations between sextupoles have to be fixed in order to exploit periodicity and symmetry conditions, or a distributed dispersion optics has to be chosen in order to make all sextupoles chromatic. In any case, these measures have a strong impact on linear lattice design and layout:

**Periodicity:** If there are $N$ cells, the ideal tune advances per cell would simultaneously make

$$N\Delta Q_x^{\text{cell}}, 3N\Delta Q_x^{\text{cell}}, 2N\Delta Q_x^{\text{cell}}, 2N\Delta Q_y^{\text{cell}}$$

integers.

For example, this works with $N = 5$, $\Delta Q_x^{\text{cell}} = 0.4 (= 144^\circ)$, $\Delta Q_y^{\text{cell}} = 0.1 (= 36^\circ)$, where $144^\circ$ phase advance per cell is a reasonable value for a light source lattice:

**Symmetry:** A lattice section versus its mirror image may suppress $2Q_x, 2Q_y$ if the tune advances per section are close to $\Delta Q_x^{\text{cell}} = \frac{2n_x+1}{4}, \Delta Q_y^{\text{cell}} = \frac{2n_y+1}{4}$ (integers). Two sections versus their mirror also suppress $Q_x, 3Q_x$. Of course this requires fixed phase advances over the straight sections which restricts the lattice flexibility. For example, the scheme applied to the Swiss Light Source [5] is sketched here ($n_x \approx 7, n_y \approx 3$):

**Dispersive straights:** If dispersion in the straight sections is accepted, also the harmonic sextupoles become chromatic sextupoles. Thus several families add up in chromaticity correction but may be tuned to cancel their contribution to the $2Q_x$ terms, in other words the linear system from Eq. (8) is not degenerate. This is an advantage of distributed dispersion lattices like SOLEIL [9], which in fact achieved very good suppression of second-order chromaticity.
Single-resonance suppression: If, in the end, a single resonance is responsible for limitation of momentum acceptance, a harmonic decomposition of the source mode may be performed for suppression of this particular resonance. This was successfully accomplished for the $3Q_x = 106$ resonance crossed at $\delta \approx \pm 2\%$ in the original ESRF lattice (see figure above) [10].

5 SECOND-ORDER SEXTUPOLE OPTIMIZATION

Second-order sextupole effects are due to cross-talk between different sextupoles. First-order effects perturb the linear betatron motion from Eq. (5). Reintroducing the perturbed motion into the sextupole Hamiltonian will lead to effects of second order in sextupole strength — and to even higher orders. Basically, sextupoles can drive resonances of any order, but the drive terms become weaker with increasing order.

In this context we should mention that in thick sextupoles, in other words, where phases and betas vary over the length of the sextupole talk to themselves, there is cross-talk between slices of the same magnet. But here we consider only thin sextupoles.

Derivation of the (quite unwieldy) formulae is done either by second-order perturbation theory [3] or by application of Lie algebra techniques [4]. Basically, the second-order modes are products of the first-order modes from Eq. (6) of type $h_{ijklmp} \cdot h_{ij'k'l'm'}$.

It turns out that there are five phase-independent terms, two of which are the second-order chromaticites that have already appeared in the previous section, and three are the amplitude-dependent tune shifts:

$$
\xi^{(2)}_x = \frac{\partial^2 Q_x}{\partial \delta^2}, \quad \xi^{(2)}_y = \frac{\partial^2 Q_y}{\partial \delta^2}, \quad \frac{\partial Q_x}{\partial J_x} \frac{\partial Q_y}{\partial J_y} = \frac{\partial Q_y}{\partial J_x} \frac{\partial Q_y}{\partial J_y}.
$$

Second-order chromaticities limit momentum acceptance. The analytic formulae based on chromatic beta functions and higher-order dispersion are rather lengthy, and codes for minimizing them run more efficiently by obtaining the values from numerical differentiation of the dispersive closed orbit [4].

Amplitude-dependent tune shifts lead to a twist in phase space and subsequent resonance overlap, and with it chaotic motion and fractal dynamic acceptance structure [11]. However, a closed expression suitable for minimization is available (Eq. (119) in Ref. [4] or Eq. (195) in Ref. [3]).

There are also eight phase-dependent terms. Calculating the long-term behaviour analogous to Eq. (7) reveals resonance denominators of the type

$$
|h^\infty_{ijklmpj'k'l'm'}| \propto \frac{\ldots}{8 \sin \pi [\bar{m} \cdot \bar{Q}] \sin \pi [\bar{m}' \cdot \bar{Q}] \sin \pi [(\bar{m} + \bar{m}') \cdot \bar{Q}]} \bar{m} := \left(\frac{j - k}{l - m}\right), \quad \bar{Q} := \left(Q_x^{\text{cell}}, Q_y^{\text{cell}}\right).
$$

The third term in the denominator drives octupolar resonances, where the corresponding mode of the octupole Hamiltonian (i.e., the first order Hamiltonian $H_4$ of octupole magnets) $h_{jkln0}$ is identified by
\[ \hat{j} = (j - k) + (j' - k') \]
\[ \hat{k} = (j - k) - (j' - k') \]
\[ \hat{l} = (l - m) + (l' - m') \]
\[ \hat{m} = (l - m) - (l' - m') . \]

The following resonances are found:
\[
\begin{align*}
h_{40000} &\rightarrow 4Q_x & h_{31000} &\rightarrow 2Q_x \\
h_{00400} &\rightarrow 4Q_y & h_{20110} &\rightarrow 2Q_x \\
h_{20200} &\rightarrow 2Q_x + 2Q_y & h_{00310} &\rightarrow 2Q_y \\
h_{20020} &\rightarrow 2Q_x - 2Q_y & h_{01110} &\rightarrow 2Q_y .
\end{align*}
\]

Practically, optimizing the dynamic aperture in a light-source lattice requires a code for minimization of first- and second-order sextupole terms, and many iterations with well-chosen weighting factors. Usually, the first-order modes and the amplitude-dependent tune shifts require large weights, whereas the octupolar resonances seem less important. The second-order chromaticities turn out to be quite resistive against optimization, because their sources are inappropriate phase advances between lattice sections. Of course, the chromaticity is kept constant during minimization by executing Eq. (3) after each step.

In fact, second-order optimization helps to improve the first-order result as proven by increased dynamic aperture in Fig. 8 (left), and further reduced tune-walk with momentum: compare Fig. 9 with Fig. 7.

However, basically it appears unnatural and inefficient to optimize the sextupoles while putting most weight on their second-order effects. Instead octupoles could be applied, which attack in first order the second-order sextupole terms. This type of scheme basically would set sextupoles based on Eq. (8) for chromaticity correction and first-order cancellation, and transfer the 13 (assuming symmetry, otherwise 21) second-order terms as right-hand-side vectors to a $13 \times P$ linear system for $P$ octupole families. Such a scheme might be required for more challenging light-source storage rings in the future. After all, blindly playing with octupoles provides promising results as shown in Figs. 8 (right) and 9 (bottom).

6 SUMMARY

A modern source lattice design must not be preceded by the linear lattice design for optimum emittance, straight length etc., but has to take the sextupole scheme into account from day one in order to find suitable places (large dispersion, decoupled betas) for the chromaticity sextupoles and to exploit symmetry and periodicity for cancellation of at least some of the sextupolar resonance drive terms.

Restoring dynamic acceptances to values sufficient for injection and for providing reasonable beam lifetime requires installation of several families of ‘harmonic’ sextupoles which mainly compensate adverse effects from the chromaticity sextupoles. Degeneration of the configuration has to be tested and avoided.

First- and second-order sextupole terms will be minimized by means of a suitable code through many trials with different weighting of the diverse quantities to be minimized. Solutions for the sextupole pattern obtained this way are only based on perturbation theory, thus validity for the required amplitudes has to be tested by tracking dynamic apertures. This should also include misalignments, magnet errors, and other effects disturbing periodicity and symmetry in order to ensure robustness of the solution.
Fig. 8: Dynamic acceptance improvement by second-order optimization of four harmonic sextupole families (left), and further improvement after adding small octupoles (right).

Fig. 9: Top figure shows tunes vs. momentum after second-order sextupole optimization. Second-order chromaticities could be slightly reduced compared to Fig. 7, but remain stubborn. Finally, they are suppressed by additional octupoles in dispersive regions, as shown in the bottom figure: Small positive linear chromaticity was also used to ‘fold’ the horizontal tune vs. momentum curve in the most narrow interval (note the different scale).

References


Lattices for light sources

Andreas Streun
Paul Scherrer Institut, CH-5232 Villigen PSI

Abstract
This paper is intended to provide a guideline on how to design the lattice (i.e., the magnet arrangement) of a high brightness light source.

In the introductory section we will consider the general framework to the design, clarify the steps of the design process and the interfaces to technical engineering, the lattice designer has to take into account.

In the second section we will examine the lattice building blocks, i.e., the magnets. A short detour to magnet design will help to understand the limitations of magnet strength.

An introduction to beam dynamics is not the subject of this paper, instead all relevant formula to understand subsequent sections are summarized in section three without any derivation.

Section four focusses on emittance and how to build a lattice for obtaining low emittance. This is first explained in a most intuitive way, then the theoretical minimum emittance and deviations from the minimum conditions are calculated in some details.

The final section dealing with the problem of acceptance considers the requirements for beam life time and injection efficiency. Physical acceptances will be discussed in some details, whereas the large subject of dynamic acceptance optimization is subject of another lecture [25] and only briefly summarized here.

1 Introduction

1.1 Global requirements

Synchrotron light users demands specify the layout of a new light source:

- The highest photon energy to be obtained from the most advanced insertion devices available (in terms of period length and harmonic content) sets the electron energy of the storage ring.
- The target performance in terms of brightness defines the natural, horizontal emittance of the electron beam, and – considering diffraction limited experiments in the X-ray range – also the emittance coupling, i.e., the ratio of vertical to horizontal emittance.
- The number and types of planned beamlines defines the shape of the machine, i.e., the number and length of straight sections and the symmetry of the lattice.
- Experiments expect micron photon beam stability at the location of the probe to be analyzed. Thermal stability of the beamline is maintained by constant photon heat load, i.e., by constant beam current. This requires a long beam life time and/or operation with top-up injection (i.e., very frequent refill).
- Low radiation background to the experiments also calls for long beam lifetime and for efficient injection into the machine, in particular in top-up operation. Both constraints require large acceptances of the machine in order to keep particles deviating in longitudinal and transverse momenta after scattering with residual gas atoms or each other, and in order to safely capture injected beams.
– The available area for building the machine and the beamlines limits the circumference of the lattice. This makes it more challenging to accommodate the requested number of straights while achieving the desired performance, and defines the type of magnetic arc to be used.
– Every light source needs some operational flexibility and some upgrade potential for new and yet unforeseeable future experiments.
– Finally, the limitations of budgets for construction and operation call for most simple and efficient solutions (in terms of number of magnets and magnet families, power consumption, etc.).

1.2 Lattice design phases and tools

The lattice design process may be divided into four phases:

1. **Preparation**: Definition of performance issues and boundary conditions. Acquisition of information on available building blocks (magnets) for composing the lattice concerning their properties and technical limitations.

2. **Linear lattice layout**: Arrangement of linear building blocks (quadrupoles and bending magnets) to obtain the desired global (i.e., concerning the lattice as a whole) quantities like circumference, emittance, etc. This phase deals with the concepts of periodic cells, matching sections, insertions etc.

3. **Nonlinearities**: Introduction of sextupoles and RF cavities for stabilization of particles with momentum deviations. Due to the nonlinearity of these elements dynamic acceptance, i.e., stability limits for transverse and longitudinal deviations from the reference orbit becomes the main design issue.

4. **Errors**: Investigation of lattice performance in presence of magnet misalignments, multipolar errors, vibrations etc. and development of correction schemes. This final design phase ends with a significant prediction on the performance of the machine, resp. tolerance requirements for the components.

Little of lattice design can be done analytically, most tasks require the aid from a computer code. During the linear layout, it is most important for the designer to have a visual, dialog-oriented code in order to ‘play’ with lattices and optimize them interactively, whereas in the later design stages the exact modeling including nonlinearities and errors is crucial in order to obtain predictive results. Typically, work proceeds by alternating fast, creative steps using reduced models, and slow, consolidating runs based on more complete models. Usually not all tasks can be fulfilled by one code and switching between two or more is required. – Here we list some keywords for lattice development code requirements:

– **Model**: complete set of elements, correct methods for tracking and concatenation, well documented approximations.
– **Elementary functions**: beta functions and dispersions, periodic solutions, closed orbit finder, energy variations, tracking, matching.
– **Toolbox**: Fourier transforms of particle data (see resonance analysis), minimization routines (see dynamic aperture optimization, coupling suppression), linear algebra package (see orbit correction).
– **User convenience**: editor functions, graphical user interface, editable text files.
– **Extended functionality**: RF dimensioning, geometry plots, lifetime calculations, injection design, alignment errors, multipolar errors, ground vibrations.
– **Connectivity**: database access, control system access (see real machine operation).

218
1.3 Interfacing

Lattice design provides an empty machine, telling how single particles move around the ring and what equilibrium will form for an ensemble of particles. However, prediction of performance in terms of brightness for light sources also includes the maximum beam current to be stored, which is a complex subject to be shared between beam dynamics, vacuum and RF departments. In the framework of this paper it will be assumed as given and not considered further. However, the lattice designer has to stay in contact with his/her colleagues from other departments for several reasons:

- **Vacuum**: Impedance of vacuum chamber affects maximum beam current, pressure affects lifetime, pumps, absorbers and flanges require space.
- **Radiofrequency**: RF parameters determine momentum acceptance and other parameters affecting directly the lattice design.
- **Diagnostics**: Beam position monitors have to be inserted at the appropriate locations (betatron phases) and also require space.
- **Magnet Design**: Technological limits and geometrical properties of magnets determine the maximum magnet strength to be used in lattice design, magnet coils need space, multipolar errors affect the acceptance.
- **Alignment**: Misalignments cause orbit distortions affecting the performance and requiring correction schemes.
- **Mechanical engineering**: Grouping of magnets on stiff girders improves the robustness of the lattice to misalignments and vibrations.
- **Construction**: All devices to be installed by the different departments meet on the design engineers blueprint and conflicts will reveal there.

Most important for the lattice designer is to include the space requirements for all devices right from the beginning. Figure 1 illustrates this by comparing the lattice designer’s and the design engineer’s view of a lattice section.

1.4 Conventions and approximations

A curvilinear coordinate system is commonly used for describing particle motion in a storage ring, where the axis \(x\) points radially to the ring outside, the \(s\)-axis is the tangent to the ring, pointing forward, and the \(y\)-axis points up. Some authors define it as a right handed system \(\{x; y; s\}\) [2], other authors as a right-handed system.
handed system \( \{x; s; y\} \) [4, 20] (which also might be viewed as a left handed \( \{x; y; s\} \) system!). The second convention has the advantages, that particles in the storage ring rotate mathematically positive (i.e., counter clockwise seen from above), and that for constant radius of curvature \( \rho \) the system maps to a cylindrical coordinate system \( \{r; \theta; z\} \) with \( r = \rho + x, \quad \theta = s/\rho, \quad z = y \) [23].

Positive particles require a negative magnetic field \( \vec{B} = -|B_y| \hat{y} \) to get the appropriate Lorentz force to form the storage ring. Since the radius of curvature of a curve in space is always positive by definition, the product of magnetic field and radius of curvature, also called magnetic rigidity \( (B\rho) \) (see Eq. (2) below) has to be negative.

For the practical design of a low emittance light source usually an idealized subset of the theoretical beam dynamics formalism is sufficient and obtained by introducing the following approximations:

- Highly relativistic beam, i.e., \( v = c, \ E = pc \)
- Small deviations from the reference axis \( (x, y \ll \rho, \ x', y' \ll 1) \) for validity of linear beam dynamics using betatron amplitudes, emittance, betafunctions, betatron phases, etc.
- Decoupling of subspaces: Synchrotron motion, i.e., dynamics in \( (\delta, \Delta s) \) subspace is slow and thus treated as a constant parameters over the timescales of betatron motion (‘adiabatic approximation’). Coupling between horizontal \( (x, x') \) and vertical \( (y, y') \) subspaces is considered to be small.
- Nonlinearities are treated as perturbations.

2 Building blocks

Elementary building blocks for linear lattice design are bending magnets and quadrupoles for guiding and focusing the beam. The nonlinear lattice design also includes sextupoles used for correction of the quadrupoles’ chromatic aberrations. The real lattice with alignment and other errors also includes small corrector dipoles and skew quadrupoles as well as beam position monitors. In addition every ring needs injection devices (kickers and septa) and one or more RF cavities for acceleration and longitudinal focusing. Light sources also contain wigglers and undulators for production of highest flux, resp. brightness synchrotron radiation.

2.1 Lattice composition: local \( \leftrightarrow \) global

It is very important to make a clear distinction between local properties of the building blocks and global properties of the lattice as pronounced by Forest and Hirata [10]:

‘A quantity is called local if it is derivable from the individual magnet irrespective of the magnet position in the ring and even irrespective of the ring itself. For example a trajectory of a particle through the magnet is local. […] Global information, on the contrary, is derivable only after the full ring is produced. For example the dynamic aperture has no meaning whatsoever if we cannot iterate the one-turn map (i.e., circulate particles in the machine).’

Concatenation of building blocks is done by coordinate transformations, i.e., translations and rotations. For example a vertical bending magnet can be described by a horizontal bending preceded by a 90° rotation around the longitudinal axis.

A building block may have any coordinate system, however in practise it is either a cartesian geometry with parallel entrance and exit planes (\( \{x, y\} \) planes, perpendicular to \( s \)) and length \( L \) or a cylindric geometry with an angle \( \phi \) between the entrance and axis planes and are length \( L \). Obviously cylindric geometry is more convenient for the description of bending magnets and cartesian geometry for other magnet types and drift spaces.
After assembling all building blocks and closing the ring the one turn map can be calculated: It is the mapping of the particle vector \( \vec{X} = (x, x', y, y', \delta, \Delta s) \) from one turn to the next \( \vec{X}_n \xrightarrow{\text{map}} \vec{X}_{n+1} \) with the order of the map corresponding to the highest power in the coordinates. Thus the closed orbit is a fixpoint of the one-turn map and the transfermatrix is a linearization of the one-turn map around the closed orbit. The so called ‘design orbit’ is just a coincidence of the closed orbit with (most) magnets’ symmetry axis for the ideal lattice but is not defined a priori.

Eventually any lattice design has to be tested by tracking particles through the lattice, since tracking concatenates all local transformation of the particle vector from block entrance to block exit including the coordinate transformations between blocks, and thus implicitly applies the full one-turn map.

### 2.2 Magnet multipole definition

In the local coordinate system of a magnet the field is given as multipole expansion around the local reference axis \((x=y=0)\) by

\[
B_y(x, s, y) + iB_x(x, s, y) = (B\rho) \sum_n (ia_n(s) + b_n(s))(x + iy)^{n-1}
\]

with \(n\) the multipole order and \(2n\) the number of poles in the magnet, i.e., \(n = 1, 2, 3, \ldots\) indicating dipole, quadrupole, sextupole, etc. The \(b_n\) are the regular multipoles \((B_x = 0 \text{ for } x = 0)\) and \(a_n\) the skew multipoles, obtained through a rotation around the \(s\)-axis by \(90\gamma/n\).

The quantity \((B\rho)\) is called the magnetic rigidity. From the Lorentz force equation it is directly derived as ratio of momentum over charge:

\[
(B\rho) = -\frac{p}{q} = -\frac{\beta E/e}{n_e c} \approx 3.356 E [\text{GeV}] \text{ for relativistic electrons } (n_e = -1)
\]

with \(n_e\) the number of elementary charges per particle and \(\beta = v/c\). By differentiation of Eq. (1) we obtain a useful expression for a pure, regular multipole:

\[
b_n = \frac{1}{B\rho (n-1)!} \frac{\partial^{(n-1)} B_y(x, y)}{\partial x^{n-1}} \bigg|_{y=0}.
\]

The radius of a cylinder around the symmetry axis touching the magnet poles is called the pole inscribed radius or magnet aperture radius \(R\). In case of dipoles the full gap \(g = 2R\) is used for characterization. The poletip field of a regular magnet is then given from Eqs. (1) and (3) by

\[
B_{pt} = (B\rho) b_n R^{n-1} = \frac{R^{n-1}}{(n-1)!} \frac{\partial^{(n-1)} B_y(x, y)}{\partial x^{n-1}} \bigg|_{y=0}.
\]

Following the conventions explained in Section 1.4 above, a positive dipole moment \(b_1\) bends both positive and negative charged particles to the ring inside, because the polarity is contained in \((B\rho)\), or other speaking, \(b_1\) is always positive otherwise we have no ring with a coordinate system \{\(x; s; y\}\) oriented as described above. Consequently, a positive quadrupole moment \(b_2\) focusses all particles horizontally.

Unfortunately there are different definitions of multipole strengths: The quadrupole strength, mostly called \(k\), usually is defined as \(k = -b_2\), however \(k = +b_2\) may also be found. For the sextupole (and higher multipole) strength, called \(m\) or \(k_s\) or other, definitions with and without the factorial are used: \(m = \mp b_3\) or \(m = \mp 2b_3\).

### 2.3 The general bending magnet

The bending magnet is a block of cylindrical symmetry with a reference radius of curvature \(\rho_{ref}\), an arc length \(L\) and a bending angle \(\Phi = L/\rho_{ref}\). The dipole moment \(b_1 = B_y/(B\rho)\) provides a radius \(\rho = 1/b_1\).
of a particle’s trajectory’s curvature. The magnetic field is adjusted that \( \rho = \rho_{\text{ref}} \) for the particular energy of the reference particle. For particles at other energies the trajectories’ curvatures do not match the coordinate system’s curvature, they thus leave the bend off-axis even if they entered on-axis, an effect called dispersion. It is important not to mix \( \rho_{\text{ref}} \) which is given by geometry and \( \rho \) which is a function of magnet current and particle energy.

The general magnet may also contain a gradient (quadrupole moment) \( b_2 \) and is called combined function magnet, because it provides both bending and focusing. The transfer matrix for propagating a vector of \textit{local} coordinates from the entry plane to the exit plane is given by

\[
\begin{pmatrix}
    x' \\
    x \\
    y' \\
    y \\
    \delta'
\end{pmatrix}
= \begin{pmatrix}
    c_x & 1/b \sqrt{K} s_x & 0 & 0 & b_2 (1 - c_x) \\
    -\sqrt{K} s_x & c_x & 0 & 0 & 0 \\
    0 & 0 & c_y & -1/b \sqrt{K} s_y & 0 \\
    0 & 0 & 1/b \sqrt{K} s_y & c_y & 0 \\
    0 & 0 & 0 & 0 & 1
\end{pmatrix} \begin{pmatrix}
    x' \\
    x \\
    y' \\
    y \\
    \delta
\end{pmatrix}
\]  

(5)

with the abbreviations

\[ c_x[s_x] = \cos[\sin((\sqrt{K} L)], c_y[s_y] = \cos[\sin((\sqrt{-b_2} L)] \text{ and } K = b_2^2 + b_2, \delta = \Delta p/p_0. \]

For \( 0 > b_2 > -b_2^2 \) the gradient bend provides horizontal and vertical focusing, for lower values of \( b_2 \) it becomes horizontally defocusing, for positive \( b_2 \) vertically defocusing. (Note: \( \cos ix = \cosh x, \sin ix = i \sinh x \)) Focusing means negative matrix elements \( m_{21} \) and \( m_{43}, \text{ i.e., a } \) positive value of \( x, \text{ resp. } y \) provides a negative increment to \( x', \text{ resp. } y' \)

For \( b_2 = 0 \), i.e., no gradient, it is a pure sector dipole magnet. For \( b_1 = 0 \), it is a quadrupole (for \( b_1 \to 0, \rho \to \infty \) the cylindrical symmetry becomes cartesian) and will not produce any dispersion. If \( b_1 = 0 \) and \( b_2 = 0 \) it is just a drift space.

Weak focusing synchrotrons used combined function magnets, the definition of the ‘field index’

\[
n = -\frac{\rho}{\frac{\partial B_y}{\partial x}} = -\frac{b_2}{b_1^2}
\]  

(6)

goes back to these times. Later, combined function magnets with strong gradients (\(|b_2| \gg b_1^2\)) became again attractive for low emittance lattices [16].

In a pure sector bend the entrance and exit edges of the magnet are orthogonal to the arc, in the general case the edges may be rotated by angles \( \zeta_1, \zeta_2 \). Rectangular bends have parallel entrance and exit edges, \( \zeta_1 = \zeta_2 = \Phi/2 \). Laminated magnets like used in synchrotrons for reasons of eddy current suppression are always rectangular since manufacturing is done by stacking the laminates.

2.4 Magnet design

A brief detour into magnet design is required in order to include limitations on magnet strengths and requirements for distances between magnets into the lattice design process:

**Length** The effective length \( L_{\text{eff}} \) of the field of an iron dominated magnet of yoke length \( L_{\text{iron}} \) is approximately given by

\[
L_{\text{eff}} := \frac{\int B(s) \, ds}{B_o} \approx L_{\text{iron}} + \frac{2R}{n}
\]

where \( B_o \) is the maximum field in the magnet’s center, \( R \) the pole inscribed radius and \( n \) the multipole order. \( L_{\text{eff}} \) and \( B_o \) are the relevant quantities for beam optics. For the total length \( L_{\text{total}} \) the coil size has to be added to the iron length. As an example consider the dipole magnet shown in Fig. 2: The required coil cross section area is given by \( A = B_g/(2j_c \mu_o) \). A conservative value for the gross average
current density in a water cooled coil (including water channels, insulations, etc.) is $j_c \approx 2 \ldots 3 \text{ A/mm}^2$. Assuming a realistic bending magnet of 1.5 T field and 40 mm gap the cross section is $A \approx 100 \text{ cm}^2$ and the coil might be quadratic with 10 cm width and height. The iron length will be shorter than the effective length by $L_{\text{iron}} \approx (L_{\text{eff}} - \text{gap})$, but the coil in our example would add 16 cm to the effective magnet length as used in lattice design. Same considerations can be done for quadrupoles and sextupoles. Figure 1 shows the spaces required by the coils. Some data for the effective coil width, i.e., $(L_{\text{total}} - L_{\text{eff}})/2$, thus to be added in advance on both sides of a magnet as free space are listed in Table 1.

**Maximum poletip field** Some data for the maximum poletip field for normalconducting iron dominated magnets are also given in Table 1. The poletip field is limited due to saturation effects somewhere in the iron. Although magnet iron saturates fully around 2.2 T it becomes nonlinear at lower fields already. In order to maintain the high field homogeneity required for modern machines and to ensure predictability and reproducability the magnet design should try to avoid saturation. The limits on poletip fields for quadrupoles and sextupoles are lower then for bending magnets since flux lines are compressed due to pole geometry and higher field values appear somewhere else in the yoke. Saturation of a $2n$-pole will create a parasitic $6n$-pole.

**Magnet apertures** Calculations as done in Fig. 2 can be done for any iron dominated multipole and lead to the result that the required current per coil expressed as (windings $\times$ current) $ NI$ or (current density $\times$ coil cross section) $j_c A$ is proportional to (poletip field $\times$ aperture radius) $B_{\text{pt}} R$. Keeping the multipole strength constant we thus obtain from Eq. (4) a proportionality of $ NI \propto R^n$, e.g., quadrupole currents increase with the square of aperture. The proportionality is the same for the power if we assume constant current density $j_c$. Thus from the magnet design point of view the apertures should be as small as possible.

On the other hand the apertures have to be as large as necessary to allow efficient injection and sufficient beam lifetime, and also sufficient cross sections for pumping. Since poletip field is the limiting quantity, larger aperture decreases the maximum acceptable multipole strength. With larger aperture a

![Fig. 2: Iron dominated dipole magnet. Integrating Maxwell’s equation $\oint H \, ds = \int j \, da$ along the path as shown at the left gives the cross section area $A$ of the coils required to create the field $B$: $A = B/(2j_c)(S_{\text{iron}}/(\mu_r \mu_0) + g/\mu_0)$ with $j_c$ the current density in the coil. Since the iron permeability $\mu_r \gg 1$ this simplifies to $A \approx Bg/(2j_c \mu_0)$. The figure at right shows the distinction between effective magnet length, iron length and total length due to addition of coil width to iron length.]

Some data for the effective coil width maximum poletip fields and aperture inscribed radii obtained from a survey on existing light source magnets.

<table>
<thead>
<tr>
<th></th>
<th>eff. coil width</th>
<th>max. poletip field</th>
<th>aperture radius</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bending magnets:</td>
<td>6.5 . . . 15 cm</td>
<td>1.5 T</td>
<td>20 . . . 35 mm (=gap/2)</td>
</tr>
<tr>
<td>Quadrupoles:</td>
<td>4 . . . 7 cm</td>
<td>0.75 T</td>
<td>30 . . . 43 mm</td>
</tr>
<tr>
<td>Sextupoles:</td>
<td>4 . . . 8 cm</td>
<td>0.6 T</td>
<td>30 . . . 50 mm</td>
</tr>
</tbody>
</table>
quadrupole has to be built longer for maintaining the integrated focusing strength, when the maximum pole tip field has been reached. In this way the whole machine size increases with magnet apertures. Some data for pole inscribed radii are also listed in Table 1.

In particular, a light source in its final configuration will run with several undulators of rather narrow vertical gap (few mm). Consequently the storage ring could be built using bending magnets of correspondingly small aperture in order to reduce operating costs. On the other hand, narrow vacuum chamber give rise to the so-called resistive wall instability. Thus the best choice of vertical aperture has to be carefully considered in the design of a light source.

3 Lattice properties

For an introduction to linear transverse dynamics we refer to refs. [20, 22, 23, 29] covering the linear Hamiltonian of betatron motion with appropriate approximations (small curvature, paraxial motion, piecewise constant fields), the equations of motion, Hill’s equation and introduction of betafunctions, betatron phases and dispersion, and how to obtain these quantities from a transfer matrix. These references also give examples of transfer matrices for particular magnets and of basic lattice cells like a triplet, a FODO cell, etc., obtained by multiplication of the concatenated magnets’ transfer matrices.

For our practical approach to lattice design only the essential formulae needed for the understanding of the following sections are summarized here:

3.1 Betafunction and emittance

For a linear lattice with negligible transverse coupling and ‘slow’ longitudinal dynamics, the betafunction $\beta(s)$ is introduced for the following motivations:

- Seen from a practical point of view, the r.m.s. beam size (assuming Gaussian particle distributions as they result from the quantum structure of synchrotron radiation), can be expressed as

$$\sigma_x = \sqrt{\frac{\text{Betafunction} \beta(s)}{\text{magnet structure}}} \times \sqrt{\frac{\text{Emittance} \epsilon}{\text{particle ensemble}}}.$$ 

The betafunction is completely defined by the magnetic fields no matter if there are particles in the machine or not, only the reference energy has to be specified. The emittance however describes the invariant phase space volume of the particles no matter which focussing forces act on them. (Of course, in long-term (thousands of turns) the emittance is also determined by the magnet structure, see Section 4)

- Seen from a theoretical point of view, the single particle Hamiltonian of the betatron oscillation is canonically transformed from coordinates $x, p_x$ (with $p_x \approx x'p_o$) to action-angle variables, namely the invariant amplitude $2J$ and the betatron phase $\phi(s)$:

$$H = \frac{p_x^2}{2} + \frac{b_2(s)x^2}{2} \quad \rightarrow \quad \tilde{H} = \frac{J}{\beta(s)}, \quad \epsilon = \langle J \rangle.$$

Here the beta function describes the variation of the Hamiltonian along the lattice, and the emittance turns out to be the average particle amplitude [20,29], (see Section 4.1).

3.2 The betatron oscillation

The linear betatron motion of a particle is described by

$$x(s) = \sqrt{2J_x \cdot \beta_x(s)} \cos \phi_x(s) + D(s) \cdot \delta$$  (7)
with $D = D_x$ the dispersion function describing the orbit translation for relative momentum deviations
$\delta = \Delta p/p_0$. The same equation applies to the vertical $y$, with $D_y(s) \equiv 0$ in case of a flat lattice (i.e., no vertical bending magnets).

$\alpha, \beta, \gamma$ are called the Twiss parameters — not to be confused with the relativistic parameters! — and related with each other and the betatron phase by

$$\phi(s) = \int \frac{1}{\beta(s)} ds \quad \alpha(s) = -\frac{1}{2} \frac{d\beta(s)}{ds} \quad \gamma(s) = \frac{1}{\beta(s)} \left(1 + \alpha(s)^2\right).$$

The angle $x'(s)$ is obtained by differentiation of Eq. (7) using the relations of Eq. (8):

$$x'(s) = \sqrt{\frac{2J_x}{\beta_x(s)}} \left(\sin \phi_x(s) + \alpha_x(s) \cos \phi_x(s)\right) + D'(s) \cdot \delta.$$  

### 3.3 Circle transformation

The Poincaré plot $\{x, x'\}$ at some location $s$ paints an ellipse in phase space during subsequent turns [perhaps with a dispersive offset of the origin]. This ellipse can be converted into a circle of radius $\sqrt{2J}$ by the geometric transformation (also see Fig. 3)

$$\begin{pmatrix} x \\ x' \end{pmatrix} = T \begin{pmatrix} x \\ x' \end{pmatrix} \quad \text{with} \quad T = \begin{pmatrix} \sqrt{\frac{\alpha_x}{\beta_x}} & 0 \\ 1 & \sqrt{\frac{\beta_x}{\alpha_x}} \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} x \\ x' \end{pmatrix} = \sqrt{2J_x} \begin{pmatrix} \cos \phi_x(s) & \sin \phi_x(s) \\ \sin \phi_x(s) & \cos \phi_x(s) \end{pmatrix} \left[1 + \frac{1}{\sqrt{\beta_x}} \begin{pmatrix} D \\ \alpha_x D + \beta_x D' \end{pmatrix} \right] \cdot \delta.$$  

### 3.4 Transfer matrix

The general transfer matrix $M_{a \rightarrow b}$ from some location $a$ to another location $b$ in the lattice is conveniently described by a circle transformation at location $a$, followed by a rotation in phase space by the betatron phase advance $\Delta \phi = \phi_b - \phi_a$ and by a backtransformation at location $b$:

$$M_{a \rightarrow b} = T_b^{-1} \begin{pmatrix} \cos \Delta \phi & \sin \Delta \phi \\ -\sin \Delta \phi & \cos \Delta \phi \end{pmatrix} T_a.$$  

Multiplication gives

$$M_{a \rightarrow b} = \begin{pmatrix} \frac{\alpha_a}{\beta_a} \cos \Delta \phi + \alpha_a \sin \Delta \phi \\ \alpha_a \cos \Delta \phi \left(1 + \frac{\alpha_a}{\beta_b} \sin \Delta \phi\right) \sin \Delta \phi \frac{\sqrt{\beta_a}}{\sqrt{\beta_b}} \sin \Delta \phi \end{pmatrix} \begin{pmatrix} \frac{\beta_a}{\beta_b} \sin \Delta \phi \frac{\sqrt{\beta_a}}{\sqrt{\beta_b}} \sin \Delta \phi \\ \frac{\beta_a}{\beta_b} \sin \Delta \phi \left(\cos \Delta \phi - \alpha_b \sin \Delta \phi\right) \end{pmatrix}.$$  

For a periodic structure, i.e., $b = a$, this simplifies to the one-turn matrix

$$M_a = \begin{pmatrix} \cos 2\pi Q + \alpha_a \sin 2\pi Q & \beta_a \sin 2\pi Q \\ -\gamma_a \sin 2\pi Q & \cos 2\pi Q - \alpha_a \sin 2\pi Q \end{pmatrix}$$

with $Q$ the machine tune, i.e., the number of betatron oscillations per revolution. This matrix further simplifies when considering a symmetry point of the machine where $\alpha = 0$.

It is common use to obtain the global transfer matrix numerically by multiplication of the local element matrices, however this method is dangerous, since it works only for the special case, that the elements are lined up in perfect alignment, so that the beam travels along the symmetry axis of every element. This case is usually given in the linear lattice design phase and therefore it works although it is not the correct procedure. In the general case, as mentioned above (see Section 2.1), the transfermatrix is a linearization around the closed orbit which can be anywhere after concatenation of the elements.
3.5 Twiss parameter propagation

The transformation of twiss parameters along the lattice is given by a $3 \times 3$ matrix composed of elements of $M_{a \rightarrow b}$, here expressed in term of sine and cosine type solutions:

$$
\begin{pmatrix}
\beta \\
\alpha \\
\gamma
\end{pmatrix}_b =
\begin{pmatrix}
C^2 & -2SC & S^2 \\
-CC' & S'C + SC' & -SS'
\end{pmatrix}
\begin{pmatrix}
\beta \\
\alpha \\
\gamma
\end{pmatrix}_a
$$

with $M_{a \rightarrow b} = \begin{pmatrix} C & S \\ C' & S' \end{pmatrix}$.

(15)

3.6 Example: Drift

As the most simple example, we may consider a drift space, the $[(x, x')$ sub-]matrix given by

$$
\begin{pmatrix}
1 & s \\
0 & 1
\end{pmatrix}
$$

(see Eq. (5) for $b_1 = 0, b_2 = 0$). From a focus, where $\beta = \beta_o, \alpha = 0$ and thus $\gamma = 1/\beta$ (see Eq. (8)), the betafunction propagates (see Eq. (15)) as

$$
\beta(s) = \beta_o + \frac{s^2}{\beta_o}.
$$

(16)

The phase advance of a drift space extending from $-L$ to $+L$ with a focus in the centre, is given by integration of Eq. (8) (left):

$$
\phi = 2 \arctan \frac{L}{\beta_o} \xrightarrow{\beta_o \rightarrow 0} 180^\circ = \frac{1}{2} 2\pi.
$$

(17)

The maximum betatron phase advance of a sharp beam focus thus is limited to $180^\circ$, or to 0.5 in tune.

4 Low emittance lattice development

For light sources low emittance is the design criterion, because the brightness of the synchrotron radiation scales quadratically with the inverse emittance in the X-ray region and at least linearly in the [diffraction limited] VUV-region [14].

4.1 Emittance definition and conventions

In an electron storage ring, the two competing synchrotron radiation effects of quantum noise excitation and classical radiation damping lead to a 6-dimensional Gaussian particle distribution with standard deviations $\sigma_x, \sigma_y$ called the [r.m.s.] beam radii, $\sigma_x', \sigma_y'$ the beam divergences, $\sigma_\delta$ the relative energy spread.
and $\sigma_s$, the bunch length. Betatron oscillations are rotations in the transverse 2-dimensional $\{x, x'\}$ and $\{y, y'\}$ sub phase spaces, which are usually considered as decoupled from each other and adiabatically decoupled from the slow synchrotron oscillation in the longitudinal $\{\delta, \Delta s\}$ sub phase space.

Figure 3 (right) shows particles in horizontal phase space with a strong and variable correlation between $x$ and $x'$ depending on the local twiss parameters. By means of the circle transformation Eq. (10) normalized coordinates $\chi, \chi'$ are introduced as shown in Fig. 3 (left): The corresponding Gaussian distributions have standard deviations which are just given by $\sigma_\chi = \sigma_{\chi'} = \sqrt{c}$. Obviously the phase space area containing particles up to 1 standard deviation is given by $A = \pi c$. This remains true for the real coordinates, since the transformation conserves phase space area ($|T| = 1$). Actually, the emittance is invariant to the betatron motion and projects locally varying beam radii and divergences and correlations of both. – In normalized coordinates, the distribution function for the particles is simply

$$\varrho (\chi, \chi') = \frac{1}{2\pi c} e^{-(\chi^2 + \chi'^2)/(2c)} .$$

Going back to real coordinates by the inverse transformation $T^{-1}$ (see Eq. (10)), a correlation term appears

$$\varrho (x, x') = \frac{1}{2\pi c} e^{-(\gamma x^2 + 2\alpha xx' + \beta x'^2)/(2c)} .$$

The observable, 1-dimensional spatial distribution is obtained by

$$\varrho (x) = \int_{-\infty}^{+\infty} \varrho (x, x') dx' = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-x^2/(2\sigma_x^2)} \text{ with } \sigma_x = \sqrt{c} \beta .$$

Transforming into action-angle variables $J$ and $\phi$ through Eq. (11) we get

$$\varrho (J, \phi) = \frac{1}{2\pi c} e^{-J/\epsilon} .$$

It follows, that emittance is the average betatron amplitude of a Gaussian distributed particle ensemble:

$$\langle J \rangle = \int_0^{2\pi} \int_0^{\infty} J \varrho (J, \phi) d\phi dJ = \epsilon .$$

There exist different conventions for emittance: Electron storage rings quote the 1-sigma-emittance as introduced here. The corresponding area of phase space (grey circle/ellipse in Fig. 3) contains only 39.3% of the particles. An interval of $[-\sigma, +\sigma]$ in one dimension however confines 68.3% of the particles since it implies integration of the other coordinates over $]-\infty, \infty[.$

Note: For proton machines, a 2-sigma-emittance is quoted, which is four times larger and confines 86.5% of the particles. This fraction is rather independent of the type of distribution, thus making the 2-sigma-emittance a more robust quantity for protons, which are not always Gaussian distributed.

There are also different opinions on whether the phase space area $A$ as shown in Fig. 3 is given by $F = \pi c$ or by $F = c$. In the latter case, the factor $\pi$ is included in the emittance unit, measuring it in $\pi \cdot \text{m-rad}$, whereas we here measure emittance in units of $\text{m-rad}$, or, since emittance is very small in light sources, in $\text{nm-rad}$.

Finally, normalized emittance $\bar{\epsilon}$ refers to true phase space in canonical coordinates $\{x; p_x\}$. For sub-relativistic beams with large spread of momentum, the difference may be substantial, but for a light source, the simple conversion $\bar{\epsilon} = m_o c^2 \gamma \epsilon$ is usually justified.

For a deeper understanding of emittance, why it is an invariant, the connection to Liouville’s theorem, its statistical interpretation and the formation of the synchrotron radiation equilibrium see Refs. [3, 7, 21].
4.2 Equilibrium emittance

No matter what has been injected into a storage ring, within a few milliseconds the synchrotron radiation effects will reach an equilibrium and shape the particle ensemble to Gaussian distributions. The natural horizontal emittance being the most relevant of these equilibrium values for a light source, is solely determined by the lattice structure, and for a flat lattice (i.e., only horizontal bending magnets) in practical units given by [21]

\[ \varepsilon_{x_0}[\text{nm-rad}] = 1470 \left( E[\text{GeV}] \right)^2 \frac{\langle H/\rho^3 \rangle}{J_x \langle 1/\rho^2 \rangle} \] (18)

with \( \langle \ldots \rangle \) an average over the lattice, and \( H \) the so called ‘lattice invariant’ or ‘dispersion’s emittance’ (actually it is the betatron amplitude of the dispersion see Eq. (11)!

\[ H(s) = \gamma_x(s)D(s)^2 + 2\alpha_x(s)D(s)D'(s) + \beta_x(s)D'(s)^2. \] (19)

The horizontal damping partition number \( J_x \) will be discussed in Section 4.5 below, in most cases it is close to unity, sometimes however intentionally pushed up to values of \( \approx 2 \) in order to halve the emittance.

In case of an isomagnetic lattice, i.e., all magnets having same bending radius, Eq. (18) simplifies to

\[ \varepsilon_{x_0}[\text{nm-rad}] = 1470 \left( E[\text{GeV}] \right)^2 \frac{\langle H \rangle_{\text{mag}}}{\rho J_x} \] (20)

with \( \langle \ldots \rangle_{\text{mag}} \) an average taken over the magnets only.

At first glance, Eq. (19) tells, that a rather sharp horizontal focus of the beam in each bending magnet’s center is the path to low emittance: With \( \alpha_x = D' = 0 \) defining a focus and a low value of dispersion at the focus, \( H \) will be small everywhere inside the bending magnets. Before examining \( H \) systematically in Section 4.4 to learn how to obtain the minimum emittance, the way to build a low emittance light source lattice will be explained from a more intuitive point of view:

4.3 Building a lattice

To work on periodic structures one may first consider a cell providing a bending angle of \( 360^\circ \) and imagine that it is repeated \( N \) times to give a ring. A solution for the betafunction exists, if there is appropriate horizontal and vertical focusing. Fig. 4 displays how to construct a light source lattice cells step by step:

a. Weak focusing: The most simple cell one could think of is one combined function magnet from Eq. (5). However the allowed range of gradients for bounded motion in both transverse planes is restricted to rather small, negative numbers of \( b_2 \). Therefore the focussing is weak and betas-functions and dispersions will be large resulting in large emittance of the beam. However this type of lattice is interesting for simple and compact industrial light sources which require flux, not brightness, for irradiation of materials [12].

b. Strong focusing: The next simple cell consists of two quadrupoles of opposite polarity: they do not cancel each other but provide focussing in both planes (this is easy to show by multiplication of the matrices). This is the principle of alternating gradient (AG) or strong focussing, such named in contrary to the weak focussing of single combined function magnets: The quadrupole gradients are orders of magnitude larger than the gradients of the weak focussing combined function magnet from the previous example, thus they provide strong variation of betasfunctions and sharp focci.

c. The FODO cell: However, a series of quadrupoles gives no ring. So the next simple cell would be alternating quadrupoles with dipole magnets between. The influence of the dipoles on the beam is small compared to the focussing forces from the quadrupoles, wich thus mainly determine the solution for the beta functions. This is the classical FODO cell: [horizontally] focussing
quadrupole ($F$), dipole (no gradient, negligible focusing, 0), defocusing quadrupole ($D$), dipole (0). FODO cells do not allow low emittance but can be built rather dense, thus they are mainly used for high energy physics machines and booster synchrotrons.

d. **Separated function low emittance cell:** The emittance in the FODO lattice is limited, because the beam has its horizontal focus not in the bending magnets but in the D-quadrupoles. So another type of cell is required which could be called FDODF: With the bending magnet in the symmetry point the horizontal focus is moved to its centre and the emittance can become very low, if the F-quadrupoles are strong enough. Several cells of this type forming a ring would already be a low emittance light source.
e. **Combined function low emittance cell:** As an alternative to the previous example, the vertical focussing may be provided by the bending magnet, which thus becomes a combined function magnet again. This cell is more compact and requires less elements, also, as will be explained in Section 4.5 below, the horizontal damping partition number $J_x$ appearing in Eq. (18) may be increased by appropriate choice of the gradient.

f. **Low emittance FODO cell:** Driving further the previous example, also the strong horizontal focussing may be provided by a combined function magnet. With clever distribution of bending angle, length and gradients on the two types of magnets it is possible to achieve low emittance and appropriate damping partitioning with a very simple lattice [16]. However, the disadvantage is the lack of flexibility since the magnet gradients are usually realized by the pole profiles and thus not adjustable.

g. **Dispersion matching:** Highest brightness synchrotron light is not obtained from bending magnets but from undulators. These devices usually have negligible focussing, but they are rather long and require suitable empty straight sections for installation, preferably without dispersion in order to hide any energy fluctuation from the users and avoid increase of the source size due to the beam’s energy spread. Thus a transition has to be constructed from the periodically oscillating dispersion to zero dispersion: Going back to the periodic separated function cell of Fig. 4.d, we see, that in the bending magnet’s center the dispersion $D$ is close to zero and $D' = 0$ by symmetry. Thus the periodic cell may be split in the center, leaving only a half bending magnet. Suppressing the dispersion completely to $D = 0$ while keeping $D' = 0$ can either be done by adjusting the strengths of the F and D quadrupoles, or, more elegantly and smoothly by increasing the distance between the last quadrupole and the half bending magnet, or by a combination of both means.

h. **Matching cell:** The dispersion suppressing cell is no longer periodic. It is a so called matching cell where some final parameters are to be obtained from some initial parameters. In order to insert this matching cell in a lattice without disturbing the periodic solution, a symmetry point ($D' = 0$, $\alpha_x = 0, \alpha_y = 0$) has to be created at its exit. Further, the undulator has to be accomodated in the straight section. The two additional constraints $\alpha_x = \alpha_y = 0$ require two more degrees of freedom, i.e., two more quadrupoles. If also a particular value of betafunction (for maximum acceptance, see Section 5.4) or phase advance (for reasons of dynamic aperture optimization, see Section 5.6) is required, three or four quadrupoles are required. Figure 4.h shows a triplet solution for matching the beam to the straight section including the undulator. Appending the mirror image of this complete matching cell, named $M$ will then form a structure $[M|M]$ which matches the periodic solution from the cell (called $[P]$) of Fig. 4.d at the transition points (entry and exit). Thus it can be inserted anywhere in a series of periodic cells: $[\ldots P|P|M|M - M|P|P\ldots]$ With periodic cells $[P]$ and matching cells $[M]$ different types of lattices can be composed:

i. **Double bend achromat:** The structure $[-M|M]$ is called a double bend achromat (DBA), because it contains two bending magnets, where the first builds up dispersion and the second one suppresses it again, making the whole structure achromatic. Most light sources are of DBA type, for example ESRF, ELETTRA, SOLEIL and others.

j. **Triple bend achromat:** The structure $[-M|P|M]$ is called a triple bend achromat (TBA). It is more compact than a DBA lattice of same emittance, however provides fewer straight sections. The TBA can be modified further by making the half bending magnets at the ends somewhat longer on expense of the center bending magnet in order to obtain the best emittance. Light sources with TBA structure are ALS, PLS and SLS. $[-M|P|M]$ is a quadruple bend achromat (QBA) [9], and so on. Any number of $[P]$ cells could be inserted between $[-M]$ and $|M|$, however this option is rather used for damping rings than for light sources, because the number of straight sections becomes too small.
Fig. 5: ESRF and SLS lattices as examples for DBA and TBA structures.
Table 2: Some high brightness light sources in operation. \( N_{\text{mag}} \) is the number of bending magnets in the lattice, \( \epsilon_{xo} \) [nm rad] is the natural horizontal emittance, and \( F \) is the ratio of emittance to its theoretical minimum, see Eq. (21))

<table>
<thead>
<tr>
<th>Name</th>
<th>Country</th>
<th>( E ) [GeV]</th>
<th>( N_{\text{mag}} )</th>
<th>( \epsilon_{xo} ) [nm rad]</th>
<th>( F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALS</td>
<td>USA</td>
<td>1.5</td>
<td>36</td>
<td>3.4</td>
<td>8.9</td>
</tr>
<tr>
<td>MAX-2</td>
<td>Sweden</td>
<td>1.5</td>
<td>20</td>
<td>8.7 (+D)</td>
<td>3.9</td>
</tr>
<tr>
<td>BESSY-2</td>
<td>Germany</td>
<td>1.7</td>
<td>32</td>
<td>5.2</td>
<td>7.5</td>
</tr>
<tr>
<td>ELETTRA</td>
<td>Italy</td>
<td>2.0</td>
<td>24</td>
<td>7.0</td>
<td>3.1</td>
</tr>
<tr>
<td>PLS</td>
<td>S.Korea</td>
<td>2.0</td>
<td>36</td>
<td>12</td>
<td>18</td>
</tr>
<tr>
<td>SLS</td>
<td>Switzer.</td>
<td>2.4</td>
<td>36</td>
<td>5.0</td>
<td>5.1</td>
</tr>
<tr>
<td>ESRF</td>
<td>Europe</td>
<td>6.0</td>
<td>64</td>
<td>4.0 (+D)</td>
<td>3.7</td>
</tr>
<tr>
<td>APS</td>
<td>USA</td>
<td>7.0</td>
<td>80</td>
<td>8.2</td>
<td>11</td>
</tr>
<tr>
<td>Spring-8</td>
<td>Japan</td>
<td>8.0</td>
<td>96</td>
<td>5.6</td>
<td>9.6</td>
</tr>
</tbody>
</table>

(\( +D \)) indicates dispersive beams in undulators

Figure 5 shows the European Synchrotron Radiation Facility ESRF and the Swiss Light Source SLS as examples for DBA and TBA lattices:

The ESRF lattice in its original, dispersion free mode shown in the figure, can be described as \( 16 \times (-M_H|M_L| - M_L|M_H) \) with two types of matching cells for high and low beta functions in the straight sections. With 5.625° bending angle for each of the 64 magnets, it provides an emittance of 8.17 nm-rad at 6 GeV beam energy and has a circumference of 846 m.

The structure \( 3 \times (-M_L|P|M_S| - M_S|P|M_L| - M_L|P|M_S| - M_S|P|M_L) \) describes the lattice of SLS: Three types of matching cells are used for matching to long, medium and short straight sections. As a modification to the basic TBA of Fig. 4.j the 'half' bends of the matching cells had been increased on the expense of the center full bend during optimization for lowest emittance. With 8° resp. 14° for the 24 end, resp. 12 center bends, the lattice provides an emittance of 5.03 nm-rad at 2.4 GeV beam energy and has a circumference of 288 m.

4.4 Minimum emittance

Solving the integral over \( \mathcal{H} \) from Eqs. (19) and (20) and minimizing the result in respect to the values of \( \alpha_x, \beta_x, D \) and \( D' \) at the magnet centre or entrance gives the theoretical minimum emittance [9,18,26,27].

Assuming that there are identical cells with only one type of bending magnet with deflection angle \( \Phi \), and further assuming that \( \Phi/2 \ll 1 \), which is valid for most light sources (\( \Phi < 20° \) gives < 1% error), the emittance can be written as

\[
\epsilon_{xo} \,[\text{nm-rad}] = 1470 \left( \frac{E[\text{GeV}]}{J_x} \right)^2 \frac{\Phi^3 F}{12 \sqrt{15}}
\]

(21)

with \( \Phi \) the deflection angle per bending magnet in radian and \( F \geq 1 \) a factor depending on the lattice type. The theoretical minimum emittance is achieved for \( F = 1 \). Table 2 gives the \( F \) values for some existing high brightness light sources: The region of operation generally is given by \( F > 3 \).

Note that emittance is independent from the bending radius, resp. the magnetic field, but increases cubically with the angle per bending magnet. That’s why light sources have many cells with relatively short bending magnets.

For two basic situations as shown in Fig. 6 the constraints on betafunction and dispersion for obtaining lowest emittance and the minimum factor \( F \) can be calculated:

- Center bending magnet: The beam has a focus \( (\alpha_{xc} = D'_c = 0) \) at the magnet center, dispersion and betafunction are symmetric in respect to the bend center and the dispersion is nonzero.
everywhere. Then we get, with $L = \rho \Phi$ the magnet length:

$$\beta_{xc} = \frac{1}{2\sqrt{15}} L \quad D_c = \frac{1}{24\rho} L^2 \quad \Rightarrow \quad F = 1 .$$

– End bending magnet: The beam enters the bending magnet with zero dispersion. Then we get a constraint for the distance $s_f$ of the focus (where $\alpha_x = 0$) from the entrance edge and for the betafunction at that focus:

$$s_f = \frac{3}{8} L \quad \beta_{xf} = \sqrt{\frac{3}{320}} L \quad \Rightarrow \quad F = 3 .$$

It is interesting to note the following:

– Zero dispersion in the center of a bending magnet does not provide the minimum emittance.

– A lattice consisting of only center bends with dispersion everywhere provides the lowest emittance achievable, it is 3 times lower than in a DBA (double bend achromat) lattice where each cell is made from two end bending magnets. As a consequence, DBA lattices starting with dispersion free straights are sometimes later tuned into a dispersive mode in order to further reduce the emittance. However the local effective emittance relevant for the brightness has to include the projection of momentum spread $\sigma_\delta$ to the horizontal dimension via dispersion and is given by

$$\epsilon_{x,\text{eff}}(s) = \sqrt{\epsilon_{z0}^2 + \epsilon_{xo} H(s) \sigma_\delta^2} . \quad (22)$$

– In TBA and higher bend achromats the end magnets should be made shorter by a factor $\sqrt[3]{3}$ in order to compensate for the factor 3 larger value of $F$ [15].

– FODO lattices are not suitable for light sources since $F \approx 100$, except modern structures with rather different types of combined function magnets used for the F- and D-magnets as shown in Fig. 4.f.

– As to be seen from Table 2 ELETTRA almost operates at the minimum emittance achievable for a DBA lattice with dispersion free straight sections. MAX-2 and ESRF also achieving low $F$-values take into account slightly dispersive straights.

Now deviations from the ideal emittance condition will be investigated [13]. This consideration is restricted to a lattice with center bends like shown in Fig. 6 (left). Defining dimensionless parameters

$$b = \frac{\beta_{xc}}{\beta_{xc,\text{min}}} \quad d = \frac{D_c}{D_{c,\text{min}}}$$
Fig. 7: Ellipses of constant ratios $F$ of emittance to minimum emittance as a function of the deviation from ideal dispersion and $\beta_x$ values for obtaining minimum emittance. Also shown are lines indicating the phase advance per cell [13].

Fig. 8: Minimum emittance cell: The 10°-gradient free sector bending magnet with optimum betafunctions and dispersion in the center ($b=d=1$) creates the minimum emittance ($E_x$) of 1.5 nm-rad at 3 GeV. The tune advance ($Q_x$) of 0.7902 corresponds to the ideal phase advance of $\Psi = 360° \cdot \Delta Q_x = 284.5°$. 
with the index \( \min \) denoting the ideal values to obtain the minimum \( F = 1 \), and introducing them into \( H \) after some algebra leads to the equation of an ellipse

\[
\frac{5}{4} (d - 1)^2 + (b - F)^2 = F^2
\]  

shown in Fig. 7. In order to learn more about the cell providing a factor \( F \) we impose constraints on periodicity, i.e., \( \alpha_x = D' = 0 \) at the entrance and exit of cell. Due to symmetry \( \beta_x \) and \( D \) have same values at entrance and exit anyway. In the approximation of small deflection angle \( \Phi/2 \ll 1 \) the matrix \( B \) transforming \( \{x, x'\} \) from center to exit of the bending magnet is given by

\[
B = \begin{pmatrix} 1 & L/2 & \rho \phi^2/8 \\ 0 & 1 & \phi/2 \end{pmatrix}
\]

with the third column describing the dispersion production (it is a submatrix from Eq. (5) containing the elements relating \( x, x' \), \( \delta \)). Now the rest of the cell, from bend exit to cell exit, may be described by a matrix \( M \) from which is only known, that it contains no other bending magnet and that it is symplectic of course, i.e., \( |M| = 1 \). Starting with the optical parameters described by \( b \) and \( d \) in the magnet center, the matrix \( M \cdot B \) has to zero \( \alpha_x \) and \( D' \), otherwise the solution would not be periodic. This provides constraints for \( M \). The detailed structure of \( M \) is less interesting than the full cell betatron phase advance \( \Psi \), which is calculated from

\[
\cos \Psi = \frac{1}{2} \text{Trace} \left( M \cdot B \left( \begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array} \right) (M \cdot B)^{-1} \left( \begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array} \right) \right)
\]

since the first half of the cell is the mirror image of the second half. The result is found as

\[
\Psi = 2 \arctan \left( \frac{6}{\sqrt{15}} \frac{b}{(d - 3)} \right).
\]

This equation describes lines of constant \( \Psi \) value in the \( (b, d) \) plane intersecting at \( d = 3, b = 0 \), as shown in Fig. 7. Reaching the minimum emittance requires a phase advance per cell of 284.5\(^\circ\). Ideal lattices based on those cells have been studied [8]: They require ‘empty cells’ alternating with the magnet cells in order to accomodate the additional focus for obtaining high phase advance. An example of a minimum emittance cell is shown in Fig. 8. Existing light sources operate at \( \Psi < 180^\circ \) and accept a larger emittance of \( F \approx 3 \ldots 5 \), see Table 2.

### 4.5 Damping partitions and energy spread

In Eq. (18) appears the quantity \( J_x \) as another possible way to reduce the emittance. The damping partition numbers are given by

\[
J_x = 1 - D \quad J_y = 1 \quad J_s = 2 + D \quad \text{with} \quad D = \frac{1}{2\pi} \int_{mag} D(s)(b_1(s))^2 + 2b_2(s) \, ds.
\]

The integral is only to be taken over the bending magnets where \( b_1 \neq 0 \). If the bending magnets also use large gradients like in some low emittance lattice concepts [16], the gradients have to be adjusted carefully in order to ensure damping, in all dimensions: \(-2 < D < 1 \). In a separate function light source lattice with no gradient in the bends, \( D \approx 1 \). Note that \( \sum J_i = 4 \), i.e., the damping may be shifted between the dimensions but the sum is limited. – The damping times depend on the energy loss per turn \( U \) and on the partition numbers:

\[
\tau_i = \frac{2CE}{cU J_i} \quad \text{with} \quad U[\text{keV}] = 26.5(E[\text{GeV}])^3 B[\text{T}]
\]

for an isomagnetic lattice (neglecting undulator radiation and wake field losses).
The r.m.s. energy (or momentum) spread of the beam is relevant for the effective emittance according to Eq. (22) if undulators are placed in dispersive sections. In practical units it is given by

$$\sigma_\delta = 6.64 \cdot 10^{-4} \cdot \sqrt{\frac{B[T] E[GeV]}{J_s}}.$$  

Thus something should be left for $J_s$ when shifting the partitions around.

4.6 Vertical emittance and beam sizes

Ideally, a flat lattice (i.e., no vertical bending magnets) has no vertical emittance, as clear from Eq. (18) with $H_y = 0$ everywhere. In reality however, spurious vertical dispersion and linear coupling create a finite vertical emittance.

Vertical dispersion is caused by roll errors of bending magnets and by vertically misaligned quadrupoles. Linear coupling between horizontal and vertical betatron oscillations is introduced by skew quadrupole fields from vertically misaligned sextupoles or from quadrupoles with roll errors.

If coupling is the main source the emittance ratio $g = \epsilon_y/\epsilon_x$, to be calculated from an integral over skew quadrupolar fields [11], tells how the natural horizontal emittance from eq. (18) is divided into horizontal and vertical emittance:

$$\epsilon_x = \frac{1}{1 + g} \epsilon_{xo} \quad \epsilon_y = \frac{g}{1 + g} \epsilon_{xo}.$$  

Careful orbit correction including beam based alignment of quadrupoles suppresses vertical dispersion, and dedicated skew quadrupole corrector magnets allow to minimize the linear coupling. Exploiting these two methods emittance ratios $g < 10^{-3}$ can be obtained. However, for VUV and soft X-ray experiments the photon diffraction phase space may be larger than the vertical emittance, thus only hard X-ray experiments profit from lowest $g$-values. On the other side, Touschek lifetime (see Section 5.1) is proportional to the bunch volume and thus scales with $\sqrt{g}$. Therefore, a moderate rather than the minimum achievable value of $g$ is often preferred.

With low coupling and negligible vertical dispersion, the beam will appear as an upright (i.e., not tilted or sheared) elliptical spot with Gaussian profiles and r.m.s. beam radii given by

$$\sigma_x(s) = \sqrt{\frac{\epsilon_x \beta_x(s)}{1 + g} (\sigma_\delta D(s))^2} \quad \sigma_y(s) = \sqrt{\frac{\epsilon_y \beta_y(s)}{1 + g}}.$$  

For further reading on coupling and beam sizes see refs. [5, 11, 28].

4.7 Chromaticity

From Eqs. (2) and (3) we see that the multipole strength $b_n$ is a function of momentum deviation $\delta$,

$$b_n = b_{no} \frac{p_o}{p} = \frac{b_n}{1 + \delta} \approx b_{no} (1 - \delta).$$  

As a consequence, the quadrupoles set to produce sharp foci in all bending magnets for the sake of low emittance, do not provide the same focussing strength for particles with momentum deviation. Thus the machine tune (total phase advance around the ring) will vary with momentum. The ratio is called chromaticity and given by [7]

$$\xi_x := \frac{dQ_x}{d\delta} = -\frac{1}{4\pi} \int_C b_2(s) \beta_x(s) \; ds \quad \xi_y := \frac{dQ_y}{d\delta} = \frac{1}{4\pi} \int_C b_2(s) \beta_y(s) \; ds.$$  

Chromaticity is naturally negative, i.e., the tunes decrease with energy since the quadrupoles become weaker. Strong quadrupoles at large betafunctions contribute most to chromaticity. Thus light sources suffer from large negative horizontal chromaticity due to the strong focusing. This must not be tolerated for two reasons:
LATTICES FOR LIGHT SOURCES

– **Momentum acceptance**: Some variation of momentum has to be accepted by the storage ring for reasons of beam lifetime (see Section 5.1). The machine tune has to stay away from integer or half integer numbers otherwise field or gradient errors will amplify coherently and destroy the beam. Thus even in the optimum case of \( \frac{1}{4 \xi} \) the momentum acceptance would be restricted to \( |\delta| < 1/(4 \xi) \) which would give an unacceptable small number for most machines.

– **Head tail instability**: Negative chromaticity excites the fundamental mode of the head tail instability, a collective oscillation of electrons in head and tail of the bunch leading to very fast beam loss. Suppression of the fundamental mode requires non negative chromaticity [6].

Sextupoles in dispersive regions of the lattice are used to compensate the chromaticity: Since a sextupole has a parabolic field \( B_y \sim x^2 \), it may be considered like a quadrupole in the vicinity of a point \( x_d \) by using the tangent: \( b_2(x_d) \approx 2b_3 x_d \). In a dispersive region, the particles are horizontally ‘ordered’ by momentum, i.e., \( x_d = D \delta \). Thus the appropriate combination of dispersion and sextupole strength will compensate the chromatic errors introduced by the quadrupole. Eventually, the chromaticities of a storage ring including sextupoles (but neglecting small contributions from bending magnets) are given by [7, 20]

\[
\xi_x = \frac{1}{4 \pi} \int_C (2b_3(s)D(s) - b_2(s))\beta_x(s)\, ds \quad \xi_y = -\frac{1}{4 \pi} \int_C (2b_3(s)D(s) - b_2(s))\beta_y(s)\, ds . \tag{31}
\]

Correcting both chromaticities requires two families of sextupoles with opposite polarities. In order to prevent the two families from counteracting each other, the sextupoles for horizontal chromaticity should be located at locations of large \( \beta_x \) and low \( \beta_y \) and vice versa. Of course, all sextupoles should be located at large dispersion.

### 4.8 Other lattice parameters

While pushing a lattice for lowest emittance, constraints on other lattice parameters have to be taken into account or side-effects have to be minimized:

– **Circumference**: For saving space and building costs one would like to make a compact machine. However small circumference is not always the best neither in lattice performance nor in cost. Inserting some space may relax the optics, improve the acceptance, etc. It is also important to take into account from the very beginning of lattice design all kinds of spaces required for magnet coils, beam position monitors, corrector magnets, absorbers, pumps, flanges, etc. as indicated by Fig. 1.

The circumference also must be an integer multiple of the RF wavelength \( C = h \lambda_{rf} \), with \( h \) called the harmonic number. \( h \) could have a nice prime factor decomposition allowing different filling patterns.

– **Periodicity**: Periodicity is the number \( N_{\text{per}} \) of identical supercells making up the total lattice (16 for ESRF, 3 for SLS in the example from Fig. 5). High periodicity has fundamental advantages:

  – simplicity: most optics calculations are based on the simple supercell;
  – stability: few systematic resonances (see below), better dynamic acceptance;
  – cost saving: larger series of a few different components.

Periodicity of existing machines ranges from 1 in DORIS to 40 in APS.

– **Working point**: The \( \{Q_x, Q_y\} \)-space, called the tune diagram is covered with resonances appearing as lines described by \( aQ_x + bQ_y = p \) as shown in Fig. 9. The resonance order is given by \( a + |b| \). If \( p \) is an integer multiple of \( N_{\text{per}} \), the resonance is systematic, i.e., amplified by the cell structure, else it is inhibited by periodicity. Even \( b \) identifies regular and odd \( b \) skew resonances. The magnets in a flat lattice are all of regular type, thus neither skew nor non-systematic resonances appear in the ideal lattice, but show up in the real lattice including multipolar and misalignment errors (see Fig. 9). – There are many constraints for placing the working point:
Fig. 9: Tune diagram for an ideal (left) and real (right) period-3 lattice. Solid lines are systematic, dashed non-systematic and dotted skew resonances. Larger thickness corresponds the lower order. Resonances up to 4th order are shown. The curved line is the tune walk for a wide range of momentum deviations. Here in this example, the working point at 20.82/8.28 turned out to be too close to the $Q_x + Q_y = 29$ sum resonance for the real machine and had to be changed.

- It must not be at integer to avoid closed orbit instability due to dipole errors.
- It must not be at half integer to avoid beam blow up due to gradient errors.
- It must not be at a 2nd order sum resonance to avoid mutual amplification of horizontal and vertical betatron oscillations.
- It has to stay away from sextupolar resonances (3rd order, see Section 5.6).
- Multiturn injection requires a fractional tune not too close to integer, $|\text{frac}(Q)| > 0.2$ in the plane of injection in order for the injected satellite to clear the septum in the turn following injection (see Fig. 10).
- Resistive wall instability requires a tune abover rather than below an integer.

Generally the fractional part of the tune is more important than the integer, but most important is the flexibility of the lattice to move the tunes independently in the tune diagram, since the best working point eventually is not found till operation. Thus lattice design has to ensure this flexibility.

**Momentum compaction factor** The relative difference in pathlength travelled by a particle at given relative momentum deviation $\delta$ within one revolution of the reference particle is expressed by the momentum compaction factor $\alpha$:

$$\frac{\Delta s}{C} = \alpha \delta, \quad \alpha = \int_C \frac{D(s)}{\rho} ds \quad (32)$$

with $C$ the machine circumference. Low emittance light sources have rather low values of $\alpha < 10^{-3}$ due to the low dispersion inside the bending magnets. They may become even isochronous, i.e., $\alpha = 0$ by introducing partially negative dispersion. At very low or even zero $\alpha$, the quadratic variation of pathlength with energy has to be taken into account. It can be controlled by means of sextupoles in order to obtain longitudinal stability, i.e., a closed RF bucket again.
5 Acceptance

Closed orbit stability is not enough. The lattice has to accept particles with some deviations from the ideal orbit in all six coordinates $x, x', y, y', \delta$ and $\Delta s$ in order to provide sufficient beam lifetime and to allow filing of the machine to the desired current. Accumulation of beam current requires mainly horizontal acceptance, and increasing the lifetime requires mainly vertical and momentum acceptance.

5.1 Lifetime

Performance in terms of brightness is only useful if the beam lifetime is sufficiently long to keep the beam current approximately constant and to keep the background from lost particles low enough to do the experiments. This requirement is only partially relieved when running top-up injections, because bad lifetime will still lead to enhanced background and to activation of injection elements. The most important processes of particle losses in light sources are:

- Touschek effect is scattering of particles within the bunch, leading to a transfer of transverse to longitudinal momentum exceeding the momentum acceptance. The loss rate is inversely proportional to the bunch volume, thus light sources with their low emittance beams suffer most from Touschek effect.
- Elastic scattering on residual gas nuclei leads to a transverse deflection and subsequent loss in regions of low aperture, which are usually the narrow vertical gaps of the undulators.
- Bremsstrahlung on residual gas nuclei leads to a change of electron momentum and thus also requires sufficient momentum acceptance, however the dependancy is much weaker than for Touschek scattering.

The three relevant lifetime times have the following, approximate scalings:

$$T_t \sim \frac{\gamma^3 \sigma_s}{I_{sb}} \epsilon_{xo} \sqrt{g} \left( \langle [\delta_{acc}(s)]^2 \rangle^2 \beta(s) \ldots \right)_C$$
$$T_{el} \sim \frac{\gamma^2 A_y P}{P}$$
$$T_{bs} \sim \frac{\delta_{acc}^{0.2}}{P}$$

with $\delta_{acc}$ the relative momentum acceptance (see Eq. (38)), $\sigma_s$ the rms bunch length, $I_{sb}$ the single bunch current, $\epsilon_{xo}$ the natural emittance from Eq. (18), $g$ the emittance ratio from Eq. (28), $P$ the residual gas pressure and $A_y$ the vertical acceptance (see Eq. (35)), assuming $A_y \ll A_x$ in any case due to the presence of narrow gap undulators in a light source lattice.

Touschek lifetime is defined as the beam half life time, since Touschek scattering is a space charge effect, i.e., a two-particle process leading to a hyperbolic beam decay. The residual gas lifetimes are defined as decay of beam current to $1/e$ of its initial value, since these effects are single particle processes leading to an exponential decay. For more information on lifetime see Refs. [17, 19, 24, 30, 31].

5.2 Injection

In order to achieve a large beam current in the storage ring, many beams ($\approx 100 \ldots 1000$) delivered by the linac/booster injection complex have to be accumulated in the storage ring. In order to keep stored particles while bringing in new particles, the scheme as sketched in Fig. 10 is applied: A closed bump of the stored beam brings it close to a current septum shielding a dipole field for inflecting the injected beam from the stored beam region. Due to technical limitations the bump has a typical duration of a few microseconds corresponding to a few turns. Beyond the end of the septum stored and injected beams propagate parallel. The bump is closed for the stored beam, whereas the injected beam performs a betatron oscillation around the stored beam orbit. This oscillation continues until radiation damping merges the injected beam into the stored beam. With damping times in the order of milliseconds, this takes some 1000 turns. During this time the injection beam requires a certain amount of horizontal acceptance as shown by the large ellipses in Fig. 10.
5.3 Acceptance definition

We distinguish physical acceptances, determined by the beam pipe diameters, and dynamic acceptances, defined by the onset of chaotic or unstable motion and particle loss beyond some critical amplitude due to nonlinear resonances. Generally acceptance is defined by the 6-dimensional phase space volume where particles are stable, i.e., where they perform bounded oscillations. In most machines the coupling between the subspaces is not too strong so we may separate horizontal, vertical and longitudinal acceptance as projections from 6D to 2D-spaces.

Then the acceptances are invariants of the lattice like the Courant-Snyder invariant in case of linear uncoupled betatron motion. Local projections of the transverse acceptance to real space \( \{x, y\} \) give the dynamic apertures, which are not invariants but depend on the local beta functions. The transverse acceptance is usually measured in \( \text{mm} \cdot \text{mrad} \), the aperture in \( \text{mm} \).

It is unusual to mention the two-dimensional longitudinal acceptance of the RF bucket, instead its projection to the axis of relative momentum deviation \( \delta \), called RF momentum acceptance, is quoted. The corresponding lattice momentum acceptance defines the \( \delta \)-range where non-zero transverse acceptances still exist.

Determination of dynamic acceptance is not trivial, since the equations of motion are nonlinear and thus not integrable in most cases. Stability of motion has to be proven by simulation, i.e., probing points in 6D-space on stability by tracking. A test particle is considered to be stable if it survives the tracking, i.e., its amplitudes stay within some limits. Of course this depends on the number of turns to be tracked. Electrons fortunately “forget” their history due to radiation damping, thus tracking one damping time usually is enough (\( 10^3 \ldots 10^4 \) turns).

It is a general criterion for lattice performance, that the pure dynamic acceptance, i.e., the phase space separatrix calculated excluding the cut off beyond beampipe, should be larger than the physical acceptance. Furthermore, the dynamic acceptance should have little nonlinear distortions, otherwise...
the actual available acceptance given as the dynamic acceptance including physical limitations, will be reduced.

5.4 Physical acceptance

An initial lattice design composed from ideal quadrupoles and bending magnets has a purely linear dynamics, and the dynamic acceptance is infinitely wide. (Actually this is not exactly true, since the equations of motions had been derived assuming paraxial motion, i.e., ”small” deviations from the closed orbit.)

A particle however will be lost if \(|x(s)| \geq a_x(s)| where \(a_x(s)| is the half width of the vacuum chamber. The linear betatron motion \(x(s)| of a particle is given by Eq. (7). Since we consider many turns we drop the betatron phase and find the physical acceptance as the maximum possible betatron amplitude, \(A = 2\alpha_{max}\), by identifying \(x(s)| with its limit \(a_x(s)| and taking the minimum from all locations in the lattice:

\[
A_x = \min \left( \frac{(a_x(s) - |D(s)\cdot \delta|)^2}{\beta_x(s)} \right). \tag{33}
\]

Since \(A_x| is an invariant of the linear betatron motion we get the local projection of the acceptance, i.e., the minimum and maximum \(x\)-values a particle can reach from Eq. (7):

\[
x(s) = \pm \sqrt{A_x \cdot \beta_x(s)} + D(s) \cdot \delta. \tag{34}
\]

Due to the absence of vertical dispersion the corresponding equations for vertical acceptance are simpler:

\[
A_y = \min \left( \frac{a_y(s)^2}{\beta_y(s)} \right), \quad y(s) = \pm \sqrt{A_y \cdot \beta_y(s)}. \tag{35}
\]

5.5 Momentum acceptance

Momentum and phase acceptance as provided by the RF bucket height and length can be considered like longitudinal physical acceptances, although they are dynamic actually, because they are almost constant along the lattice and decoupled from the transverse dynamics. But momentum acceptance is also restricted by the transverse acceptance of the lattice: From Eq. (33) we also see that \(A_x| disappears for momentum deviations \(|\delta| > \min(a_x(s)/|D(s)|)|, i.e., when the closed orbit hits the vacuum chamber.

Momentum acceptance of the lattice is relevant for the Touschek beam lifetime (see Section 5.1): The scattering events cause a sudden change in particle momentum while leaving the transverse coordinates almost constant. After the event, the two interacting particles’ vectors are given by \((\approx 0, \approx 0, \approx 0, \approx 0, \pm \delta, 0)| since the scattered particles come from the beam core where the transverse coordinates are very small. Now the particle will start a betatron oscillation relative to the dispersive closed orbit. For a linear lattice the amplitude of this oscillation is given by

\[
A_x = \gamma_{xo}(D_o \delta)^2 + 2\alpha_{xo}(D_o \delta)D_o' \delta + \beta_{xo}(D_o' \delta)^2 = \mathcal{H}_o \delta^2 \tag{36}
\]

with \(\alpha_{xo}, \beta_{xo}, \gamma_{xo}| the twiss paramaters at location ‘0’ where the scattering event occurred, \(D_o, D_o'\) the dispersion and its slope, and \(\mathcal{H}_o\) the lattice invariant from Eq. (19).

The particles will perform oscillations according to Eq. (7) resulting at another location \(s\) in the maximum excursion

\[
x(s) = \sqrt{A_x \beta_x(s) + D(s)\delta} \left( \sqrt{\mathcal{H}_o \beta_x(s) + D(s)} \right) \cdot \delta \tag{37}
\]

Like in derivation of Eq. (33) we identify \(x| with \(a_x| as loss criterion and get the local momentum acceptance for location ‘0’ as

\[
\delta_{acc}(s_0) = \pm \min \left( \frac{a_x(s)}{\sqrt{\mathcal{H}_o \beta_x(s) + |D(s)|}} \right). \tag{38}
\]
Thus the momentum acceptance provided by the lattice varies for different locations. From Eq. (38) it can be easily shown that the lattice momentum acceptance at location of maximum dispersion is half of its value for the dispersion free region.

Finally, the local momentum acceptance is the minor of the values from Eq. (38) and the momentum acceptance provided by the RF system, i.e., the height of the RF bucket. The RF momentum acceptance does not vary along the lattice and is given by

\[ \delta_{\text{RF}}^{\text{acc}} = \sqrt{\frac{2U\lambda_{\text{rf}}}{\pi E \alpha C} (\cot \varphi_s + \varphi_s - \pi/2)} \]  

with \( \sin \varphi_s = U/V_{\text{rf}} \) and \( \lambda_{\text{rf}}, V_{\text{rf}} \) RF wavelength and peak voltage, and \( U \) the energy loss per turn from Eq. (26). Beam lifetime calculations finally have to integrate scattering rates and momentum acceptances over the lattice [24].

5.6 Dynamic acceptance

The sextupoles to be installed for chromaticity correction (see Section 4.7) have dramatic side-effects, because they apply a parabolic, i.e., a nonlinear field. Thus the equations of motion become nonlinear and usually can not be integrated any longer. As a consequence, beyond some transverse amplitude the motion will become unstable thus defining finite dynamic acceptances.

Finding a distribution of sextupoles for correcting the large chromaticity and leaving sufficient dynamic acceptance, is the most challenging task in light source design! How to attack this problem systematically is described elsewhere [1, 25]. Here we only give a brief outline of the approach:

Analysis of the sextupole Hamiltonian reveals nine terms of first order in sextupole strength that affect the equations of motion. Two of them are the chromaticities, they are independant of betatron phases, i.e., quadrupoles and sextupoles contribute additive as expressed by Eq. (31). The seven other terms are phase dependant and thus show a resonant behaviour: The kicks from single sextupoles on the particles may partially cancel due to their different phase advances, however any residual kick may be amplified more or less, depending on the machine tune. As a consequence the so-called sextupolar resonances will be excited (see Fig. 9): There are two terms driving integer resonances of type \( Q_x \), one term driving third integer resonances \( 3Q_x \), two terms driving coupling resonances \( Q_x \pm 2Q_y \) and two terms driving chromatic half integer resonances \( 2Q_x, 2Q_y \) for off-momentum particles.

Quadrupoles contribute to four of the nine terms: They are the source of the two chromaticities, i.e., the variation of tunes with momentum, and they also contribute to the \( 2Q_x, 2Q_y \) terms, which cause variation of the betafunctions with momentum and subsequently higher order chromaticities. Clearly, the ideal distribution of sextupoles in a lattice would compensate the four chromatic quadrupole terms while maintaining a total cancellation of the five adverse sextupole terms through appropriate phase advances.

Since all nine terms are of first order (i.e., linear) in sextupole strength, they form a linear system of equations. If we have \( M \) sextupole families, a procedure like singular value decomposition (SVD) would return the \( M \)-vector of sextupole strengths for the optimum sextupole pattern. Basically, for \( M \geq 9 \) a solution should exist that corrects all chromatic effects while not exciting any resonances.

In practise however, in particular in low emittance light sources, the linear system tends to degenerate, because the horizontal betatron phase advance per cell is rather close to \( 180^\circ \) due to strong horizontal focussing for achieving low emittance (see Fig. 7), thus the individual sextupole contributions to the phase dependant \( 2Q_x \) term add up coherently, and no solution will be found for the sextupole pattern. Instead appropriate phase advances have to be inserted into the machine, mainly by exploiting the straight sections of the storage ring for that purpose, to create a sextupole pattern (in terms of betatron phases) which shows less degeneracy and may have a chance to find a set of strengths that works. Clearly, manipulating the straights for the sake of phase advances has a large impact on the general machine layout and performance. As a consequence, the following guideline emerges: for lattices with large
chromaticites linear and nonlinear lattice design are not decoupled. One must not proceed by designing the bending magnet and quadrupole arrangement first and add the sextupoles later. Instead from the early planning of the lattice the sextupole pattern has to be taken into account.

Another complication arises from the crosstalk between sextupoles causing higher order effects. If the sextupoles are rather strong, as it is the case in a low emittance light source, also the second order sextupole Hamiltonian has to be considered. It consists of 13 terms causing tune shifts with amplitude, second order chromaticities and excitation of octupolar resonances. Any light source design has to consider at least the amplitude dependant tune shifts in order to achieve sufficient dynamic acceptance.

In practice, a minimization procedure will vary the vector of sextupole family strengths in order to suppress a penalty function constructed by application of suitable weight factors to first and second order terms. Results achieved during this procedure have to be checked by tracking calculations.

Figure 11 gives an example of successful dynamic acceptance recovery through minimization of the 9 first order and 13 second order sextupole terms by means of 9 sextupole families.

References
[4] Cern accelerator school convention, see for example Internal report CERN 85-19, p.xv or CERN 87-03, p.3
[8] D.Einfeld et al., A lattice design to reach the theoretical minimum emittance in a storage ring, Proc.EPAC’96, p.638


[18] A.Ropert, Low emittance lattices, CERN 95-06, p.147, CERN 1995


[29] E.J.N.Wilson, Nonlinearities and resonances, CERN 95-06 (1995) 15


SOURCES OF EMITTANCE GROWTH

D. Möhl
CERN, Geneva, Switzerland

Abstract
This note discusses a variety of mechanisms that can lead to a blow-up of the normalized emittance of a particle beam: mismatch at transfer from one machine to the next; scattering on a foil or an internal target, interaction with the residual gas; crossing of resonances; power supply ripple; collective instabilities; intra-beam scattering.

1 INTRODUCTION
The notion of beam emittance has been introduced in previous lectures at this school. From these you know that ideally the normalized emittance \( \varepsilon^* = \varepsilon \beta \gamma \), i.e., the emittance \( \varepsilon \) multiplied by the relativistic factors \( \beta \) and \( \gamma \) is invariant during acceleration. Reality is different! Take the example of the Large Hadron Collider (LHC) under construction at CERN [1]. It comprises in its injector chain the existing cascade of: proton source, Radio-Frequency-Quadrupole linac (RFQ), Alvarez linac, Booster (PSB), Proton Synchrotron (PS) and Super Proton Synchrotron (SPS). Ideally, the normalized emittance in the whole cascade up to the LHC at top energy should be constant and equal to that at the exit of the ion source. In reality, a blow-up of 20 altogether is expected. It builds up in the various stages of acceleration and transfer from one machine to the next (Fig. 1). Yet, in the design of the LHC, utmost care was taken to minimize the emittance blow-up in order to reach the highest possible luminosity (beam density) of the colliding beams. What then is the source of this large dilution?

In this paper we point out various phenomena that lead to beam blow-up. In Section 2 we recall the definition of emittance and introduce a basic mechanism, called filamentation, that is involved in virtually all emittance blow-up processes. The rest of the paper is concerned with various single-particle and collective effects leading to filamentation and its consequence, emittance increase.

Fig. 1: Emittance history in the LHC and its injector chain (LHC design assumptions [1]). The blow-up, concentrated in the drawing at the transit from one machine to the next, occurs partly at transfer and partly during acceleration in the subsequent machine. The large increase at injection to the booster is partly explained by the multturn injection used there (three turns but with losses).
2  EMITTANCE DEFINITION

2.1  Betatron equation

For convenience we recall here the equation of linear betatron motion [2]–[6], which naturally leads to the definition of emittance. The transverse deviation \( x(s) \) of a particle from the design orbit in a beam line or a storage or accelerator ring is governed by the betatron equation

\[
x(s)' + K(s) x(s) = 0
\]

Here \( s \) is the distance along the design orbit, the dash (') denotes derivation with respect to \( s \). It is assumed that the particle moves with constant (or very slowly changing) velocity \( (s) \) in the \( s \) direction. The function \( K(s) \), in linear approximation a function of only \( s \), is given by the arrangement of focusing elements.

The solution for particle ‘\( i \)’ is conveniently written in ‘quasi sinusoidal’ form as:

\[
x = A_i \sqrt{\beta_s} \cos(\Psi + \delta_i)
\]

or equivalently

\[
x = A_i \sqrt{\beta_s} \cos(\Psi + \delta_i)
\]

\[
\beta_i \equiv \alpha_s x + \beta_i x' = -A_i \sqrt{\beta_s} \sin(\Psi + \delta_i)
\]

The amplitude function \( \beta_s \), required to perform the transformation from Eq. (1) to Eqs. (2) and (3), is related to \( K(s) \) by another differential equation (the envelope equation [2]–[6]). Usually \( \beta_s(s) \) as well as other optical parameters are obtained by lattice programs like MAD [7]. All one has to do is to input the focusing structure \( K(s) \) and appropriate boundary conditions. In a circular machine (with circumference \( C \)), the ‘cyclic’ condition \((\beta(s), \beta'(s)) (s=0)=(\beta(s), \beta'(s))(s=C)\) leads automatically to a unique solution. In a beam line, usually \( \beta_s \) and \( \beta'_s \) at the entrance are prescribed in order to match the \( \beta \) function of the preceding stage. With this or any other two-boundary conditions, the \( \beta \) function of a beam line with a given arrangement of focusing elements is also fully determined.

The phase function \( \Psi(s) \) and the function \( \alpha_s(s) \) follow directly from \( \beta_s(s) \):

\[
\Psi = \int \frac{1}{\beta_s(s)} ds, \quad \alpha_s = -\frac{1}{2} \beta'(s)
\]

The following function is also frequently used:

\[
\gamma_s(s) = \frac{1 + \left(\frac{1}{2} \beta'(s)\right)^2}{\beta(s)}
\]

Finally, the constants \( A_i \) and \( \delta_i \) in Eqs. (2) and (3) are given by the initial conditions of particle \( i \).

All this has been treated in numerous articles, lecture notes, and textbooks on transverse beam dynamics (see, for example Refs. [2]–[5]) since the classical paper of Courant and Snyder [6]. The essentials are repeated here to be self-contained and to introduce the notation.
2.2 Single-particle emittance

Equation (2) represents an ellipse in the \((x, x')\) ‘phase plane’, as sketched in Fig. 2, with \(\Psi(s) = 0 \rightarrow 2\pi\) as parameter. The inclination of the ellipse is given by \(\alpha_x\) (in fact the angle of the axis against the \(x\)-direction is \(\xi = -1/2 \arctan \left(2\alpha_x/\left(\beta_x - \gamma_x\right)\right)\)). Some other properties are as indicated in Fig. 2. For a circular machine you may think of sampling \(x'_i\) and \(x_i\) turn by turn at a fixed position on the circumference. The phase \(\Psi(s)\) increases by \(\Delta\Psi = \Psi(C) - \Psi(0) = 2\pi Q\) during each turn. Then, for a non-integer betatron tune \(Q\), the sampling points trace the ellipse (Fig. 3).

**Fig. 2:** Single-particle \((x, x')\) phase-space trajectory (from Ref [2])

**Fig. 3:** Phase-space coordinates \((x, x')\) on consecutive turns observed at a fixed azimuthal position (adapted from Ref. [8])
The area of this ellipse, \( \pi A_i^2 \), is entirely determined by the amplitude \( A_i \) of the motion. This area is closely related to the emittance. We use the convention that the \( \pi \) is ‘absorbed into the units’. Thus \( \varepsilon_i = A_i^2 \) (with units \( \pi \) m rad) is the definition of the ‘single-particle emittance’ or ‘Courant and Snyder invariant’. It can be expressed in various different forms (see Fig. 2) e.g., \( \varepsilon_i = A_i^2 = \frac{x_{i_{\text{max}}}}{\gamma_i} \), or [for any pair \((x_i, p_i)\)], \( \varepsilon_i = \frac{(x_i^2 + p_i^2)}{\beta_i} \), or especially
\[
\varepsilon_i = \frac{x_{i_{\text{max}}}}{\beta_x}.
\] (6)

Please remember that throughout we define the emittance as ‘area’ in phase space with units ‘\( \pi \) radian metre’ [\( \pi \) rad m]! In some publications (e.g. [2]) \( \pi A_i^2 \) is defined as emittance.

Turning to Eq. (3), we note that it represents a circle in the \((x, p_x)\) plane (Fig. 4) with radius \( r_i = A_i \sqrt{\beta_i} \) and ‘area’ \( A_i^2 \beta_i = \varepsilon_i \beta_i \). In the following we shall widely use the representation in the \((x, p_x)\) plane (called ‘normalized phase space’). Closely related to it is the \( \left[ (1/\sqrt{\beta_i})x, (1/\sqrt{\beta_i})p_x \right] \) plane (which we will call ‘emittance plane’), where the betatron motion Eq. (3) is a circle with radius \( A_i \) and ‘area’ \( \varepsilon_i \). Note that \( (1/\sqrt{\beta_i})x \) is called \( \eta \) by Courant and Snyder [6] and hence
\[
\frac{d\eta}{d\psi} = \beta_x \frac{d\eta}{ds} = \frac{1}{\sqrt{\beta_x}} p_x.
\]

**Fig. 4:** Single-particle trajectory in normalized \([x, p_x]\) phase space

### 2.3 Beam emittance

For a beam, the trajectories of different particles in \((x, p_x)\) space are concentric circles (Fig. 5), with radii given by the amplitudes of the particles. In a ‘matched beam’ (which is adapted to the phase-space trajectories given by the \( \beta \) function of the machine or the beam line, see Section 2.4 below), there is no correlation between the amplitudes and the phases. In other words: for each amplitude value there is a set of particles with their betatron phases \( \Psi \) distributed uniformly from 0 to \( 2\pi \).
The beam emittance is defined as ‘some average’ of the single-particle emittances. Two definitions, referring to different averaging processes, are frequently used: 

i) Emittance definition referring to a fixed fraction of particles:

- The ‘geometrical beam emittance’ \( \varepsilon_\% \) is the area/\( \beta_x \) of the circle in \((x, p_x)\) space, that contains the motion of a given fraction \( F \) of the particles. Frequently used are: \( F = 40\% \), \( 85\% \) or \( 95\% \). For low-energy devices, \( F = 63\% \) (corresponding to the fraction \( 1 - 1/e \)) is sometimes used and the ‘brilliance’ is defined as the intensity contained in \( \varepsilon_{63}\% \).

ii) Emittance definition referring to the standard deviation of the projected distribution:

- Let \( \sigma_x \) be the standard deviation of the particle density in \((x, p_x)\) space, projected on the \( x \)-axis (i.e. the r.m.s. beam size measured, for example, by a profile detector). Then the ‘\( k \) r.m.s. emittance’ (also called ‘\( k \sigma \) emittance’) is the area/\( \beta_x \) of the circle of radius \( r_c = k \sigma_x \) in \((x, p_x)\) space namely

\[
\varepsilon_{k \sigma} = (k \sigma_x)^2 / \beta_x .
\] (7)

- Frequently used are: \( k = 1 \) (‘r.m.s. emittance’), \( k = 1 \) or \( 2.5 \).

For hadrons, i.e., proton and ion beams, the geometrical emittance is appropriate to determine aperture requirements (loss) whereas for both hadrons and leptons the r.m.s. emittance is useful to determine the luminosity (density) of beams. The conversion of one into the other form depends on the particle distribution. For beam with a Gaussian distribution of the projected density, one has \( F_{\text{Gauss}} = 1 - \exp(-k^2 / 2) \). Examples are given in Table 1.
Table 1: Fraction of particles of a Gaussian beam with their motion circumscribed by a circle of radius \( k\sigma_s / \sqrt{\beta_s} \) in emittance space [i.e. contained in the \( k\sigma \) emittance; \( \varepsilon = (k\sigma^2_s) / \beta_s \)]

<table>
<thead>
<tr>
<th>( k )</th>
<th>( F_{\text{Gauss}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>39.3%</td>
</tr>
<tr>
<td>2</td>
<td>86.5%</td>
</tr>
<tr>
<td>2.5</td>
<td>95.6%</td>
</tr>
</tbody>
</table>

Note the \( k^2 \) in definition (7) and remember also that \( \sigma_s \) is the r.m.s. of the projected density. Some authors call \( \varepsilon = k(\sigma^2_s) / \beta_s \) the \( k \) r.m.s. or \( k \) emittance and there is further confusion as ‘r.m.s. emittance’ is sometimes used for the r.m.s. value of the single-particle emittances \( \sqrt{\langle \varepsilon^2_i \rangle} \) which is different from definition (7) with \( k = 1 \).

2.4 Acceptance

The half-aperture \( (a_s) \) and the \( \beta \) function of a beam channel determine the quantity

\[
A_s = a_s^2 / \beta_s
\]

which resembles the emittance definition [Eq. (6) or (7)]. Both \( a_s \) and \( \beta_s \) vary with position \( (s) \). The minimum of (8) defines the acceptance (also called ‘admittance’), i.e. the maximum single-particle emittance that can be transmitted. If a certain percentage of losses (e.g. 5%) should not be exceeded, the corresponding geometrical emittance (e.g. \( \varepsilon_{\text{geo}} \)) can be at most equal to the acceptance. This determines the aperture required in hadron machines. Once tails of the beam are chopped off, the truncated distribution remains in the machine unless ‘emittance heating’, e.g., by scattering processes (discussed in Section 4) comes into play.

For lepton (electron, positron …) machines, where synchrotron radiation is important, the situation is somewhat different: the emittance results from the equilibrium between radiation damping and quantum excitation and each particle has a finite chance to attain large amplitude during the random-walk process. In this case, the aperture has to be ±7\( \sigma \) or even ±10\( \sigma \) for a good lifetime of the beam.

2.5 Filamentation

Having defined the beam emittance, we now turn to the dilution mechanism called filamentation. In Fig. 6 we illustrate it for the example of mismatch at injection, to be discussed in more detail in Subsection 3.3 below.

The \( \beta \) function and its derivative \( \beta'(s) \) at the injection position in a circular machine define phase ellipses sketched in Fig. 6(a). When the transfer line has the same \( \beta \) and \( \beta' \) at the injection point (i.e. at the exit of the line) then the ellipses of the injected particles coincide with those of the machine and the beam can be matched without blow-up. For perfect matching, the dispersion functions and their derivatives have also to be equal (see Subsection 3.2 below). If, on the contrary, \( \beta(s) \) and/or \( \beta'(s) \) are different, then the emittance ellipses of the incoming beam have a shape and/or an orientation different from the machine phase-space [Fig. 6(b)]. As the beam circulates in the machine, each particle will turn around in transverse phase-space on the ellipse corresponding to its injection conditions \((x, x')\) with an angular frequency \( Q \).

For a strictly linear machine, \( Q \) is the same for all particles; but with (even small) non-linearity, the betatron frequency depends on the amplitude of the particle. Thus particles move with different
speed on different phase ellipses. The beam (shaded area in Fig. 6) changes from its original form first into an S-shaped area [Fig. 6(c)] and then after long-enough time into a diluted distribution that fills the machine phase-space almost uniformly [Fig. 6(d)]. In other words, the initial correlation between amplitudes and phases of the particles has practically vanished. This ‘smearing-out’ process is called filamentation. Clearly the phase-space area needed to contain the beam after filamentation is larger than that of the injected beam.

Filamentation occurs in circular machines. In a beam line or a linac, the number of betatron oscillations is usually not large enough to lead to complete ‘smear-out’. Yet here also, extra aperture has to be provided to contain the phase-space rotations of a mismatched beam. After mismatch in an injection line or a linac, full filamentation takes place in the subsequent circular machine.

In the following, we shall discuss various mechanisms that lead to mismatch and thus emittance dilution through filamentation.

![Fig. 6: Filamentation of a mismatched beam (from Ref. [2])](image)

### 3 EMITTANCE INCREASE DUE TO MISMATCH AT TRANSFER

#### 3.1 Steering error

 Assume a beam injected into a circular machine with an error in position $(\Delta x)$ and angle $(\Delta x')$ with respect to the design orbit. Then, in the normalized $(x, p_x)$ phase space, the emittance circles of the beam are no longer concentric with the admittance circles of the machine (Fig. 7). Rather, they are radially displaced by an amount

\[
\Delta x = \sqrt{(\Delta x)^2 + (\Delta p_x)^2} = \sqrt{(\Delta x)^2 + (\alpha_x \Delta x + \beta_x \Delta x')^2}.
\]

After filamentation, a beam with a phase-space area which was originally $r_0^2$ fills the area $(r_0 + \Delta r)^2$ (Fig. 7). The increase in geometrical emittance is

\[
\varepsilon_y \rightarrow \varepsilon_y (1 + \Delta r / r_0)^2.
\]

Similarly for the r.m.s. emittance one finds (Appendix A)

\[
\varepsilon_{\sigma_x} \rightarrow \varepsilon_{\sigma_x} \left(1 + \frac{1}{2} \left(\Delta r^2 / \sigma_{x,0}^2\right)\right).
\]

These results are independent of the initial and final distribution. However, the shape of the final distribution depends strongly on the magnitude of the error. For large errors (compared to the standard deviation of the original distribution), the beam centre becomes depopulated and double-humped distributions emerge (Fig. 8).
Fig. 7: Increase of emittance due to a steering error at injection

Fig. 8: Projected distribution of an originally Gaussian beam after filamentation of a position error $\Delta x$ at injection (From Ref. [2].)
These double humps are a sign of big steering errors at transfer in contrast with a mismatch of the $\beta$ function (to be discussed in Subsection 3.3), which leads mainly to a widening of the distribution.

### 3.2 Momentum error and dispersion-function mismatch

Assume that a beam with a momentum $p + \delta p$ is injected into a machine adjusted for the momentum $p$ on its central orbit. If the dispersion $D$ of the line and the machine are matched ($D$ and $D'$ are the same) and the beam is injected onto the off-momentum orbit with $\delta x = D\delta p/p$ and $\delta x' = D'\delta p/p$, then perfect matching can be achieved. If, however, injection is made directly onto the central orbit, then this is equivalent to a steering error $\Delta x = D\delta p/p$, $\Delta x' = D'\delta p/p$ and our previous results [Eqs. (10) and (11)] can be used putting

$$\Delta r = \left\{D^2 + (\beta_x D' + \alpha_x D)^2\right\}^{1/2} \cdot \frac{\delta p}{p}. \quad (12)$$

Next, imagine that the dispersion functions of the line and the machine at the injection location differ. We can assume without loss of generality that the machine has $D = 0$ and $D' = 0$ at the injection point (in the general case replace $D \rightarrow D + D'$ in the following). For each momentum $p + \Delta p$ we have an equivalent steering error $\Delta x = D\Delta p/p$ and $\Delta x' = D'\Delta p/p$. Thus the momentum distribution is folded into the transverse phase-space. The blow-up in geometrical emittance after filamentation depends on both the transverse and the momentum distribution. An approximation is to calculate the blow-up from Eq. (11) putting

$$\Delta r = \left\{D^2 + (\beta_x D' + \alpha_x D)^2\right\}^{1/2} \cdot \frac{\delta p}{p} \cdot \left(\frac{\sigma^2}{\sigma_0^2}\right). \quad (13)$$

Here $\pm (\Delta p/p)_b$ is momentum spread containing the prescribed fraction of the momentum distribution.

For the increase of the r.m.s. emittance one obtains (Appendix B)

$$\epsilon_{\sigma} \rightarrow \epsilon_{\sigma} \left\{1 + \frac{1}{2} \left\{D^2 + (\beta_x D' + \alpha_x D)^2\right\} \left(\frac{\sigma^2}{\sigma_0^2}\right) \right\}. \quad (14)$$

This result is independent of the shape of the momentum distribution and also independent of the initial and final transverse distribution. Equation (14) indicates that in the case of dispersion mismatch the momentum distribution is folded into the transverse emittance.

### 3.3 Focusing errors

Let the functions $\beta$ and $\beta'$ of the transfer line (subscript l) at the ‘hand-over point’ (exit of the line $\rightarrow$ entrance of the ring) differ from those of the ring (subscript r). Normalizing the ring phase-space trajectories to circles, the trajectories of the mismatched injected beam are ellipses. Particles contained in a given ellipse fill the circle of the ring phase-space circumscribing the corresponding ellipse, after filamentation. The area of this circle (and the new geometrical beam emittance) is by the factor $b/a$ (major axis/minor axis of the ellipse, Fig. 9) larger than that of the corresponding beam ellipse before filamentation: $\epsilon_{\sigma} \rightarrow \epsilon_{\sigma} \cdot (b/a)$. A similar relation is obtained for the increase of the r.m.s. emittance (Appendix C). In both cases the blow-up can be expressed in terms of a factor $F$, [2]–[3] (called $H$ in Ref. [3]) given by

$$F = \frac{1}{2} \left(\beta_y \gamma' + \beta_y - 2 \alpha_2 \alpha_1 \right) \quad \text{or equivalently} \quad F = \frac{1}{2} \left(\frac{\beta_x \gamma - \beta_x}{\beta} + \frac{\alpha_x - \alpha_1}{\beta} \right) \beta_x \beta_y \beta. \quad (15)$$

One obtains ([2]–[3] and Appendix C)

$$\epsilon_{\sigma} \rightarrow \epsilon_{\sigma} \left(1 + \sqrt{F^2 - 1}\right). \quad (16)$$
Fig. 9: Mismatch due to focusing errors

Table 2 gives a summary and a numerical example of the results for steering and focusing errors. Several observations can be made:

– the blow-up of the geometrical emittance is always larger than the blow-up of the r.m.s. emittance;
– the relative blow-up (of both geometrical and r.m.s. emittance) due to steering errors depends on the initial emittance, whereas
– the relative increase due to focusing errors depends only on the beta-function mismatch, and is independent of the initial emittance.

Table 2: Influence of steering and focusing errors on geometrical and r.m.s. emittance, comparison of formulae and a numerical example

<table>
<thead>
<tr>
<th>Type of emittance</th>
<th>Steering error, general and example for $\Delta r^2/\varepsilon_{\Delta x} = 0.1^a$</th>
<th>Focusing error, general and example for $F = 1.1^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blow-up $\dot{\varepsilon}<em>x/\varepsilon</em>{x,0}$ of emittance containing a given fraction of beam</td>
<td>general $\left(1 + \Delta r / \sqrt{F \beta_x}\right)^2$</td>
<td>$F\sqrt{F^2 - 1}$</td>
</tr>
<tr>
<td></td>
<td>example 1.73</td>
<td>1.56</td>
</tr>
<tr>
<td>Blow-up $\dot{\varepsilon}<em>{\text{rms}}/\varepsilon</em>{\text{rms},0}$ of r.m.s. emittance</td>
<td>general $\left(1 + \Delta r^2 / \varepsilon_{\text{rms}} \beta_x\right)$</td>
<td>$F$</td>
</tr>
<tr>
<td></td>
<td>example 1.05</td>
<td>1.1</td>
</tr>
</tbody>
</table>

$^a \Delta r = \sqrt{(\Delta x)^2 + (\alpha_x \Delta x + \beta_x \Delta v)^2}$ [Eq. (9)].

$^b F = \frac{1}{2} \left( \frac{\beta_t}{\beta_r} + \beta_t \left( \frac{\alpha_r - \alpha_t}{\beta_r \beta_t} \right)^2 \beta_t \beta_t \right)$ [Eq. (15)].
3.4 Mismatch due to non-linearity and space charge

Until now we have tacitly assumed that linear focusing predominates in both the transfer line and the machine. Weak (unavoidable) non-linearity was only considered as the source of filamentation. Stronger non-linearity complicates the matching because it also leads to a ‘distortion’ of the phase-space trajectories (Fig. 10). For example, an additional term of the form $K'(s)x^2$ in the betatron Eq. (1) tends to introduce triangular instead of circular phase-space trajectories (see, for example, Ref. [9]). Therefore, if the machine and/or the transfer line are significantly non-linear, then it becomes difficult, if not impossible, to keep the beam matched and thus to avoid emittance blow-up during transfer. In this context, space charge is important. It introduces both linear and non-linear terms, which complicate the matching.

Fig. 10: Phase space in the presence of a sextupolar non-linearity. For small-amplitude particles, the trajectories are circles. For larger amplitude they are no longer closed but tend towards infinity, thus indicating unstable motion. The triangular separatrix delimits the region of stable motion; outside it the motion is unstable motion. If the entire region inside the separatrix is filled by an injected beam with originally circular trajectories, the relative increase in geometrical emittance is $6/\pi = 2$. The area enclosed by the separatrix depends on the strength of the non-linearity and on the distance of the tune $Q$ from a third integer. (Adapted from Ref. [9].)

3.5 Feedback to damp oscillations due to injection errors

In the early part of the filamentation process, the beam centre (in the case of steering errors, see Fig. 11) or the beam size (in the case of focusing errors) oscillates. These oscillations can be detected by position (dipolar) pick-up electrodes or by beam profile (quadrupolar) electrodes, respectively. Similar dipolar or quadrupolar kicker electrodes can act upon the oscillation.
Fig. 11: A beam injected with a steering error (a) rotates in phase space. The projected density oscillates between the two extreme positions (b) and (c). The oscillation can be observed with position pick-up electrodes that record the centre of the projected distribution. As the beam filaments, the oscillation gradually ‘damps out’

A suitable arrangement of pick-up(s) and kicker(s) is used to reduce the effect of injection errors. The action of the feedback has to be fast compared to the filamentation time (given by the coherence time of the particles) which depends on the tune spread in the beam. The coherence time \( t = 1/\omega_{\text{rec}} \Delta Q \) typically extends over several hundred revolutions so that the feedback system has to correct only a small fraction of the injection error on each turn. In this way one can avoid or reduce the emittance blow-up. But the larger aperture (determined by the circumscribed circle in Figs. 7 and 11 that indicates the final emittance if filamentation were to occur) is still required to accommodate the oscillation.

Feedback systems (see e.g., Refs. [10,11]) are frequently used to damp the dipole oscillations at injection. Detection and damping of quadrupolar oscillations is more difficult and less popular. In fact, to my knowledge, complete feedback systems to damp beam-shape oscillations at injection have not yet been implemented in any existing machine.

4 EMITTANCE INCREASE DUE TO SCATTERING ON ‘OBSTACLES’

4.1 Scattering in a foil or a window or an internal gas target

Windows or foils are placed into transfer lines to separate regions of different vacuum, or to strip ions. The Coulomb scattering upon transition through such an ‘obstacle’ leads to an increased angular divergence of the beam (Fig. 12).

The blow-up of the r.m.s. emittance can be estimated using the classical multiple scattering formula, Ref. [12]:

\[
\theta_{\text{rms}} = \frac{13.6 \text{ MeV} / c}{p \beta_p q_p} \sqrt{\frac{L}{L_{\text{rad}}}} \left(1 + \delta_{\text{corr}}\right).
\]  

(18)
It gives the r.m.s. scattering angle (in each of the transverse planes) for a particle of charge number \( q_p \) (= 1 for protons), total momentum \( p \) [MeV/c], velocity \( \beta_p = v_p / c \), traversing a foil of a material with a ‘radiation length’ \( L_{\text{rad}} \) and of thickness \( L \); \( \delta_{\text{corr}} \) is a small correction factor to be neglected from here on. The increase in r.m.s. emittance of a traversing beam after filamentation (in the subsequent machine) can be calculated using (18) (Appendix D):

\[
\Delta \epsilon_{k\sigma} = \frac{1}{2} (k \theta_{\text{rms}})^2 \beta_s^2 .
\]  

(19)

This assumes that the beam is matched in the absence of the foil. By readjusting the focusing in the downstream part of the line, the increase can be somewhat reduced (Appendix D).

Equation (19) also applies to the blow-up per turn due to an internal gas target in a circular machine. Here the ratio \( L_{\text{rad}} / L \) in Eq. (19) is conveniently obtained if both the ‘target thickness’ (as is customary for targets) and the radiation length in the gas are given in g/cm\(^2\). Two observations can be made from Eq. (19):

- the blow-up is independent of the initial emittance and
- it is proportional to the value of the focusing function \( (\beta_s) \) at the foil or the target.

Hence the interest in having a low \( \beta \) at the location of the scatterer.

Examples are given in Tables 3 and 4. The stripping foil in the PS to SPS transfer line (Table 3) was used for the final stripping of lead ions. The emittance blow-up was acceptable for the fixed-target experiments at the SPS. For the LHC ion beam with a much stricter impedance budget, a low \( \beta \) (about 5 m instead of 20 m) will be created at the stripper position to reduce the emittance increase to tolerable values [13]. To this end, four additional quadrupoles will be inserted into the line.
Table 3: Calculated influence of the stripping foil in the transfer line between the PS and SPS. The scattering angle Eq. (18) is calculated taking the average \( q_p = 68 \) of the charge states before and after stripping.

<table>
<thead>
<tr>
<th>Beam</th>
<th>lead ions (( \text{Pb}^{208} ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy/nucleon</td>
<td>4.25 GeV/u</td>
</tr>
<tr>
<td>Momentum/nucleon</td>
<td>5.1 GeV/c/u</td>
</tr>
<tr>
<td>Total momentum/ion ( p )</td>
<td>1060 GeV/c/ion</td>
</tr>
<tr>
<td>Relativistic velocity factor ( \beta = v/c )</td>
<td>0.98</td>
</tr>
<tr>
<td>Relativistic mass factor ( \gamma = (1 - \beta^2)^{-1/2} )</td>
<td>5.56</td>
</tr>
<tr>
<td>Focusing function at target ( \beta_h = \beta_v )</td>
<td>23 m</td>
</tr>
<tr>
<td>Charge state ((q)) before/after stripping</td>
<td>53/82</td>
</tr>
<tr>
<td>Stripping target material</td>
<td>aluminium</td>
</tr>
<tr>
<td>Stripper thickness ( L )</td>
<td>0.08 cm (0.22 g/cm(^2))</td>
</tr>
<tr>
<td>Radiation length in aluminium ( L_{rad} )</td>
<td>8.9 cm (24 g/cm(^2))</td>
</tr>
<tr>
<td>r.m.s. scattering angle/traversal ( \theta_{\text{rms}} ) (for ( q = 68 ))</td>
<td>( 8.7 \times 10^{-5} ) rad</td>
</tr>
<tr>
<td>Blow-up of r.m.s. emittance ( \Delta \varepsilon_{\text{rms}, h,v} ) (&quot;&quot; )</td>
<td>( 8.5 \times 10^{-8} \pi ) rad m</td>
</tr>
<tr>
<td>Blow-up of normalized r.m.s. emittance ( \Delta \varepsilon_{\text{rms}, \beta \gamma} ) (&quot;&quot; )</td>
<td>( 0.47 \times 10^{-6} \pi ) rad m</td>
</tr>
<tr>
<td>Tolerable normalized emittance for SPS fixed-target experiments</td>
<td>( 4 \times 10^{-6} \pi ) rad m</td>
</tr>
<tr>
<td>Tolerable normalized emittance in the LHC at collision energy</td>
<td>( 1.5 \times 10^{-6} \pi ) rad m</td>
</tr>
</tbody>
</table>

Table 4: Calculated influence (at 180 MeV) of the hydrogen jet target used by the ‘JETSET’ experiment in LEAR.

<table>
<thead>
<tr>
<th>Circulating beam</th>
<th>antiprotons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Momentum ( p )</td>
<td>0.61 GeV/c</td>
</tr>
<tr>
<td>Velocity ( \beta_p = v/c )</td>
<td>0.55</td>
</tr>
<tr>
<td>Revolution frequency ( f )</td>
<td>2.1 MHz</td>
</tr>
<tr>
<td>Focusing function (low ( \beta )) at target ( \beta_h = \beta_h )</td>
<td>1 m</td>
</tr>
<tr>
<td>Internal target</td>
<td>hydrogen jet</td>
</tr>
<tr>
<td>Target thickness ( L )</td>
<td>( 8 \times 10^{-13} ) atoms/cm(^2) (( 1.3 \times 10^{-13} ) g/cm(^2))</td>
</tr>
<tr>
<td>Radiation length in hydrogen ( L_{rad} )</td>
<td>63 g/cm(^2)</td>
</tr>
<tr>
<td>r.m.s. scattering angle/traversal ( \theta_{\text{rms}, h,v} )</td>
<td>( 6.1 \times 10^{-8} ) rad</td>
</tr>
<tr>
<td>Blow-up/turn of r.m.s. emittance ( \Delta \varepsilon_{\text{rms}, h,v} )</td>
<td>( 1.8 \times 10^{-15} \pi ) rad m</td>
</tr>
<tr>
<td>Blow-up/sec of r.m.s. emittance ( \text{d} \varepsilon_{\text{rms}} / \text{d} t )</td>
<td>( 3.9 \times 10^{-9} \pi ) rad m s(^{-1})</td>
</tr>
<tr>
<td>Tolerable blow-up/sec (cooling strength)</td>
<td>( &lt; 5 \times 10^{-9} \pi ) rad m s(^{-1})</td>
</tr>
</tbody>
</table>
The hydrogen gas jet [14] was used in LEAR up to 1996. A small equilibrium beam size was desirable. It results from the ‘beam heating’ \( \frac{d\varepsilon_{\text{rms}}}{dt} \) at the target and ‘beam cooling’ by the stochastic cooling system. For cooling times (for \( 10^{10} \) particles) of the order of 200 s, equilibrium emittances of \( 1 \pi \text{ mm mrad} \) were attainable when the heating was less than \( 5 \times 10^{-9} \pi \text{ mm mrad s}^{-1} \) (corresponding to a target thickness of less than \( 10^{14} \) atoms/cm\(^2\)).

**4.2 Scattering on the residual gas**

This subject is treated in many papers of which I find the internal report by W. Hardt [15] especially instructive. Here we can use our previous results by taking the residual gas atmosphere as a ‘distributed thin scatterer’. For pure nitrogen (N\(_2\)) at pressure \( P \) the radiation length is \( L_{\text{rad}, N_2} = 327 \text{ m} / (P/760 \text{ Torr}) \) and the thickness traversed by the beam in time \( t \) is \( L = \beta_p c t \). Then from Eqs. (18) and (19) we get the blow-up of the \( k\sigma \) emittance as

\[
\Delta \varepsilon_{k\sigma} = \frac{1}{2} k^2 q_p^2 \left( \frac{14 \text{ MeV}/c}{p \beta_p} \right)^2 \bar{\beta}_s \frac{\beta_p c t}{327/(P/760 \text{ Torr})}.
\]

Here \( \bar{\beta}_s \) is the average around the ring of the beta function as scattering occurs all around the circumference; \( p = 931 \text{ MeV}/c \cdot A_p \cdot \beta_p \gamma_p \) is the total momentum, where \( A_p \) is the mass number of the ion (≈ 1 for protons). Equation (20) leads directly to Hardt’s formula [with a slight difference in the numerical factor as he takes 15 MeV/c in the scattering angle formula (18)]

\[
\Delta \varepsilon_{k\sigma} = 0.14 k^2 q_p^2 \left( \frac{\beta_p c}{A_p \gamma_p} \right) P t \frac{P}{\beta_s \gamma_p} (P \text{ [Torr]}, \beta_s \text{ [m]}, t \text{ [sec]}).
\]  

Equation (20) is widely used to determine the vacuum requirement in a storage ring. For a synchrotron, one has to integrate \( dt/\beta_p^2 \gamma_p \) over the acceleration cycle to obtain the blow-up of the normalized emittance. For a residual atmosphere with different gas components of partial pressures \( P_i \), we can define the N\(_2\) equivalent pressure for multiple Coulomb scattering as

\[
P_{N_2, \text{ equ}} = \sum P_i \left( \frac{L_{\text{rad}, N_2}}{L_{\text{rad}, i}} \right).
\]

The radiation length for some gases at atmospheric pressure is given in Table 5.

**Table 5:** Radiation length and density (at 760 Torr and 20 °C) for some gases (adapted from Ref. [12])

<table>
<thead>
<tr>
<th>Gas component</th>
<th>H(_2)</th>
<th>He</th>
<th>N(_2)</th>
<th>Ne</th>
<th>Ar</th>
<th>Air</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density (at 20 °C and 760 Torr)</td>
<td>0.084</td>
<td>0.166</td>
<td>0.116</td>
<td>0.839</td>
<td>1.66</td>
<td>1.20</td>
<td>g/dm(^3)</td>
</tr>
<tr>
<td>‘Radiation density’</td>
<td>63</td>
<td>94.3</td>
<td>38</td>
<td>28.9</td>
<td>19.6</td>
<td>36.1</td>
<td>g/cm(^2)</td>
</tr>
<tr>
<td>Radiation length (at 20 °C, 760 Torr)</td>
<td>7500</td>
<td>568</td>
<td>327</td>
<td>344</td>
<td>118</td>
<td>301</td>
<td>m</td>
</tr>
</tbody>
</table>

As an example for vacuum requirements in a storage ring, Table 6 summarizes the situation in LEAR at its lowest energy (5.3 MeV i.e. 100 MeV/c). At this energy, stochastic cooling worked with time constants (\( \tau \)) of say 3 minutes for \( 10^9 \) antiprotons when the r.m.s. emittances were \( 1\pi \text{ mm mrad} \). Thus a heating rate \( \Delta \varepsilon_{\text{rms}}/dt \) up to \( 5 \times 10^3 \pi \text{ mm mrad s}^{-1} \) could be tolerated leading to an equilibrium \( \pi \Delta \varepsilon_{\text{rms}}/dt = 1 \pi \text{ mm mrad} \). If smaller equilibrium emittances are desired, a lower vacuum or faster cooling (e.g., electron cooling) is required.
Table 6: Vacuum conditions in LEAR at its lowest energy

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circulating beam antiprotons</td>
<td></td>
</tr>
<tr>
<td>Momentum $p$</td>
<td>100 MeV/c</td>
</tr>
<tr>
<td>Velocity $\beta_p = v/c$</td>
<td>0.1</td>
</tr>
<tr>
<td>Revolution frequency $f$</td>
<td>0.4 MHz</td>
</tr>
<tr>
<td>Vertical acceptance</td>
<td>$125 \pi$ mm mrad</td>
</tr>
<tr>
<td>Average beta function $\bar{\beta}$</td>
<td>10 m</td>
</tr>
<tr>
<td>r.m.s. angle for $\Delta e_{rms} = 125 \times 10^{-6} \pi$ m rad</td>
<td>$5 \times 10^{-3}$ rad</td>
</tr>
<tr>
<td>Residual gas pressure (N$_2$ equiv.) $P$</td>
<td>$10^{-12}$ Torr</td>
</tr>
<tr>
<td>Radiation length in N$_2$ at $10^{-12}$ Torr</td>
<td>$2.3 \times 10^{17}$ m (25 light years)</td>
</tr>
<tr>
<td>Time for beam to traverse $2.3 \times 10^{17}$ m</td>
<td>27 h ($4 \times 10^{10}$ turns)</td>
</tr>
<tr>
<td>Blow-up/sec of r.m.s. emittance $d\epsilon_{rms}/dt$</td>
<td>$1.3 \times 10^{-9} \pi$ rad m s$^{-1}$</td>
</tr>
<tr>
<td>Tolerable blow-up/sec (cooling strength)</td>
<td>$&lt;5 \times 10^{-9} \pi$ rad m s$^{-1}$</td>
</tr>
</tbody>
</table>

5 EMITTANCE INCREASE DUE TO BENDING AND TUNE FLUCTUATIONS

5.1 Crossing of betatron resonances

Dangerous low-order resonances are usually avoided by choosing an appropriate working point ($Q_x$, $Q_y$). However, high-order resonances may be traversed due to small, unavoidable tune fluctuations or programmed tune changes. A handy relation for the blow-up by a rapid traversal of a resonance $lQ = \text{integer}$ is given by A. Wrulich [16]. The amplitude increase (for small $\Delta a/a \ll 1/l$) may be written as:

$$\frac{(\Delta a/a)_{rms}}{l \sqrt{\Delta Q_t}} = \frac{\pi \Delta e}{I \sqrt{\Delta Q_t}}.$$  \hspace{1cm} (21)

Here $l$ is the order, $\Delta e$ the width of the resonance, $\Delta Q_t$ the tune change per turn. The emittance growth after filamentation ($\Delta e/\epsilon = 1/2 \Delta a^2/a^2 = \Delta a/a$) is given by the same expression. Typically $\Delta e$ is of the order of $10^4$ to $10^5$ and for fast crossing $\Delta Q_t$ is of the same order (e.g. $dQ/dt = 0.1$ ms, $f_{rev} = 1$ MHz $\rightarrow \Delta Q_t = 10^3$). Hence only a few transitions can be tolerated even of high-order resonances. For repeated random crossings the amplitude growth, Eq. (21), is multiplied by the square root of the number of crossings.

For slow tune variation, particles can be trapped in resonance ‘bands’, which move them outwards, eventually even to the aperture limit. This can happen through momentum diffusion (i.e. due to scattering on the residual gas), leading to a tune diffusion via the chromaticity $\xi$: $\Delta Q / Q = \xi \Delta p / p$, or due to other sorts of ‘tune ripple’ (caused, for example, by power supply fluctuations treated in Section 5.2). A careful choice of the working point and a very high stability of the tune are therefore essential, especially in storage rings where a long beam life is required.

5.2 Power supply ripple

Ripple in the dipole and quadrupole fields can be treated by solving

$$x'' + K(s)x = d(s) + g(s)x$$ \hspace{1cm} (22)
where \(d(s)\) and \(g(s)\) are small random functions \([17]\) representing bending and focusing ripples. In the context of accelerators, such differential equations are treated in Refs. \([18]\) and \([19]\). Instead, we can use the previous results, Eqs. (11) and (17), for steering and focusing errors assuming that the particles receive many small random kicks.

We shall consider only dipole errors here, the effect of ripple in the quadrupole supplies can be treated along similar lines. Let \(\theta(n)\) be the random deflection error per turn with the variance \[\langle \theta(n)^2 \rangle = \theta_{\text{rms}}^2.\] Each kick introduces an increase in the \(k\sigma\) -emittance of \(\Delta \varepsilon_{\text{ker}} = \frac{1}{2} \beta_s (k\theta_{\text{rms}})^2 f_{\text{rev}} t.\) (23)

To get an idea of the orders of magnitude involved, assume that storage for one day is required. For a storage ring like the ISR this means \(n = f_{\text{rev}} t \approx 2.5 \times 10^{10}\) turns. Take a typical beta function: \(\beta_s = 20\) m, and assume one magnet giving a nominal angle of \(\theta_\phi = 10\) mrad with field jitter \((B/B)_{\text{rms}} = 10^{-6} \to \theta_{\text{rms}}^2 = (10^5\text{ mrad})^2.\) Then Eq. (21) yields \(\Delta \varepsilon_{\sigma} = 25\pi \text{ mm mrad.}\) This is a sizeable fraction of the acceptance, which is typically 100 \(\pi \text{ mm mrad.}\) These considerations suggest a necessary field stability \(\Delta B/B\) of \(1/10^6.\) Does this require a similar current stability \((\Delta I/I)\) of the power supply?

In fact, the frequency content of the noise is very important \([18,19]\). It can be shown that the beam is most sensitive to noise components that coincide in their frequency \(f \pm m Q\) with the ‘betatron sidebands’ of the particles, i.e., if \(f_k = (m \pm Q) f_{\text{rev}}\) (with \(m = 0, 1, 2, \ldots\)). The magnet and its connections act like a filter with a transfer function \(H(\omega).\) In the simplest case \(H(\omega) = (i\omega L)^{-1}\) is given by the large inductance of the magnet. Then the beam response to ripple in the power supply current is reduced by a filtering factor \(\tilde{F}\)

\[
\tilde{F} = \sum_{m=-\infty}^{\infty} H^2(m\omega_{\text{rev}} + Q\omega_{\text{rev}}) = \sum_{m=-\infty}^{\infty} \frac{1}{k(m\omega_{\text{rev}} + Q\omega_{\text{rev}})^2}.
\] (24)

The lowest sideband \((m \approx -Q)\) which contributes the dominating effect to the sum has already a frequency of \(0.1 \times f_{\text{rev}}\) to \(0.3 \times f_{\text{rev}}\), typically in the range of 100 kHz for small and of 1 kHz for very large machines. Then the effective noise, compared to the supply ripple at 50 Hz, is reduced by 3 to 6 orders of magnitude. These qualitative considerations may show the way to evaluate the sensitivity to power supply ripple.

6 INTRA-BEAM SCATTERING

Small-angle (multiple) Coulomb scattering between particles of the same beam (‘intra-beam scattering’, Refs. \([20]\)–\([25]\)) is important for dense beams. In the collisions, energy transfer back and forth between all three planes can occur. In a strong-focusing accelerator or storage ring this leads not only to a re-distribution, but unavoidably to a growth of the six-dimensional phase-space volume. In fact in modern storage rings and colliders, intra-beam scattering is one of the dominant effects limiting the luminosity lifetime.
There are many papers and excellent Accelerator School articles about intra-beam scattering. I shall therefore only sketch the way of derivation and then point to some particularities. With reference to Fig. 13, the calculations can proceed in the following steps (due to A. Piwinski, 1974):

– Transform the momenta of the two colliding particles into their centre-of-mass system. Calculate the change of the momenta using the Rutherford cross-section.

– Transform the changed momenta back into the laboratory system.

– Calculate the change of the emittances due to the change of momenta at the given location of the collision.

– Take the average over all possible scattering angles (impact parameters from the size of the nucleus to the beam radius).

– Assume a ‘Gaussian beam’ (in all three planes). Take the average over momenta and transverse positions of the particles at the given location on the ring circumference.

– Finally, calculate the average around the circumference (taking into account the lattice function of the ring) to determine the change per turn.

There is (to my knowledge) no simple analytical relation to estimate the effect over a large range of parameters. However, various useful approximations for special cases, e.g. for high or low energy, have been published. The situation is well summarized in the Handbook article by Piwinski [22]. For more exact calculations, computer codes have been developed and successfully applied to various existing machines. Usually one determines ‘form factors’ $F$ and expresses the individual growth $1/\tau_x$, $1/\tau_y$, $1/\tau_p$, by

$$1/\tau_{x,y,p} = 1/\tau_0 * F_{x,y,p} \cdot$$  

(25)
The ‘growth constant’ $1/\tau_0$ (denoted by $A$ in some of the publications) is given by

$$1/\tau_0 = \frac{N_b \left(E_0 r_0^2\right) \left(q^2/A\right)}{(64\pi) \gamma \varepsilon_x \varepsilon_y \varepsilon_i^2} \propto \frac{N_b \left(q^2/A\right)}{\gamma \varepsilon_x \varepsilon_y \varepsilon_i^2}. \quad (26)$$

Here the variables have the following significance: $N_b$, number of particles per bunch; $E_0$, nucleon rest mass (931 MeV); $r_0$, classical proton radius; $q$, $A$ charge number and mass number of the particle; $\varepsilon_{x,y}^* = \beta \gamma \sigma_{x,y}^2/\beta_{x,y}$, $\varepsilon_i^* = \beta \gamma \sigma_i, \sigma_{xp}(E_0/c)$ normalized $\sigma$ emittances of the bunch; $\beta, \gamma$ relativistic factors; $\sigma_i$ and $\sigma_{xp}$ r.m.s. bunch length and r.m.s. $\Delta p/p$.

The scaling of the true growth rate is more complicated than one would deduce from Eq. (26) because, in general, the form factors in Eq. (25) depend also on the energy and on the emittances. In fact G. Parzen gives a relation, valid at very high energy, which we write as

$$1/\tau_x = \frac{1.25 N_b \left(E_0 r_0^2\right) \left(q^2/A\right)}{\varepsilon_x \varepsilon_y \varepsilon_i^2} \frac{\sigma_{xp} 1}{\sigma_{at} Q_x} \propto \frac{N_b \left(q^2/A\right)^2}{\varepsilon_x \varepsilon_y \varepsilon_i^2} \quad (27)$$

$$1/\tau_p = 1/\tau_x \left(\frac{\sigma_x}{\sigma_{xp}}\right)^2.$$

In Eq. (27) $\sigma_x = \sqrt{\varepsilon_x \beta_x}, \sigma_{xp} = D\sigma_x/p$ and $\sigma_{at} = \sqrt{\sigma_x^2 + \sigma_{xp}^2}$ are the average horizontal beam width due to betatron oscillation, due to momentum spread and the total width, respectively, and we have assumed that for the averages $\beta_{x,y,av} = R/Q_{x,y}$ and $D_{av} = R/Q_x^2$ holds, and that $\varepsilon_x/\beta_{x,av} = \varepsilon_y/\beta_{y,av}$.

Equation (27) exhibits independence of $\gamma$ rather than the $1/\gamma$ dependence in Eq. (26) and suggest form factors [Eq. (25)] of the order of $\gamma/Q_x$.

One notes from Eqs. (26) and (27)

- the strong dependence on ion charge ($q^2/A^2$),

- the linear dependence on the normalized phase-space density [$N_b/(\varepsilon_x^* \varepsilon_y^* \varepsilon_i^*)$],

- the weak dependence on energy ($\gamma^0$ or at most $\gamma^{-1}$).

As an example we compare, in Table 7, the situation for protons and lead ions in the LHC. One notes that, although the bunch intensity is lower by more than 3 orders of magnitude, the growth rate is 6 to 10 times faster for lead ions compared to protons. One also notes that Parzen’s formula [Eq. (27)] gives the correct order of magnitude. The growth constant (26) gives much longer times, suggesting that at least in this case the form factors in Eq. (25) are much different from unity (in fact much smaller than 1).
Table 7: LHC parameters for protons and ions at collision energy and resulting IBS growth times from Parzen’s formula [Eq. (27) and values (from computer results) quoted in the LHC parameter list [1].

<table>
<thead>
<tr>
<th></th>
<th>Protons</th>
<th>Lead ions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy (E) (TeV/A)</td>
<td>7</td>
<td>2.76</td>
</tr>
<tr>
<td>(\gamma)</td>
<td>7461</td>
<td>2963</td>
</tr>
<tr>
<td>Particle charge state (q)</td>
<td>1</td>
<td>82</td>
</tr>
<tr>
<td>Particle mass number (A)</td>
<td>1</td>
<td>208</td>
</tr>
<tr>
<td>Particles/bunch (N_b)</td>
<td>(1.15 \times 10^{11})</td>
<td>(7 \times 10^7)</td>
</tr>
<tr>
<td>Transverse 1(\sigma) emittances (\beta \gamma \varepsilon_h = \beta \gamma \varepsilon_v (\pi \text{ mm mrad}))</td>
<td>3.75</td>
<td>1.5</td>
</tr>
<tr>
<td>r.m.s. bunch length (l) (m)</td>
<td>0.075</td>
<td>0.079</td>
</tr>
<tr>
<td>r.m.s. momentum spread (\sigma p/p)</td>
<td>(1.1 \times 10^{-4})</td>
<td>(1.1 \times 10^{-4})</td>
</tr>
<tr>
<td>Machine radius (R) (km)</td>
<td>4.24</td>
<td>4.24</td>
</tr>
<tr>
<td>Betatron tunes (Q_x)</td>
<td>(\approx 64)</td>
<td>(\approx 64)</td>
</tr>
<tr>
<td>(Q_y)</td>
<td>(\approx 59)</td>
<td>(\approx 59)</td>
</tr>
<tr>
<td>Average (\beta_x = \beta_y)</td>
<td>(R/Q) (m)</td>
<td>70</td>
</tr>
<tr>
<td>Average dispersion (D)</td>
<td>(R/Q^2) (m)</td>
<td>1.2</td>
</tr>
<tr>
<td>Horizontal growth time Eq. (27) (\tau_x) (h)</td>
<td>254</td>
<td>26</td>
</tr>
<tr>
<td>Longitudinal growth time Eq. (27) (\tau_p) (h)</td>
<td>127</td>
<td>12</td>
</tr>
<tr>
<td>Horizontal growth time (Ref. [1]) (\tau_x) (h)</td>
<td>80</td>
<td>13</td>
</tr>
<tr>
<td>Longitudinal growth time (Ref. [1]) (\tau_p) (h)</td>
<td>61</td>
<td>7.7</td>
</tr>
</tbody>
</table>

To conclude this section we mention further particularities of intra-beam scattering concerning the sum of the three emittances.

- For constant lattice functions (weak focusing) and below transition energy, the sum of the three emittances is constant (the beam behaves like a ‘gas in a box’).
- Above transition, the sum of the emittances always grows (due to the negative mass effect, i.e., particles ‘being pushed go around slower’).
- In any strong-focusing lattice the sum of the emittances always grows (also below transition because of the ‘friction’ due to the derivatives of the lattice functions).
- The increase of the 6-dimensional phase-space volume can be explained by transfer of energy from the common longitudinal motion into transverse energy spread.
- Although the sum grows, there can be strong transfer of emittance and theoretically even reduction in one plane at the expense of fast growth in another plane (in practice the reduction in one plane has, to my knowledge, not been observed).

7 CONCLUSION

Emittance (beam density) preservation is a major concern in the design and the operation of modern accelerators and colliders. There is a great number of ‘single-particle’ and ‘collective’ effects which have to be strictly controlled to avoid excessive ‘beam heating’. We have discussed only the most common of them here.
Beam cooling (not treated in this note) can, to some extent, be used to fight emittance growth and even lead to very small equilibrium emittances. They result from the balance between the cooling and (many of) the heating mechanisms discussed above.

**Appendix A: Increase of r.m.s. emittance due to a steering error at injection**

We use the following notation (Fig. A.1): index m, machine phase-space (origin at 0); index i, injected beam phase-space [origin displaced by $\Delta r = (\Delta r, \Delta \Psi)$].

![Phase space for a steering error](image)

**Fig. A.1:** Phase space for a steering error

A test particle has:

$$\vec{r}_m = \vec{r}_i + \Delta \vec{r},$$

i.e.

$$x_m = r_i \cos \Psi_i + \Delta r \cos \Delta \Psi$$

$$p_m = r_i \sin \Psi_i + \Delta r \sin \Delta \Psi$$

(A.1)

The injection error ($\Delta r$ and $\Delta \Psi$) is fixed but the phases $\Psi_i$ of the injected particles are uniformly distributed from $0 \to 2\pi$.

Averaging yields

$$\langle r_m^2 \rangle = \langle x_i^2 + p_i^2 \rangle =$$

$$= \langle r_i^2 \rangle + \Delta r^2 .$$

After filamentation, the phases are again uniformly distributed and $x = r_m \cos \Psi$, i.e.:

$$\langle x^2 \rangle = \frac{1}{2} \langle r_m^2 \rangle = \frac{1}{2} \langle r_i^2 \rangle + \frac{1}{2} \Delta r^2$$

$$\sigma^2 = \sigma_i^2 + \frac{1}{2} \Delta r^2$$

$$\epsilon_{r_{ms}} = \epsilon_{rms,i} + \frac{1}{2} \Delta r^2 / \beta_x .$$
This result is derived and discussed in more detail in Refs. [2], [3], [26] and [27].

Appendix B: Increase of r.m.s. emittance due to mismatch of dispersion

As before we assume that the machine has zero \( D \), \( D' \) and the line has finite \( D \), \( D' \) at the injection point; in the general case one has just to substitute the difference between line and machine \( D \rightarrow \Delta D \) and \( D' \rightarrow \Delta D' \).

For any particle:
\[
\begin{align*}
x &= r_i \cos(\Psi_i) + D\Delta p / p, \\
p_x &= r_i \sin(\Psi_i) + (\beta_i D' + \alpha_i D)\Delta p / p.
\end{align*}
\]

For each value of \( \Delta p / p \) the phases \( \Psi_i \) are uniformly distributed around an off-momentum orbit with its centre displaced by \( \Delta \vec{r} = \left[ D\Delta p / p, (\beta_i D' + \alpha_i D) \right] p / p \). After filamentation the phases are uniformly distributed around the origin \( \Delta \vec{r} = 0 \). Then [by virtue of Eqs. (3) above]
\[
\sigma_x^2 \langle x^2 \rangle = \frac{1}{2} \langle A_i^2 \rangle = \langle p_x^2 \rangle
\]
and hence
\[
\begin{align*}
2\sigma_x^2 &= \langle x^2 + p_x^2 \rangle = \langle r_i^2 + (D^2 + (\beta_i D' + \alpha_i D)^2 \Delta p / p)^2 \rangle, \\
\sigma_x^2 &= \sigma_0^2 + \frac{1}{2} \langle (D^2 + (\beta_i D' + \alpha_i D)^2 \rangle (\sigma_p / p)^2 \\
\varepsilon_{\text{rms}} &\rightarrow \varepsilon_{\text{rms},i} \left[ 1 + \frac{1}{2} \langle (D^2 + (\beta_i D' + \alpha_i D)^2 \rangle (\sigma_p / p)^2 / \sigma_0^2 \right].
\end{align*}
\]

Appendix C: Increase of emittance due to \( \beta \)-function mismatch at transfer

Consider the incoming beam (Fig. 9, above) in the normalized (circular) phase-space coordinates of a machine. To calculate the increase of the geometrical emittance, observe that after filamentation the incoming phase-space ellipse (containing a given fraction of the beam) fills the circumscribed circles of the machine phase-space. The increase in area and thus in geometrical emittance is \( b/a \) (major/minor ellipse axis):
\[
\varepsilon_{\text{gs}} \rightarrow \varepsilon_{\text{gs}} (b/a).
\]

To calculate the increase of the r.m.s. emittance as a function of \( b/a \) we can assume that the incoming beam ellipses are in the upright position (this amounts to rotating the normalized machine phase-space by a constant angle which does not affect the resulting blow-up). Then the coordinates of an incoming test particle can be written as
\[
\begin{align*}
x_i &= A_i \sqrt{b/a} \sin \Psi_i, \\
p_i &= A_i \sqrt{a/b} \cos \Psi_i.
\end{align*}
\]
Here \( A_i^2 \) is the single-particle emittance and the beam average \( \langle A_i^2 \rangle = 2 \sigma_i^2 / \beta_i \) is [by virtue of Eqs. (2) and (7) above] twice the r.m.s. emittance in the line. The coordinates \( x_i, p_i \) define the initial conditions for the circular phase-space motion with an amplitude \( r_i^2 = (x_i^2 + p_i^2) \) in the ring. Averaging over all particles assuming that the phases are random we find \( \langle r_i^2 \rangle / \beta_i = \langle A_i^2 \rangle (b/a + a/b) / 2 \). This is (twice) the r.m.s. emittance after filamentation in the ring. Thus the blow-up of the k r.m.s. emittance is
To express the aspect ratio \( b/a \) by the mismatch of the Twiss parameters we refer to Refs. [2] and [3] where it is shown that, in terms of the mismatch function \( F \) [Eq. (15) above],
\[
\left( \frac{b}{a} \right) \equiv \left( F + \sqrt{F^2 - 1} \right), \quad \left( \frac{a}{b} \right) = \left( F - \sqrt{F^2 - 1} \right).
\]
From Eqs. (C.1) and (C.2) one then finds Eqs. (16) and (17) above for the blow-up of the geometrical and the r.m.s. emittance, respectively.

Appendix D: Increase of r.m.s. emittance due to scattering on a foil

The normalized circular phase-space trajectories of a matched beam at the entrance of the foil are converted into elliptical trajectories at the exit (Fig. 12). In fact for any particle scattered by an angle \( \theta_i \)
\[
x_i \rightarrow x_i = r_{i0} \sin(\Psi_i) \\
p_{xi} \rightarrow p_{xi} + \Delta p = r_{i0} \cos(\Psi_i) + \beta_0 \theta_i \quad \text{ (remember: } p_x = \alpha x + \beta x' \).
\]
These are the initial conditions for the new betatron oscillation after the scattering event. Averaging over the beam, we take the phases \( \Psi_i \) of the original oscillation uniformly distributed and uncorrelated with the scattering angle. After filamentation in the line and/or the subsequent ring, the new betatron phases are also uniformly distributed and one obtains
\[
\langle \sigma_x^2 \rangle = \langle x_i^2 + p_{xi}^2 \rangle = \langle r_{i0}^2 \rangle + \langle \beta_0^2 \theta_i^2 \rangle \\
\rightarrow 2\sigma_x^2 = 2\sigma_{x0}^2 + \beta_0^2 \sigma_{ms}^2 \quad \text{ or } \quad \epsilon_{k\sigma} = (k\sigma_x)^2 / \beta_0 = \epsilon_{k\sigma0} \frac{1}{2} (k\theta_{ms})^2 \beta_0
\]
Thus the blow-up of \( k\sigma \) emittance is
\[
\Delta \epsilon_{k\sigma} = \frac{1}{2} (k\theta_{ms})^2 \beta_0.
\]
We have tacitly assumed above that the beta function is the same at the entrance and exit so that the beam remains matched in the absence of the scatterer. By re-adjusting the optics in the downstream part of the line, the blow-up can in fact be somewhat reduced. The idea is to provide a smaller \( \beta \) function (\( \beta_i < \beta_0 \)) at exit (1) than at the entrance (0) of the scatterer. For simplicity, we perform the calculation only for \( \alpha = 0 \). We chose \( \beta_i \) such that it matches the ellipse in Fig. 12 (right) after scattering instead of the circumscribed circle. This is called re-matching at the scatterer. Then, for any particle
\[
x = A_i \sqrt{\beta_0} \cos(\Psi_i) = A_i \sqrt{\beta_i} \cos(\Psi_i) \\
x' = -A_i \sqrt{1/\beta_0} \sin(\Psi_i) + \theta = -A_i \sqrt{1/\beta_i} \sin(\Psi_i)
\]
Taking the squares and averaging with the assumption that the beam remains matched before and after the foil (so that all the \( \cos^2 \) and \( \sin^2 \) terms average to 1/2), one obtains
\[ \langle x^2 \rangle = \frac{\langle A_0^2 \rangle \beta_0}{2} = \frac{\langle A_1^2 \rangle \beta_1}{2} \]

\[ \langle x^2 \rangle = \frac{\langle A_0^2 \rangle}{2 \beta_0} = \frac{\langle A_1^2 \rangle}{2 \beta_1} . \]

The solution of this system is

\[ \langle A_1^2 \rangle^2 - \langle A_0^2 \rangle^2 = 2 \beta_0 \theta_{rns} \langle A_0^2 \rangle \]

\[ \frac{\beta_1}{\beta_0} = \frac{\langle A_1^2 \rangle}{\langle A_0^2 \rangle} . \]

Remember that the 1σ emittance is \( \epsilon_{rns} = \langle A^2 \rangle / 2 \). Therefore one obtains

\[ \epsilon_1 = \epsilon_0 \sqrt{1 + \frac{\beta_0 \theta_{rns}^2}{\epsilon_0}} = \epsilon_0 \left[ 1 + \frac{1}{2} \frac{\beta_0 \theta_{rns}^2}{\epsilon_0} - \frac{1}{8} \left( \frac{\beta_0 \theta_{rns}^2}{\epsilon_0} \right)^2 \right] \]

\[ \frac{\beta_1}{\beta_0} = \frac{1}{\sqrt{1 + \frac{\beta_0 \theta_{rns}^2}{\epsilon_0}}} . \]

Examples of the ratio \( \epsilon_1 / \epsilon_0 \) without and with re-matching are given in Table D.1. For small scattering angle, \( \epsilon_1 / \epsilon_0 \) with re-matching is by \( 1/8(\beta_0 \theta_{rns}^2 / \epsilon_0)^2 \) smaller than in the case of constant \( (\beta_1 = \beta_0) \).

One concludes that re-matching is only worth while for large values of \( (\beta_0 \theta_{rns}^2 / \epsilon_0) \).

<table>
<thead>
<tr>
<th>Scattering angle ( (\beta_0 \theta_{rns}^2 / \epsilon_0) )</th>
<th>Increase of r.m.s. emittance without re-matching ( \epsilon_1 / \epsilon_0 )</th>
<th>Increase of r.m.s. emittance with re-matching ( \epsilon_1 / \epsilon_0 )</th>
<th>Change of β-function for re-matching ( \beta_1 / \beta_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.5</td>
<td>1.41</td>
<td>0.71</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1.73</td>
<td>0.58</td>
</tr>
<tr>
<td>3</td>
<td>2.5</td>
<td>2</td>
<td>0.50</td>
</tr>
</tbody>
</table>

References


[27] H.G. Hereward, How good is the r.m.s. as a measure of beam size, internal report CERN MPS/DL 69-15.
Abstract
Landau damping is the suppression of an instability by a spread of frequencies in the beam. It is treated here from an experimental point of view. To introduce the concept we consider a set of oscillators having a spread in resonant frequencies $\omega_r$, and calculate the response of their there center-of-mass to an external driving force. A pulse excitation gives each oscillator the same initial velocity but, due to their different frequencies, the center-of-mass motion will decay with time. A harmonic excitation with a frequency $\omega$ being inside the distribution in $\omega_r$ results in oscillators responding with different phases and only a few of them having $\omega_r \approx \omega$ will grow to large amplitudes and absorb energy. The oscillator response to a pulse excitation, called Green function, and the one to a harmonic excitation, called transfer function, serve as a basis to calculate Landau damping which suppresses an instability at infinitesimal level before any large amplitudes are reached. This is illustrated by a negative feed-back system acting on the center-of-mass of the oscillators. In the absence of a frequency spread this leads in an exponential growth, but a sufficiently large spread will provide stability. Landau damping is then applied to the transverse motion of a coasting (un-bunched) beam. The dependence of revolution frequency and betatron tune on momentum establishes a spread of the frequencies contained in the upper and lower betatron side-band. A transverse impedance present in the ring gets excited by the beam and acts back on it, causing an instability. For a given frequency distribution a maximum value of this impedance, being still conform with stability, is calculated and plotted in a stability diagram. Finally, longitudinal instabilities in a coasting beam are investigated. They can be stabilized by a spread of revolution frequencies caused by an energy distribution of the particles in the beam.

1 INTRODUCTION
1.1 Mechanism of Landau damping
A single oscillator with resonant frequency $\omega_r$ reacts to a pulse excitation with a free oscillation lasting for a long time or being slightly damped. A lasting harmonic excitation with frequency $\omega$ results in a forced oscillation at the same frequency $\omega$ but a phase which depends on the difference $\omega - \omega_r$. For $\omega = \omega_r$ the oscillation amplitude grows linear with time.

We consider a set of oscillators having different resonant frequencies $\omega_{rj}$ with distribution $f(\omega_{rj})$. A pulse excitation results in an oscillation of each oscillator with the same initial velocity $\dot{x}(0)$ followed by a free oscillation with individual frequencies $\omega_{rj}$. For instabilities or beam observation the center-of-mass motion of the particles is relevant. Due to their different $\omega_{rj}$ the freely oscillating particles change their phase with respect to each other and the center-of-mass motion is slowly reduced. In case of a harmonic excitation the phases of the individual particle oscillations are different and the center-of-mass motion has a smaller amplitude than the individual particles.

This represents a kind of damping where the coherent center-of-mass motion is reduced compared to the incoherent motion of the particles. This damping is usually
not exponential and differs in many respect from other damping mechanism. It depends on the form of the resonant frequency distribution \( f(\omega_{rj}) \) but mainly on its width, i.e. the frequency spread. For a frequency distribution given by an external parameter the damping is proportional to \( f(\omega) \) itself. If this distribution is determined by the amplitude dependence of \( \omega_{rj} \) it is affected by the excitation giving a damping proportional to the derivative \( df(\omega_r)/d\omega_r \).

1.2 Treatment of Landau damping

Landau damping can be understood from different points of view and is often presented in different ways, [1, 2, 3]. We treat it here in a manner close to beam observation and experiment.

We calculate the center-of-mass response of a beam with resonant frequency distribution \( f(\omega_r) \) to a pulse or harmonic excitation and compare it with experiments.

This center-of-mass motion induces fields in the beam surroundings which act back on the beam and can enhance the excitation. The electrical properties of the components surrounding the beam, relevant for this effect, can be expressed by an impedance. The fields induced in this impedance can be sufficiently large to keep this process going leading to a self excitation. This leads to an instability having a threshold determined by beam response and impedance. Below this threshold the frequency spread eliminates any coherent motion at infinitesimal small amplitudes before it can grow - we have stability. Above, the voltage induced in the resistive part of the impedance leads to an increase of initial coherent motion and we have an instability.

The amount of Landau damping depends on the frequency distribution \( f(\omega_r) \) or its derivative at the frequency \( \omega \) at which the instability occurs. It can happen that the coherent (center-of-mass) motion has a different frequency than the incoherent individual particle frequencies. In this case Landau damping might become ineffective and we can get an instability for a very small resistive impedance.

We will calculate the beam response and Landau damping for transverse and longitudinal instabilities of a coasting (un-bunched) beam. From this we can determine the maximum transverse or longitudinal impedance, which still does not create an instability, and represent it in the so-called stability diagram.

2 RESPONSE OF AN OSCILLATOR SET TO AN EXCITATION

2.1 Response to a pulse excitation – Green function

A set of oscillators \( j \) of different resonant frequencies \( \omega_{rj} \) receive at \( t = 0 \) a kick giving each the same velocity \( \dot{x}_j(0^+) = \dot{x}_0 \). Each performs a harmonic oscillation of the
as a series of infinitesimal kicks with harmonic modulation
\[ \dot{x}_0 = \hat{\dot{x}}_0 \cos(\omega t_0)dt_0 = Gdt_0 = \hat{G}\cos(\omega t_0)dt_0 \]

The observer measures only the center-of-mass motion
\[ \langle x(t) \rangle = \dot{x}_0 \int f(\omega_r)\cos(\omega_r t) d\omega_r \approx \dot{x}_0 \int f(\omega_r) \sin(\omega_r t) d\omega_r \]

Expressing the velocity response by the difference frequency \( \Delta \omega_r \)
\[ g(t) = \frac{\langle \dot{x}(t) \rangle}{\dot{x}_0} = \cos(\omega_{r0} t) \int f(\Delta \omega_r) \cos(\Delta \omega_r t) d\omega_r \]
\[ -\sin(\omega_{r0} t) \int f(\Delta \omega_r) \sin(\Delta \omega_r t) d\omega_r \]

The integrals represent the inverse sine and cosine Fourier transforms
\[ g(t) = \mathcal{F}^{-1}_{\cos}(f(\Delta \omega_r)) \cos(\omega_{r0} t) - \mathcal{F}^{-1}_{\sin}(f(\Delta \omega_r)) \sin(\omega_{r0} t) \]
\[ = \mathcal{F}^{-1}(f(\Delta \omega_r)) e^{i\omega_{r0} t} \]

The center-of-mass velocity response \( g(t) \) of an oscillator set with resonance frequency distribution \( f(\Delta \omega_r) \) to a pulse excitation (Green function) is the inverse Fourier transform of this distribution times an oscillation with the central frequency \( \omega_{r0} \).

**2.2 Response to a harmonic excitation – transfer function**

We assume oscillators having a very small damping and excite them at the time \( t = t_0 \) with a short pulse. The resulting velocity response of a single particle is given by
\[ \dot{x}_j(t) = e^{-a(t-t_0)} \cos(\omega_{rj}(t-t_0)) \]
and shown in Fig. 1. A harmonic excitation with frequency \( \omega \) starting at \( t_0 = -\infty \), lasting until observation time \( t \) is treated as a series of infinitesimal kicks with harmonic modulation
\[ d\dot{x}_0 = \frac{d\hat{\dot{x}}_0}{dt_0} \cos(\omega t_0)dt_0 = Gdt_0 = \hat{G}\cos(\omega t_0)dt_0 \]
where $G$ is an acceleration. The velocity obtained at the time $t$

$$
\dot{x}(t) = \hat{G} \int_{-\infty}^{t} \cos(\omega t_0) e^{-a(t-t_0)} \cos(\omega_r(t - t_0)) dt_0
$$

substituting $T = t - t_0$ and developing $\cos(\omega t_0) = \cos(\omega(t - T))$

$$
\dot{x}(t) = \hat{G} \int_0^{\infty} (\cos(\omega t) \cos(\omega T) + \sin(\omega t) \sin(\omega T)) e^{-aT} \cos(\omega_r T) dT
$$

$$
= \hat{G} \left( I_1 \cos(\omega t) + I_2 \sin(\omega t) \right)
$$

with the two integrals

$$
I_1 = \int_0^{\infty} e^{-aT} \cos(\omega T) \cos(\omega_r T) dT = \left( \frac{a/2}{a^2 + (\omega - \omega_r)^2} + \frac{a/2}{a^2 + (\omega + \omega_r)^2} \right)
$$

$$
I_2 = \int_0^{\infty} e^{-aT} \sin(\omega T) \cos(\omega_r T) dT = \left( \frac{(\omega - \omega_r)/2}{a^2 + (\omega - \omega_r)^2} + \frac{(\omega + \omega_r)/2}{a^2 + (\omega + \omega_r)^2} \right)
$$

We assume $\omega \approx \omega_r$ and weak damping $a \ll \omega$.

$$
I_1 \approx \frac{a}{2(a^2 + (\omega - \omega_r)^2)} \, I_2 \approx \frac{\omega - \omega_r}{2(a^2 + (\omega - \omega_r)^2)}
$$

with $\int_{-\infty}^{\infty} I_1 d\omega_r = \frac{1}{2} \arctan \left( \frac{\omega_r - \omega}{a} \right)_{-\infty}^{\infty} = \frac{\pi}{2}$.

For $a \to 0$ the first expression vanishes except for the point $\omega - \omega_r \to 0$ where it becomes infinite. We approximate it with the Dirac $\delta$-function. The second integral can be simplified. For vanishing $a$ it jumps from $-\infty$ to $+\infty$ when $\omega$ goes through $\omega_r$

$$
I_1 = \frac{1}{2} \pi \delta(\omega - \omega_r) \text{ for } a \to 0 \, , \, I_2 = \frac{1}{2(\omega - \omega_r)} \text{ for } a \to 0 \, , \, \omega - \omega_r \neq 0
$$

We get for the single particle response to a harmonic excitation

$$
\ddot{x} = \frac{\hat{G}}{2} \left( \pi \delta(\omega - \omega_r) \cos(\omega t) + \frac{1}{(\omega - \omega_r)} \sin(\omega t) \right)
$$

and for the set of oscillators with resonant frequency distribution $f(\Delta \omega_r)$

$$
\langle \dot{x} \rangle = \int_{-\infty}^{\infty} \dot{x} f(\Delta \omega_r) d\omega_r
$$

$$
= \frac{\hat{G}}{2} \left( \pi f(\omega) \cos(\omega t) + PV \int_{-\infty}^{\infty} \frac{f(\Delta \omega_r) d\omega_r}{\omega - \omega_r} \sin(\omega t) \right)
$$

This response to a harmonic excitation is often called transfer function.

### 2.3 Short derivation of the transfer function using complex notation

For oscillators with symmetric resonant frequency distribution, no damping and complex notation with positive and negative frequencies

$$
f(\omega_r) = f(-\omega_r) \, , \, \ddot{x} + \omega_r^2 x = \frac{\hat{G}}{2} e^{j\omega t} \, , \, (-\omega_r^2 + \omega_r^2)x = \frac{\hat{G}}{2} e^{j\omega t} \quad (2)
$$
we have a displacement response
\[ x = -\frac{\hat{G} e^{j\omega t}}{2(\omega^2 - \omega^2_r)} = -\frac{\hat{G} e^{j\omega t}}{2(\omega - \omega_r)(\omega + \omega_r)} = -\frac{\hat{G} e^{j\omega t}}{4\omega} \left( \frac{1}{\omega - \omega_r} + \frac{1}{\omega + \omega_r} \right) \]
\[ \langle x \rangle = \frac{-\hat{G} e^{j\omega t}}{4\omega} \int_{-\infty}^{\infty} \left( \frac{f(\Delta \omega_r)}{\omega - \omega_r} + \frac{f(\Delta \omega_r)}{\omega + \omega_r} \right) d\omega_r \]
\[ = \frac{-\hat{G} e^{j\omega t}}{2\omega} \int_{-\infty}^{\infty} \frac{f(\Delta \omega_r)}{\omega - \omega_r} d\omega_r \]  
(3)

If the exciting frequency \( \omega \) is within the distribution \( f(\omega_r) \) the integral contains a pole leading to a residue and principle value integral
\[ \langle x \rangle = \frac{-\hat{G} e^{j\omega t}}{4\omega} \left( \pm j\pi f(\omega) + \text{PV} \int \frac{f(\Delta \omega_r)}{\omega - \omega_r} d\omega_r \right) \]
with
\[ \text{PV} \int \frac{f(\Delta \omega_r)}{\omega - \omega_r} d\omega_r = \lim_{\varepsilon \to 0} \left[ \int_{-\infty}^{-\varepsilon} \frac{f(\Delta \omega_r)}{\omega - \omega_r} d\omega_r + \int_{\varepsilon}^{\infty} \frac{f(\Delta \omega_r)}{\omega - \omega_r} d\omega_r \right] \]  
(4)

The velocity response with is \( \langle \dot{x} \rangle = j\omega \langle x \rangle \)
\[ \langle \dot{x} \rangle = \frac{\hat{G}}{2} e^{j\omega t} \left( \pm \pi f(\omega) - j\text{PV} \int \frac{f(\Delta \omega_r)}{\omega - \omega_r} d\omega_r \right) \]  
(5)

The real (resistive) part of the velocity response is in phase with the excitation and can absorb energy. The imaginary (reactive) part is out of phase with the excitation and does not absorb energy. There is an ambiguity in the sign of the resistive part because the initial condition is not specified. If there is no oscillation to begin with, the exciter gives energy to the oscillators (+ sign), however, there could already exist a coherent oscillation at an early time giving now energy to the exciter (- sign).

Caution: Sometimes for the oscillation in complex notation \( \exp(-i\omega t) \) is used instead of \( \exp(j\omega t) \). The corresponding results can be converted by exchanging \( j = -i \).

2.4 Time evolution of the response using real notation

We saw that the resistive part of the beam response to a harmonic excitation is due to the residue of the integral. Since it is this term which leads to Landau damping we would like to better understand the underlying physics. For this purpose we now discuss the time evolution of the response using real notation [4]. This leads to derivations which are more lengthy but more transparent. The differential equation describing the excitation of the beam is in analogy with (2)
\[ \frac{d^2y}{dt^2} + \omega_\beta^2 y = \hat{G} \sin(\omega t). \]
This equation has a homogeneous solution
\[ y_h = A \sin(\omega_\beta t) + B \cos(\omega_\beta t) \]
and a particular solution
\[ y_p = \frac{\hat{G}}{\omega_\beta^2 - \omega^2} \sin(\omega t). \]
The general solution is a combination of the two
\[ y = A \sin(\omega_\beta t) + B \cos(\omega_\beta t) + \frac{\hat{G}}{\omega_\beta^2 - \omega^2} \sin(\omega t). \]
To determine the integration constants $A$ and $B$ we have to specify the initial conditions. We take the case where the oscillators are at rest until the time $t = 0$ when the excitation starts

$$y(0) = \dot{y}(0) = 0 \rightarrow A = -\frac{\omega}{\omega_\beta (\omega_\beta^2 - \omega^2)}; \quad B = 0.$$ 

This gives for the general solution

$$y = \frac{\hat{G}}{\omega_\beta^2 - \omega^2} \left( \sin(\omega t) - \frac{\omega}{\omega_\beta} \sin(\omega_\beta t) \right) \quad (6)$$

$$= \frac{\hat{G}}{2\omega} \left( \frac{1}{(\omega_\beta - \omega)} - \frac{1}{\omega_\beta + \omega} \right) \left( \sin(\omega t) - \frac{\omega}{\omega_\beta} \sin(\omega_\beta t) \right).$$

We differentiate with respect to $t$ to get the velocity of the single oscillator response

$$\dot{y} = \frac{\hat{G}}{2} \left( \cos(\omega t) - \cos(\omega_\beta t) \frac{\omega}{\omega_\beta} \sin(\omega_\beta t) \right) - \cos(\omega t) - \cos(\omega_\beta t) \frac{\omega}{\omega_\beta} \sin(\omega_\beta t) \right) \right)$$

and rewrite this equation by substituting $\omega_\beta$ with $\omega_\beta = \omega + (\omega_\beta - \omega)$ in the first term and $\omega_\beta = (\omega_\beta + \omega) - \omega$ in the second term of the parenthesis,

$$\dot{y} = \frac{\hat{G}}{2} \left[ \cos(\omega t) \left( \frac{1 - \cos((\omega_\beta - \omega)t)}{(\omega_\beta - \omega)} - \frac{1 - \cos((\omega_\beta + \omega)t)}{\omega_\beta + \omega} \right) \right.$$ 

$$+ \sin(\omega t) \left( \frac{\sin((\omega_\beta - \omega)t)}{(\omega_\beta - \omega)} + \frac{\sin((\omega_\beta + \omega)t)}{\omega_\beta + \omega} \right) \right]. \quad (7)$$

The above equation gives the velocity response for a single oscillator with resonant frequency $\omega_\beta$. For the first term in the square bracket the velocity and the acceleration (2) are basically out of phase while for the second term they are in phase. This statement is not exact since both terms also contain an oscillatory term with frequency $(\omega_\beta - \omega)$ which has to be discussed further. Before we integrate to obtain the center-of-mass response we discuss these two components of the single oscillator response (8). We concentrate on a region in the vicinity of the exciting positive frequency $\omega$, i.e. $\omega_\beta \approx \omega$ where the first term inside each of the two round brackets is dominant. A corresponding discussion for $\omega_\beta \approx -\omega$ could easily be carried out in addition. In Fig. 3 the envelope of the executed oscillation is shown as a function of the difference $(\omega_\beta - \omega)$ between the resonant and the exciting frequency for different times $t$ after the start of the excitation. As this time increases particles oscillating with opposite phase are close together in frequency leading to some cancelation in the integration to follow. For the resistive (in phase) term the oscillators with resonant frequency close to the exciting frequency gain large amplitudes.

To calculate the center-of-mass response we have to integrate the single particle response (8) over the resonant frequency $\omega_\beta$ weighted with the distribution $f(\omega_\beta)$. Using again the symmetry property of the distribution $f(\omega_\beta) = f(-\omega_\beta)$ we can show that

$$\int_{-\infty}^{\infty} \frac{1 - \cos((\omega_\beta - \omega)t)}{(\omega_\beta - \omega)} d\omega_\beta = -\int_{-\infty}^{\infty} \frac{1 - \cos(\omega_\beta + \omega)t}{\omega_\beta + \omega} d\omega_\beta$$

and

$$\int_{-\infty}^{\infty} \frac{\sin((\omega_\beta - \omega)t)}{(\omega_\beta - \omega)} d\omega_\beta = \int_{-\infty}^{\infty} \frac{\sin(\omega_\beta + \omega)t}{\omega_\beta + \omega} d\omega_\beta$$

276
Figure 3: Time evolution of the beam response to a harmonic excitation which gives for the velocity response

\[ \langle \dot{y} \rangle = \hat{G} \cos(\omega t) \int_{-\infty}^{\infty} f(\omega) \frac{1 - \cos((\omega_\beta - \omega)t)}{\omega_\beta - \omega} d\omega_\beta 
+ \hat{G} \sin(\omega t) \int_{-\infty}^{\infty} f(\omega) \frac{\sin((\omega_\beta - \omega)t)}{\omega_\beta - \omega} d\omega_\beta. \]  

(8)

We start with the out of phase (reactive) part which has an oscillatory term of the form \( \cos((\omega_\beta - \omega)t) \). As the time \( t \) increases this term will have opposite phase for smaller and smaller frequency differences in \( \omega_\beta \). The integration over \( \omega_\beta \) will therefore vanish and we can replace the numerator of the reactive term by unity \((1 - \cos((\omega_\beta - \omega)t) \to 1)\) except for the oscillators with \( \omega_\beta \approx \omega \). This central part \( \omega_\beta \approx \omega \) becomes more and more narrow as the time \( t \) increases and we can replace the integral over the reactive term by the principle value integral \((4)\)

\[
\lim_{t \to \infty} \int f(\omega) \frac{1 - \cos((\omega_\beta - \omega)t)}{\omega_\beta - \omega} d\omega_\beta = \int \frac{f(\omega)}{\omega_\beta - \omega} d\omega_\beta 
= \lim_{\epsilon \to 0} \left( \int_{-\infty}^{\epsilon} f(\omega) d\omega_\beta + \int_{\epsilon}^{\infty} f(\omega) d\omega_\beta \right) = \text{PV} \int \frac{f(\omega)}{\omega_\beta - \omega} d\omega_\beta. 
\]  

(9)

The resistive (in phase) term of (8) contains an oscillatory term under the integral of the form

\[ \frac{\sin((\omega_\beta - \omega)t)}{\omega_\beta - \omega} = t \frac{\sin((\omega_\beta - \omega)t)}{(\omega_\beta - \omega)t} = t \text{sinc}((\omega_\beta - \omega)t). \]  

(10)

For \( \omega_\beta \neq \omega \) and large \( t \) the above expression oscillates with opposite phase for small changes in \( \omega_\beta \). The integration over this frequency will vanish as long as the distribution...
\( f(\omega_\beta) \) is sufficiently smooth. For \( \omega_\beta \approx \omega \) and very large \( t \) the function sinc((\( \omega_\beta - \omega \))t) will be about unity and the above expression grows with increasing time \( t \) without limits. Furthermore the integral

\[
\int_{-\infty}^{\infty} \frac{\sin((\omega_\beta - \omega)t)}{(\omega_\beta - \omega)} d\omega_\beta = \pi
\]

is independent of \( t \). We can therefore replace (10) by the \( \delta \)-function

\[
\lim_{t \to \infty} \sin((\omega_\beta - \omega)t) = \pi \delta(\omega_\beta - \omega).
\]

Collecting the results obtained for the reactive and resistive part of the beam response we get

\[
\lim_{t \to \infty} \langle \dot{y} \rangle = \hat{G} \left[ \sin(\omega t)\pi f(\omega) + \cos(\omega t)\text{PV} \int f(\omega_\beta) \frac{d\omega_\beta}{(\omega_\beta - \omega)} \right].
\]

This is the same result as that already obtained more quickly using complex notation. However, in this subsection we learned basically three things:

- The result (5) is only correct if the excitation has lasted for a long time. How long this time has to be depends on the resolution with which the distribution \( f(\omega_\beta) \) has to be considered. If this distribution does not change significantly over a frequency range of \( \Delta \omega_\beta \) it is sufficient to excite for a time \( t \gg 1/\Delta \omega_\beta \).

- The sign of the residue can be determined from the initial conditions. Usually one excites a set of oscillators being initially at rest. However, it is in principle possible to have a set of particles oscillating initially with a particular distribution in amplitude and phase such that the exciting acceleration takes energy out of the beam. Since in the complex notation the initial conditions were not specified both possibilities are contained in the beam response equation.

- From Fig. 3 it is clear that a few oscillators, having resonant frequencies close to the exciting frequency \( \omega_\beta \approx \omega \), attain large amplitudes. The energy absorbed by the beam from the exciter goes, therefore, into large oscillation amplitudes obtained by a small fraction of the oscillators.

2.5 Relation between the two responses

For the velocity response to a harmonic excitation we obtain from the previous section (5)

\[
r(\omega) = \frac{\langle \dot{y}(\omega) \rangle}{G} = \left[ \pi f(\omega) - i\text{PV} \int \frac{f(\omega_\beta)}{(\omega_\beta - \omega)} d\omega_\beta \right] = r_r(\omega) + ir_i(\omega).
\]

This response of an oscillator set to a harmonic excitation is called the transfer function.

The pulse response, also called Green function, is according to (1)

\[
g(t) = \frac{\langle \dot{y}(t) \rangle}{\dot{y}_0} = \int f(\omega_\beta) \cos(\omega_\beta t) d\omega_\beta
\]

Comparing these two responses one suspects that they are related to each other by the Fourier transform and its inverse

\[
F(\omega) = \mathcal{F} [f(t)] = \int f(t)e^{i\omega t} dt = \int f(t) \cos(\omega t) dt + i \int f(t) \sin(\omega t) dt \quad (11)
\]

\[
f(t) = \mathcal{F}^{-1} [F(\omega)] = \frac{1}{2\pi} \int F(\omega) e^{-i\omega t} d\omega.
\]
The integration goes usually from $-\infty$ to $\infty$, however for our case it has to cover just the distribution $f(\omega_\beta)$ for the Fourier transform and it goes from 0 to $\infty$ for its inverse since the pulse response vanishes for negative times. Since we used an excitation of the form $G(t) = \tilde{G} e^{-i\omega t}$ we use in the Fourier transform $e^{i\omega t}$. The factor in front of the integration is chosen for convenience to be unity for the transform and $1/2\pi$ for the inverse. This choice is arbitrary but the product of the two factors has to be $1/2\pi$.

We take the cosine Fourier transform of the Green function
\[
\mathcal{F}_{\cos}[g(t)] = \int_0^\infty g(t) \cos(\omega t) dt = \int_{-\infty}^\infty f(\omega_\beta) \int_0^\infty \cos(\omega_\beta t) \cos(\omega t) dt d\omega_\beta.
\]

The integration over $t$ results in the $\delta$-function
\[
\int_0^\infty \cos(\omega_\beta t) \cos(\omega t) d\omega_\beta = \frac{1}{2} \int_{-\infty}^\infty \cos(\omega_\beta t) \cos(\omega t) d\omega_\beta = \pi \delta(\omega_\beta - \omega).
\]

This gives
\[
\int_0^\infty g(t) \cos(\omega t) dt = \pi \int_{-\infty}^\infty f(\omega_\beta) \delta(\omega_\beta - \omega) d\omega_\beta = \pi f(\omega) = r_r(\omega),
\]
For the imaginary part $r_i(\omega)$ we get
\[
\begin{align*}
  r_i &= -\text{PV} \int_{-\infty}^\infty \frac{f(\omega_\beta)}{\omega_\beta - \omega} d\omega_\beta = -\frac{1}{\pi} \text{PV} \int_{-\infty}^\infty \frac{r_r(\omega_\beta)}{\omega_\beta - \omega} d\omega_\beta \\
  &= -\frac{1}{\pi} \text{PV} \int_0^\infty g(t) \int_{-\infty}^\infty \frac{\cos(\omega_\beta t)}{\omega_\beta - \omega} d\omega_\beta dt.
\end{align*}
\]

The integral over $\omega_\beta$ is [5]
\[
\int_{-\infty}^\infty \frac{\cos(\omega_\beta t)}{\omega_\beta - \omega} d\omega_\beta = -\pi \sin(\omega t)
\]
results for the imaginary part of the transfer function
\[
r_i(\omega) = \int_0^\infty g(t) \sin(\omega t) dt = \mathcal{F}_{\sin}[g(t)],
\]
which is the sine Fourier transform of the pulse response $g(t)$. We conclude therefore that the transfer function is the Fourier transformation of the Green function and vice versa
\[
r(\omega) = \int_{-\infty}^\infty g(t) e^{i\omega t} dt = \int_0^\infty g(t) e^{i\omega t} dt \quad \text{and} \quad g(t) = \frac{1}{2\pi} \int_{-\infty}^\infty r(\omega) e^{-i\omega t} d\omega
\]
and the transfer function is related to the particle distribution $f(\omega_\beta)$ in frequency
\[
r(\omega) = r_r(\omega) + ir_i(\omega) = \frac{\dot{g}(\omega)}{G} = \pi f(\omega) - i\text{PV} \int \frac{f(\omega_\beta)}{\omega_\beta - \omega} d\omega_\beta.
\]

The relation between the real and imaginary part of the transfer function
\[
r_i(\omega) = -\int_{-\infty}^\infty \frac{f(\omega_\beta)}{\omega_\beta - \omega} d\omega_\beta = -2 \int_{-\infty}^\infty \frac{r_r(\omega_\beta)}{\omega_\beta - \omega} d\omega_\beta
\]
is called a Hilbert transform and is a consequence of the fact that the pulse response $g(t)$ is real and that it vanishes for negative arguments.

The transfer function and the Green function of a beam can be measured directly [7, 8, 9, 10]. It is a powerful method to find the distribution of frequencies and to estimate the stability of a beam.
2.6 Examples of transfer and Green functions

We give now two examples of frequency distributions \( f(\omega \beta) \) and calculate their responses to a harmonic excitation (transfer function \( r(\omega) \)) and to a pulse excitation (Green function \( g(t) \)). They are presented in Fig. 5 which contains also the stability diagrams which will be discussed later.

We start with a Gaussian distribution

\[
f(\omega \beta) = \frac{1}{\sqrt{2\pi}\sigma_\omega} \exp \left( -\frac{\omega_0^2}{2\sigma_\omega^2} \right) \quad \text{with} \quad \int_{-\infty}^{\infty} f(\omega \beta) d\omega \beta = 1.
\]

We calculate first the Green function

\[
g(t) = \frac{1}{\sqrt{2\pi}\sigma_\omega} \int_{-\infty}^{\infty} \exp \left( -\frac{\omega_0^2}{2\sigma_\omega^2} \right) \cos(\omega_0 t) d\omega \beta = \exp \left( -\frac{\sigma_\omega^2 t^2}{2} \right).
\]

The transfer function is best obtained by a Fourier transform of the Green function. For the real part we find

\[
r_r(\omega) = \int_{0}^{\infty} \exp \left( -\frac{\sigma_\omega^2 t^2}{2} \right) \cos(\omega t) dt = \frac{\sqrt{2\pi}}{2\sigma_\omega} \exp \left( -\frac{\omega^2}{2\sigma_\omega^2} \right).
\]

The imaginary part is

\[
r_i(\omega) = \int_{0}^{\infty} \exp \left( -\frac{\sigma_\omega^2 t^2}{2} \right) \sin(\omega t) dt = \frac{\sqrt{2}}{\sigma_\omega} \exp \left( -\frac{\omega^2}{2\sigma_\omega^2} \right) \int_{0}^{\omega/\sqrt{2}\sigma_\omega} \exp(t^2) dt',
\]

where the last integral is called Dawson integral \([11]\). The complex transfer function becomes

\[
r(\omega) = \frac{\sqrt{2\pi}}{2\sigma_\omega} \exp \left( -\frac{\omega^2}{2\sigma_\omega^2} \right) \left[ 1 + i \frac{2}{\sqrt{\pi}} \int_{0}^{\omega/\sqrt{2}\sigma_\omega} \exp(t^2) dt' \right].
\]

The transfer function can also be represented by giving its amplitude \( A(\omega) \) and phase \( \phi(\omega) \) instead of its real and imaginary part. Both representations of the transfer function as well as the Green function of a Gaussian distribution are shown in the upper part of Fig. 5.

Next we investigate the response of a set of particles having a Lorentz distribution in frequency

\[
f(\omega \beta) = \frac{2a}{a^2 + \omega_0^2}.
\]

The Green function \( g(t) \) is

\[
g(t) = \int_{-\infty}^{\infty} \frac{2a}{a^2 + \omega_0^2} \cos(\omega_0 t) d\omega \beta = \pi e^{-at}.
\]

From this we get the transfer function \( r(\omega) \) which can be expressed by its real and imaginary part or by its in amplitude \( A(\omega) \) and phase \( \phi(\omega) \)

\[
r(\omega) = \frac{2\pi a + i\omega}{a^2 + \omega^2}, \quad A(\omega) = \frac{2\pi}{\sqrt{a^2 + \omega^2}}, \quad \tan(\phi) = \frac{\omega}{a}.
\]

This Green function and the transfer function are shown in the lower part of Fig. 5. It is interesting to note that for the Lorentz distribution the Green function is an exponential, i.e. after a pulse excitation the center-of-mass of this beam will move exponentially to zero.
Figure 4: Response of oscillators with a Gaussian frequency distribution

Figure 5: Response of oscillators with a Lorentz frequency distribution
2.7 Narrow distributions around a central frequency

For accelerator applications the frequency distributions are not centered around the origin and they are usually very narrow, such that the width of this distribution is much smaller than the frequency itself. This is certainly the case for synchrotron and betatron side bands. We will adapt our results for the beam response to this condition. We choose a certain central frequency $\omega_{\beta c}$ of the distribution and introduce $\Delta \omega_{\beta} = \omega_{\beta} - \omega_{\beta c}$ as a variable and express the distribution

$$f(\omega_{\beta}) \rightarrow F(\Delta \omega_{\beta}).$$

For simplicity we assume the distribution $F(\Delta \omega_{\beta})$ to be symmetric in $\Delta \omega_{\beta}$ but the treatment could easily be generalized. To find the Green function we use the inverse Fourier transform (1)

$$g(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\Delta \omega_{\beta}) \cos(\omega_{\beta c} + \Delta \omega_{\beta} t) d\Delta \omega_{\beta}$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\Delta \omega_{\beta}) \left[ \cos(\omega_{\beta c} t) \cos(\Delta \omega_{\beta} t) - \sin(\omega_{\beta c} t) \sin(\Delta \omega_{\beta} t) \right] d\Delta \omega_{\beta}$$

$$= \frac{1}{2\pi} \left[ \cos(\omega_{\beta c} t) \int_{-\infty}^{\infty} F(\Delta \omega_{\beta}) \cos(\Delta \omega_{\beta} t) d\Delta \omega_{\beta}$$
LANDAU DAMPING

\[- \sin(\omega_{\beta}t) \int_{-\infty}^{\infty} F(\Delta \omega_{\beta}) \sin(\Delta \omega_{\beta}t) d\Delta \omega_{\beta} \].

The integrations have to include the whole narrow distribution for the positive and negative frequencies. It can be seen that the second integral in the square bracket vanishes due to the symmetry of the distribution. This leads to the result

\[ g(t) = \frac{1}{2\pi} \cos(\omega_{\beta}t) \int F(\Delta \omega_{\beta}) \cos(\Delta \omega_{\beta}t) d\Delta \omega_{\beta} = \cos(\omega_{\beta}t) E(t). \]

The pulse response of a narrow distribution consists of a fast oscillation with frequency \( \omega_{\beta} \) equal to the central frequency of the distribution modulated in amplitude by the envelope function \( E(t) \). The latter is the cosine Fourier transform of the narrow distribution \( F(\Delta \omega_{\beta}) \)

\[ E(t) = \frac{1}{2\pi} \int F(\Delta \omega_{\beta}) \cos(\Delta \omega_{\beta}t) d\Delta \omega_{\beta}. \]

It should be noted that this treatment is not exact since the central frequency is not known and has to be chosen. In most accelerator applications one is only interested in this envelope of the beam response to a \( \delta \)-pulse excitation. We show the transfer function and this envelope of the Green functions for a narrow Gaussian and Lorentz distribution are shown in Fig. 7.

\[ \text{2.8 Landau damping of oscillator set} \]

\[
\begin{align*}
    k \langle \dot{x} \rangle & \quad \text{amplifier} \quad \langle \dot{x} \rangle \\
    \text{kicker} & \quad \text{detector}
\end{align*}
\]

Figure 8: Feed-back system and Landau damping

Based on the center-of-mass response of a set of oscillators we illustrate how the frequency spread leads to Landau damping of a coherent oscillation which would otherwise grow. The velocity of the center-of-mass motion of a set of oscillators is measured by a detector, the signal is amplified and fed to a kicker to produce an acceleration \( G_s \) in phase with the velocity, as shown in Fig. 8. This represents a negative feedback system which can lead to a growing oscillation, i.e. an instability. The center-of-mass velocity response to an acceleration \( G = \hat{G} \exp(j\omega t) \) is in general and for a Gaussian distribution

\[
\langle \dot{x} \rangle = G [r_r + j r_i] = G \frac{1}{2} \left( \pi f(\omega) - j PV \int \frac{f(\Delta \omega_r)}{\omega - \omega_r} d\omega_r \right) \\
= G \left( \frac{\pi}{\sqrt{2\pi \sigma_{\omega}}} e^{-\Delta \omega^2/2\sigma_{\omega}^2} + j \frac{\sqrt{2}}{\sigma_{\omega}} e^{-(\Delta \omega/\sigma_{\omega})^2/2} \int_{0}^{\Delta \omega/(\sqrt{2}\sigma_{\omega})} e^{-t'^2} dt' \right). \]

We assume now that the excitation happens at the central frequency for the Gaussian distribution \( \Delta \omega = 0 \) for which \( r_i = 0 \)

\[ \langle \dot{x} \rangle = G \frac{\pi}{\sqrt{2\pi \sigma_{\omega}}}. \]
We replace the external excitation \( G \) by the one of the feedback 
\[ G_s = k \langle \dot{x} \rangle \]
and assume a gain \( k \) just sufficient to keep the oscillation going. This is the limit of stability since a slightly larger gain would increase the oscillation leading to an exponential growth
\[ \langle \dot{x} \rangle = k \langle \dot{x} \rangle \frac{\pi}{\sqrt{2\pi}\sigma_\omega} \rightarrow k \leq \sqrt{\frac{2}{\pi}}\sigma_\omega. \]
This maximum gain \( k \) still giving stability is proportional to the frequency spread. Landau damping works by making an accidental coherent oscillation incoherent at infinitesimal levels without having first a growth reaching finite amplitudes. It does not lead to a growth of the incoherent oscillations.

3 TRANSVERSE BEAM RESPONSE TO AN EXCITATION

3.1 Transverse oscillation modes of a coasting beam

A uniform coasting beam of \( N \) particles circulates with revolution frequency \( \omega_0 \), current \( I = eN\omega_0/(2\pi) \) in a ring of uniform focusing. Each particle executes a betatron oscillation of frequency \( Q\omega_0 \)
\[ \theta_i = \theta_{0i} + \omega_0 t \quad y_i(t) = \hat{y} \cos(Q\omega_0(t - t_i)). \]
Depending on the phase \( Q\omega_0 t_{ij} \) between adjacent particle we have different modes. We choose a set having a form as seen at a fixed location \( \theta \)
\[ y(t) = \hat{y} \cos(n\theta - \omega t) \quad y(0) = \hat{y} \cos(n\theta). \]
Frozen in time, \( t = 0 \), we have a closed wave with \( n \) periods as shown in Fig. 9. Following a particle \( \theta_s(t) = \theta_0 + \omega_0 t \) on its motion gives us the betatron oscillation with frequency \( Q\omega_0 \).
\[ y_s = \hat{y} \cos(n\theta_0 - (\omega - n\omega_0)t) = \hat{y} \cos(n\theta_s - Q\omega_0 t) \]
giving for the frequency \( \omega \) seen by a stationary observer
\[ \omega = (n + Q)\omega_0 = \omega_\beta \quad \text{with} \quad -\infty < n < \infty. \]

We divide modes into fast and slow waves according to the sign of the phase difference between adjacent particle
\[ \omega_{\beta f} = (n_f + Q)\omega_0 \quad n_f > -Q \]
\[ \omega_{\beta s} = (n_s - Q)\omega_0 \quad n_s > Q. \]
3.2 Distribution in momentum and betatron frequencies

The betatron frequencies of a beam with nominal momentum are:

$$\omega_{bf} = (n_f + Q)\omega_0, \quad \omega_{bs} = (n_s - Q)\omega_0.$$ 

A deviation $\Delta p$ results in a change of these frequencies through the relation

$$\frac{\Delta E}{E} = \beta^2 \frac{\Delta p}{p} = -\frac{\beta^2 \Delta \omega_0}{\eta_c \omega_0}, \quad \Delta Q = Q' \frac{\Delta p}{p},$$

resulting in

$$\Delta \omega_{bf} = (Q' - \eta_c (n_f + Q))\omega_0 \frac{\Delta p}{p}, \quad \Delta \omega_{bs} = (Q' - \eta_c (n_s - Q))\omega_0 \frac{\Delta p}{p}. \quad (14)$$

We assume a symmetric distribution $F_p(\Delta p)$ which also results in a revolution frequency distribution $F_{\omega}(\Delta \omega_0)$ as well as in a distribution of the two betatron side band frequencies $f(\omega_{bf}), f(\omega_{bs})$, shown in Fig. 10. These latter two behave like a set of oscillators with a spread in resonant frequency.

3.3 Response of particle string with very small momentum spread

We consider now a narrow string of beam with all particles having the same momentum. We excite it

$$\ddot{y} + \omega_0 Q^2 y = \hat{G} \cos(\omega)$$
and seek a solution \( y(t) = \hat{y} \cos(n\theta - \omega) \), 11. To excite such a mode \( n \), each particle has to be driven by the proper phase corresponding to its longitudinal position \( \theta \). Therefore, we expect to find a excitation frequency which is not \( \omega_0 Q \) but close to the fast or slow wave frequencies \( \omega_{bf} = (n_f + Q)\omega_0 \) or \( \omega_{bs} = (n_s - Q)\omega_0 \). Substituting the desired solution form in the differential equation form gives

\[
(- (n\omega_0 - \omega)^2 + Q^2 \omega_0^2) \dot{\hat{y}} \cos(n\theta - \omega) = \hat{G} \cos(\omega t).
\]

We assume excitation and observation is done at the location \( \theta = 0 \)

\[
\dot{\hat{y}} = \frac{\hat{G}}{\omega_0^2 Q^2 - (n\omega_0 - \omega)^2} = \frac{-\hat{G}}{(\omega - \omega_0(n + Q)(\omega - \omega_0(n - Q))} = \frac{-\hat{G}}{(\omega - \omega_{bf})(\omega - \omega_{bs})} = \frac{\hat{G}}{2\omega_0 Q} \left( \frac{1}{\omega - \omega_{bs}} - \frac{1}{\omega - \omega_{bf}} \right).
\]

to excite the fast wave we use \( \omega \approx (n_f + Q)\omega_0 \) and the first term is much smaller than the second one. Correspondingly for the slow wave we use \( \omega \approx (n_s - Q)\omega_0 \) and the second term is much smaller than the first one. We approximate for the two waves

\[
\left( \frac{\dot{\hat{y}}}{G} \right)_f \approx - \frac{1}{2\omega_0 Q} \left( \frac{1}{\omega - \omega_{bf}} \right), \quad \left( \frac{\dot{\hat{y}}}{G} \right)_s \approx \frac{1}{2\omega_0 Q} \left( \frac{1}{\omega - \omega_{bs}} \right).
\]

The two responses have opposite sign, this will be discussed later.

### 3.4 Response of a beam with finite momentum spread

The whole beam has frequency distribution \( f(\omega_{bf}) \) and \( f(\omega_{bs}) \) The center-of-mass responses in displacement and velocity are related \( \langle \dot{y} \rangle = j\omega \langle y \rangle \)

\[
\langle \dot{\hat{y}} \rangle_f = -\frac{\hat{G} \omega}{2Q \omega_0} \int \frac{f(\omega_{bf})}{\omega_{bf} - \omega} d\omega_{bf} = -\frac{\hat{G} \omega}{2Q \omega_0} \left( \int \frac{f(\omega_{bf})}{\omega - \omega_{bf}} \right) d\omega_{bf},
\]

\[
\langle \dot{\hat{y}} \rangle_s = \frac{\hat{G} \omega}{2Q \omega_0} \int \frac{f(\omega_{bs})}{\omega_{bs} - \omega} d\omega_{bs} = \frac{\hat{G} \omega}{2Q \omega_0} \left( \int \frac{f(\omega_{bs})}{\omega - \omega_{bs}} \right) d\omega_{bs}.
\]

The term \( \pi f(\omega) \) is real, exciting acceleration and responding velocity are in phase resulting in an absorption of energy and damping, called Landau damping. It is only present if the excitation frequency \( \omega \) is within the frequency distribution of the individual particles. The second term is imaginary and gives the out-of-phase response being of less interest.

The spread in betatron frequencies is given by the momentum spread and the dependence of revolution frequency \( \omega_0 \) and betatron tune \( Q \) on momentum deviation \( \Delta p/p \). It is therefore determined by an external parameter which is not affected by the excitation of betatron oscillations.

### 3.5 Measuring the beam transfer function

It is often useful to measure the beam response to a harmonic excitation to obtain information about the betatron frequency distribution and to estimate the amount of Landau damping. The center-of-mass displacement of the beam due to an external excitation can directly be measured with a network analyzer as shown in Fig. 12. This device compares the relative amplitude and phase of the beam signal from a position monitor with
the harmonic driving signal of frequency $\omega$ sent to the beam. The frequency $\omega$ is swept through the though the range of interest to cover the lower or upper betatron side band. It is also possible to drive the beam with white noise and compare the Fourier transform of response and excitation directly using a Fast Fourier Transform (FFT) circuit. In the theory we liked to use the velocity response which is more transparent for understanding the resistive and reactive behavior of the beam. In measurements the displacement is observed and our equation have to be converted to analyze the results. Due to cable delays the real and imaginary part of the response are often mixed. It is easier to measure amplitude and phase response and correct the latter off-line.

In the first example [8], the lower betatron side band of a coasting beam is excited and its measured response, show in Fig. 13. The amplitude, plotted in logarithmic scale, shows a maximum response at the frequencies contained in the beam, but with long tails.

The second example [9], shown in Fig. 14, involves the measurement of both side bands. On the upper part the response in amplitude and phase is plotted. Knowing that each side band involves a phase change of $\pi$, we can get an absolute value of the phase and calculate the resistive (real) and reactive (imaginary) responses which are plotted on the bottom of the figure. It confirms the positive and negative sign of the resistive response at the lower and upper side band obtained already in the calculation.

4 TRANSVERSE BEAM INSTABILITY AND LANDAU DAMPING

4.1 Transverse impedance

It is well known that a beam can excite longitudinal modes in a cavity which than act back on the beam as illustrated on the left of Fig 15. The property of the cavity, relevant for this effect, is quantified by a frequency dependent longitudinal impedance $Z_T(\omega)$ which has a resistive (real) and a reactive (imaginary) part, in which the induced voltage is in or out of phase compared to the exciting current. It is also possible that the beam excites so-called deflecting modes which give a transverse force to the beam. A simple case is illustrated on the right of Fig 15. Here, a positive charge is going through a cavity with a displacement $y$ from the axis and excites a mode which has first a longitudinal electric field with a transverse gradient. A quarter of an oscillation, $T_r/4$, later this mode is converted into a transverse magnetic field which can deflect particles. The relevant cavity property is quantified by a transverse impedance, defined as the integrated deflecting field per unit dipole moment of the exciting current

$$Z_T(\omega) = -\frac{i \frac{2\pi R}{I_0} \left[ \vec{E}(\omega) + [\vec{\beta} \times \vec{B}(\omega)] \right]_{T} ds}{I_y(\omega)}.$$ 

The right hand side is multiplied with $i$ which indicates that the driving dipole moment $I_y$ is out of phase with the deflecting field. The above expression is matched to a complex description of an oscillation with $e^{-jwt}$. If one uses the convention $e^{jwt}$ instead, $i$ has to be replaced by $-j$. This impedance might become more clear if we relate the fields to the
vertical velocity $\dot{y}$ rather than to the position $y$. With $y = \dot{y}e^{-i\omega t}$ we have $\dot{y} = -i\omega y$ and

$$Z_T(\omega) = -\omega \int_{0}^{2\pi R} [\vec{E}(\omega) + [\vec{B} - \vec{B}]T] ds.$$

A real transverse impedance means that the transverse deflecting fields are in phase with the transverse velocity and transfer energy to or from the transverse motion of the beam.

### 4.2 Transverse beam stability criterion

We calculated the center-of-charge (coherent) response of the beam to external transverse excitation with acceleration $G$

$$G = \frac{\langle e[\vec{E} + [\vec{B}]T] \rangle}{m_0 \gamma}$$

which gives for the fast wave

$$\langle y \rangle = -\frac{\dot{G}e^{-i\omega t}}{2Q\omega_0} \int \frac{f(\omega_{3f})}{\omega_{3f} - \omega} d\omega_{3f}.$$

If the beam surroundings represent a transverse impedance $Z_T$

$$Z_T(\omega) = -i \int_{0}^{2\pi R} [\vec{E}(\omega) + [\vec{B} - \vec{B}]T] T ds.$$
Figure 14: Measure beam transfer function of the lower and upper side betatron band; top: amplitude and phase, bottom: resistive part

the oscillating beam induces fields in it which will act back on the beam. We assume now that the acceleration $G$ is not external but due to the self fields induced

$$G = \frac{ieZ_T I(y)}{\gamma m_0 2\pi R}.$$ 

Substituting this in the expression for the beam response we get for the fast and the slow wave

$$1 = -\frac{ieIZ_T}{4\pi Q m_0 c^2 \gamma} \int \frac{f(\omega_{\beta f})}{\omega_{\beta f} - \omega} d\omega_{\beta f} \quad \text{and} \quad 1 = \frac{ieIZ_T}{4\pi Q m_0 c^2 \gamma} \int \frac{f(\omega_{\beta s})}{\omega_{\beta s} - \omega} d\omega_{\beta s}.$$ 

This gives a condition for which the beam is just at the limit of stability [13, 14]. By replacing the external excitation with the self induced one we assume that an oscillation, once started, is just kept going by the self forces. To make it more applicable we introduce some normalization to separate the term which depends on the different beam parameters
from the one which is just given by the form of the distribution. We introduce the half width at half height $S$ of the distribution $f(\omega_\beta)$ and normalize the two frequencies $\omega$ and $\omega_\beta$ with it

$$\xi_f = \frac{\omega_f}{S}, \xi_s = \frac{\omega_s}{S}, \xi = \frac{\omega}{S}, f(\xi_f) = S f(\omega_\beta f), f(\xi_s) = S f(\omega_\beta s)$$

and get

$$1 = \frac{iecIZ_T}{4\pi EQS} \int \frac{f(\xi_f)}{\xi - \xi_f} d\xi_f$$

$$1 = -\frac{iecIZ_T}{4\pi EQS} \int \frac{f(\xi_s)}{\xi - \xi_s} d\xi_s.$$

This expression allows us to find, for given beam parameters, for each value of the driving frequency $\omega$ (or its normalized value $\xi$), the maximum impedance $Z_T$ which still does not lead to an instability. In this expression the integral and the impedance are complex numbers. It is therefore convenient to visualize the stability criterion by mapping the impedance to the frequency $\omega$ or $\xi$. As long as this frequency has no imaginary part there is no growing instability since we expressed the oscillation as $e^{-i\omega t}$. We write the above condition slightly different

$$\frac{iecIZ_T}{4\pi EQS} = V_T + iU_T = \frac{1}{i \int \frac{f(\xi_f)}{\xi_f} d\xi_f}$$

and

$$\frac{iecIZ_T}{4\pi EQS} = V_T + iU_T = -\frac{1}{i \int \frac{f(\xi_s)}{\xi_s} d\xi_s}.$$

Plotting the above equality for a real frequency $\xi_1$ gives the stability diagram shown in Fig. 16. Unfortunately the above equations rarely give solutions which can be expressed in elementary function for realistic distributions. We give in Fig. 16 the example of a Gaussian distribution which we have calculated earlier. The diagram consists of two curves; at the right, one for the slow wave with positive values of the impedance; at the left, one for the fast wave with negative values for the resistive impedance. These curves represent a situation which is at the limit of stability. A slightly larger impedance than the one corresponding to this limit will lead to an instability. For the slow wave this means that values for the complex impedance $Z_T$, lying on the right of the stability diagram, lead to instability while for corresponding values lying on the left we still have stability. For the fast wave only negative impedances lead to instabilities if the resulting impedance lies on the left of the fast wave stability curve. To summarize the situation for both waves, we have stability as long as the impedance is inside the stability diagram bounded by the fast and the slow wave stability limit curves.
4.3 Landau damping for bunched beams and for longitudinal instabilities

Without going into any details concerning the Landau damping for bunched beams we should make a few remarks on the limitation of the treatment presented here. The Landau damping against transverse instabilities in an un-bunched beam, which has a betatron frequency spread due to the energy spread and chromaticity (14), is a special case. The betatron oscillation we excite will not influence the frequency spread in the beam. In other words, the frequency $\omega_\beta$ of a particle is given by parameters like $\Delta p$, $\eta$ and $Q'$ which are not influenced by the excitation of betatron oscillations. This situation is different and more complicated in cases where the betatron frequency spread is determined by octupole fields which give a dependence of the betatron frequency on amplitude. This is always the case for a bunched beam but it can also be the dominant effect in un-bunched beams. Exciting a betatron oscillation will, at the same time, influence the betatron frequency distribution. Going through this calculation one finds that the integrals, (3), determining the beam response and the stability diagram does not contain the distribution $f(\omega_\beta)$ but rather its derivative [13, 14]. The same situation is present for longitudinal stability for un-bunched as we will see in the next section.

5 LONGITUDINAL BEAM INSTABILITY AND LANDAU DAMPING

5.1 Introduction

We consider now the same coasting (un-bunched) beam and investigate its longitudinal stability provided by Landau damping. Again we first have to study the response of the beam to an excitation in the longitudinal direction or in energy. In the transverse case, treated before, the frequency spread was determined by the energy distribution of the particles combined with the betatron frequency dependence on energy. The excita-
tion affected the amplitude of the betatron oscillations but not their frequencies. The longitudinal response is more complicated. The spread in revolution frequencies is given by the energy distribution combined with the relation between energy and revolution frequency deviations. The beam is excited in energy which directly affects the frequency distribution. This situation is more complicated and the longitudinal response is usually calculated with the Vlasov equation. Before using this method we make a short study of the longitudinal dynamics and try to illustrate the beam response by a more qualitative description.

5.2 Longitudinal dynamics

We consider an un-bunched (coasting) beam circulating in a storage ring with the nominal radius $R_0$, the nominal revolution frequency $\omega_0 = \beta_0 c / R_0$, the nominal energy and momentum $E_0 = m_0 c^2 \gamma_0$, $p_0 = m_0 c \beta_0 \gamma_0$. A particle with a different momentum $p = p_0 + \Delta p$ has a relative energy deviation

$$\frac{\Delta E}{E_0} = \frac{E - E_0}{E_0} = \beta_0^2 \frac{\Delta p}{p_0}.$$  

This particle will move on an orbit which is different from the nominal one by the radial distance

$$\Delta R = D_x \frac{\Delta p}{p_0} = D_x \frac{\Delta E}{E_0},$$

where $D_x$ is the horizontal dispersion. The length $L$ of the orbit has a relative difference compared to the nominal one by

$$\frac{\Delta L}{L_0} = \frac{\Delta p}{p_0} \int_0^{2\pi} \frac{D_x(\theta)}{\rho} d\theta = \alpha_c \frac{\Delta p}{p_0},$$

where $\rho$ is the radius of curvature and $\alpha_c$ the momentum compaction. The revolution time $T = L / \beta c$ is also changed

$$\frac{\Delta T}{T_0} = \frac{\Delta L}{L_0} - \frac{\Delta \beta}{\beta_0} = \left( \alpha - \frac{1}{\gamma^2} \right) \frac{\Delta p}{p_0},$$

where we used the relation

$$\frac{\Delta p}{p_0} = \gamma^2 \frac{\Delta \beta}{\beta}.$$  

As a consequence the revolution frequency $\omega_r$ is changed by

$$\frac{\Delta \omega_r}{\omega_0} = -\frac{\Delta T}{T_0} = - \left( \alpha - \frac{1}{\gamma^2} \right) \frac{\Delta p}{p_0} = -\eta \frac{\Delta p}{p_0} = - \frac{\eta}{\beta^2} \frac{\Delta E}{E_0},$$

where we introduced $\eta = \alpha - 1 / \gamma^2$.

A realistic beam has many particles with a certain distribution

$$f(\Delta E, \theta) = \frac{1}{N} \frac{d^2 N}{d\theta dE} \text{ with } \int_0^{2\pi} d\theta \int_{-\infty}^{\infty} f(\theta, \Delta E) dE = 1$$

in the energy deviation $\Delta E$. In general this distribution can depend on the azimuthal angle $\theta$. However, the equilibrium distribution is independent of $\theta$, $g_0 = f_0(\Delta E)$. Instead of giving the distribution with respect to $\Delta E$ we can also give it with respect to the revolution frequency $\omega_r$

$$F(\theta, \Delta \omega_r) = \frac{1}{N} \frac{d^2 N}{d\theta d\omega_r} = \frac{1}{N} \frac{d \omega_r}{d\theta} \frac{d^2 N}{d\omega_r dE} = \frac{\beta^2 E_0}{\eta \omega_0} f(\theta, \Delta E).$$

For convenience we assumed a symmetric equilibrium distribution

$$f_0(-\Delta E) = f_0(\Delta E).$$

292
5.3 Illustration of the beam response to a pulse excitation

We take now a coasting beam having a certain energy distribution $f_0(\Delta E)$ with RMS value $\sigma_E$. We introduce at the time $t = 0$ a disturbance in energy of the form

$$\delta E(\theta) = \delta E_0 \sin(n\theta)$$

for all particles in the beam. This is illustrated in Fig. 17 for the case $n = 2$. It shows the beam over one turn, $\theta = 0 \rightarrow 2\pi$ with the lines indicating points of equal density $f(\Delta E, \theta)$. Two particles are marked by a point to give their position and by an arrow indicating their motion with respect to a nominal particle at $\theta = \pi$ and $\Delta E = 0$. A cross section through the beam at $\theta = \pi$ is shown on the right. At the bottom we show the current $I(\theta)$ represented by this beam which is given by the integral

$$I(\theta) = \frac{Ne\omega_0}{2\pi} \int f(\theta, \Delta E)d(\Delta E).$$

The first frame gives the situation just after the excitation. Since the disturbance affects the energy only, the current $I$ is not changed by it and is still uniform $I(\theta) = I_0$ and equal to its average value. The dependence of the revolution frequency on energy deviation of the particles

$$\frac{\Delta \omega_r}{\omega_0} = \frac{\eta}{\beta^2 E_0} \Delta E$$

results in a relative longitudinal displacement of regions with different energy as time goes on. This is illustrated in the second frame of the figure which shows the situation after $N$ turns for which

$$2\pi N \frac{\eta}{\beta^2 E_0} \sigma_E = 0.3$$

which means that after $N$ turns a particle with an energy deviation $\Delta E = \sigma_E$ has slipped by an angle 0.3 rad compared to a particle with nominal energy. Due to this shearing the original energy disturbance is partially converted into a modulation of the beam current. This gets even more pronounced a little later as shown in the third frame. However, as time goes on the shearing becomes more pronounced and the modulations at the different energy deviations $\Delta E$ overlap. This reduces the current modulation again which will slowly disappear as illustrated in the last two frames.

5.4 The Vlasov equation

The application of kinetic theory and the related Vlasov equation has been treated in earlier CAS courses [15, 16, 17]. Here, we only make a short and not rigorous derivation and put all the emphasis on the application to Landau damping. We start with Liouville’s theorem which says that the phase space density $\psi(q, p)$ in a small volume around a given particle does not change with time

$$\frac{d\psi(q, p)}{dt} = 0.$$ (15)

Usually we associate $q$ with a space coordinate and $p$ with a momentum or a velocity but the situation can be more general. These coordinates $q$ and $p$ should be canonically conjugate which means that they are derived from a Hamiltonian $H(q, p)$ by the canonical equations

$$\dot{q} = \frac{\partial H(q, p)}{\partial p}, \quad \dot{p} = -\frac{\partial H(q, p)}{\partial q}.$$ 

The theorem of Liouville is visualized in Fig. 18 where the phase-space trajectories of three particles are drawn. Their positions and the surface element, they determine, are
Figure 17: Development of an initial energy disturbance of an un-bunched beam
shown for two different times $t$. The form of this surface element changes but its area stays constant. According to Liouville’s theorem the particles move in phase space like an incompressible liquid. This is also illustrated in the development of the energy disturbance shown in Fig. 17. The constancy of the phase density $\psi(q, p)$ is expressed by (15) where the absolute differentiation indicates that one follows the particle while measuring the density of its immediate neighborhood. We would like to know the development of this density as seen by a stationary observer (like a beam monitor) which does not follow the particle. It depends now not only directly on the time $t$ but also indirectly through the coordinates $q$ and $p$ of the moving particles which change with time. Therefore, we have to express the absolute differential with respect to $t$ by the partial differentiations with respect to time as well as with respect to the coordinates $q$ and $p$ multiplied with the time derivative of these coordinates

$$\frac{d\psi(q, p)}{dt} = \frac{\partial\psi(q, p)}{\partial t} + q\frac{\partial\psi(q, p)}{\partial q} + \dot{p}\frac{\partial\psi(q, p)}{\partial p} = 0.$$  

This expression is the Vlasov equation in its most simple form and is nothing else but an expression for Liouville’s conservation of phase-space density seen by a stationary observer. Rigorous derivation of the Vlasov equation can be found in the literature e.g. [15, 19]. It is interesting to apply the Vlasov equation to a stationary distributions which does not depend explicitly on time. For example a stationary bunch looks the same each revolution as it is observed through an intensity monitor. For this stationary case the Vlasov equation becomes

$$\frac{\partial\psi(q, p)}{\partial t} = 0 \rightarrow q\frac{\partial\psi(q, p)}{\partial q} + \dot{p}\frac{\partial\psi(q, p)}{\partial p} = 0.$$  

Expressing $\dot{q}$ and $\dot{p}$ with the canonical equations we get

$$\frac{\partial\psi}{\partial q}\frac{\partial H}{\partial p} - \frac{\partial\psi}{\partial p}\frac{\partial H}{\partial q} = 0.$$  

From this it can be shown that a stationary distribution $\psi$ is a function of the Hamiltonian $H$ only $\psi = \psi(H)$. This fact is often useful in applications, specially in finding stationary distributions [20, 21]. In this presentation we apply it to determine the response of a

Figure 18: Phase space trajectories and Liouville’s theorem
beam to an external excitation. This will be one of the ingredients to treat the Landau damping of longitudinal instabilities.

Before we go to the practical applications we have to make some remarks on the coordinates \( q \) and \( p \) we use. In Liouville’s theorem the phase-space area is only conserved if expressed in canonically conjugate variables \( q, p \). The same criterion applies to the validity of the Vlasov equation. Examples of such pairs are position and momentum \( x, p \) or time and energy \( t, E \). These variables are often not very practical for accelerator applications. We could for example choose the longitudinal position \( s \) and the corresponding momentum \( p_s \) as coordinates to describe the longitudinal phase-space. However the path length \( s \) around the ring is different for particles with different momenta. As a result, individual particles would enter a certain machine element, such as a quadrupole, at different values of the coordinate \( s \). From this point of view other coordinate pairs in which the position of the machine element is given for all particles by one coordinate only are more suitable. Such a variable is for example the angle \( \theta \) of the motion around the ring. However the conjugate variable is quite complicated [22]

\[
W = \int_{E_0}^{(E_0 + \Delta E)} \frac{\Delta E}{\omega} d(\Delta E)
\]

where \( \Delta E \) is the deviation from the nominal energy \( E_0 \) and \( \omega(\Delta E) \) is the revolution frequency which depends on this energy deviation. In our applications we only use terms up to first order in \( \Delta E \) and use as an approximation \( \theta \) and \( \Delta E \) as variables. However we should always remember that this is a linear approximation in \( \Delta E/E_0 \) which can become very inaccurate close to transition energy. As a further convenience we will use the RF phase angle \( \phi \) instead of \( \theta \) to which it is related through \( \phi = h\theta \) with \( h \) being the harmonic number.

5.5 Response of the beam to a longitudinal excitation

We follow here the methods used in [14, 23] and consider a storage ring containing a cavity which is driven with frequency \( \omega \approx n\omega_0 \) and amplitude \( U_0 \)

\[
U(t) = U_0 e^{-i\omega_0 t}.
\]

At each revolution the energy of each particle is changed by

\[
\Delta E = eU(t_i) = eU_0 e^{-i\omega_0 t_i}
\]
due to the cavity voltage where \( t_i \) is the time the particle passes through the cavity. Assuming a relatively small energy change per turn we can make a smooth approximation as if the cavity voltage was distributed uniformly around the ring

\[
\Delta \dot{E} = \dot{E} = \frac{\omega}{2\pi} \Delta E = \frac{\omega_0 U_0}{2\pi} e^{i(n\theta - \omega t)}.
\]

(16)

We use now the Vlasov equation to describe the development of the particle distribution. As mentioned before we should express it in canonically conjugate variables \( \theta \) and \( W \). However for our approximation of small energy deviations we use \( \theta \) and \( \Delta E \) as variables and get

\[
\frac{\partial f(\theta, \Delta E)}{\partial t} + \frac{\partial f(\theta, \Delta E)}{\partial \theta} \dot{\theta} + \frac{\partial f(\theta, \Delta E)}{\partial \Delta E} \dot{E} = 0
\]

The energy change \( \dot{E} \) (16) has to be used in the Vlasov equation. We expect that it will establish an energy modulation of the particles of the same form resulting in a modification of the distribution function

\[
f(\theta, \Delta E) = f_0(\Delta E) + f_1(\Delta E) e^{i(n\theta - \omega t)}
\]

296
which we assume to be small \( f_1 \ll f_0 \) such that quadratic terms in \( f_1 \) as well as terms containing products \( f_1 U_0 \) can be neglected. With this we get from the Vlasov equation

\[
i(\omega - n\omega_r) f_1(\Delta E)e^{i(n\theta - \omega t)} = -\frac{d_0(\Delta E)}{d(E)} \frac{e^{i\omega_0 U_0}}{2\pi} e^{i(n\theta - \omega t)}.
\]

We can replace the revolution frequency \( \omega_r \) of the individual particles on the right hand side by the nominal frequency \( \omega_0 \) (but not on the left hand side) leading to

\[
f_1(\Delta E)e^{i(n\theta - \omega t)} = \frac{ie^{i\omega_0 \omega t}}{2\pi} \frac{d_0(\Delta E)}{\omega - n\omega_r}.
\]

We cannot observe the distribution \( f(\theta, \Delta E) \) easily but we can measure the beam current \( I(\theta, t) \) which is given by

\[
I(\theta, t) = N e^{i\omega_0 \omega t} \int_{-\infty}^{\infty} f(\theta, \Delta E) d(\Delta E).
\]

The above integration does not really go from \(-\infty\) to \(\infty\) but has just to cover the distribution \( f(\theta, \Delta E) \). We split the current into the unperturbed part \( I_0 \) and a perturbation \( I_1(\theta, t) \). For convenience we chose \( \theta = 0 \) and get for the latter

\[
I_1(0, t) = \hat{I}_1 e^{-\omega t} = \frac{iN e^{2\omega_0^2 U_0 e^{-i\omega t}}}{2\pi} \int \frac{d_0(\Delta E)}{\omega - n\omega_r} d\Delta E.
\]

The response will be large if \( \omega \approx n\omega_0 \). The revolution frequency \( \omega_r \) of the individual particles depends on its energy deviation \( \Delta E \)

\[
\frac{\Delta \omega_r}{\omega_0} = -\frac{\eta \Delta E}{\beta^2 E_0}.
\]

The above integration might contain a pole. To investigate this we use \( \omega_r \) as an independent variable

\[
\hat{I}_1 = iN e^{2\omega_0^2 \eta U_0} \frac{1}{2\pi \beta^2 E_0} \int \frac{dF_0(\omega_r)}{\omega - n\omega_r} d\omega_r.
\]

Such integrals containing a pole have a principle value and a residue as we explained before

\[
\int \frac{dF_0(\omega_r)}{\omega - n\omega_r} d\omega_r = PV \int \frac{dF_0(\omega_r)}{\omega - n\omega_r} d\omega_r \pm i\pi \frac{dF_0(\omega_r)}{d\omega_r}(\omega).
\]

The principle value integral is obtained by leaving a small gap around the pole during integration and letting this gap go to zero

\[
PV \int \frac{dF_0(\omega_r)}{\omega - n\omega_r} d\omega_r = \int_{-\infty}^{\omega/n-\epsilon} \frac{dF_0(\omega_r)}{d\omega_r} d\omega + \int_{\omega/n+\epsilon}^{\infty} \frac{dF_0(\omega_r)}{d\omega_r} d\omega + \int_{\omega/n+\epsilon}^{\infty} \frac{dF_0(\omega_r)}{d\omega_r} d\omega
\]

with \( \epsilon \to 0 \).

The residue is just the derivative \( dF_0/\omega \) taken at the frequency \( \omega \) of excitation. We get for the perturbed current

\[
\hat{I}_1(\omega) = \frac{N e^{2\omega_0^2 \eta U_0}}{2\pi \beta^2 E_0} \left( \mp \pi \frac{dF_0(\omega_r)}{d\omega_r}(\omega) + iPV \int \frac{dF_0(\omega_r)}{d\omega_r}(\omega) d\omega \right) \tag{17}
\]

The current \( \hat{I}_1(\omega) \) induced by the excitation has a real and an imaginary part. The real component of the current is in phase with the excitation voltage and therefore absorbs
power. This can lead to damping, called Landau damping. The sign of this term seems to be ambiguous. This is due to the fact that we did not specify the initial condition. In a natural experiment there will be no perturbed current when one starts the excitation. It will appear only later as a consequence of the external voltage. In this case the positive sign should be taken. It is however also possible that the beam has already a modulation at the start and the external voltage can be such as to reduce this perturbation which would be expressed by the negative sign. We assume now the first case in which energy is absorbed by the beam and take the positive sign.

The imaginary part of the induced current is $90^\circ$ out of phase compared to the excitation and does not lead to energy absorption and damping.

5.6 Response of the beam to a pulse excitation

Instead of investigating the beam to a harmonic excitation (transfer function) we can also establish a sudden energy modulation and observe the free motion of the particles thereafter. This is the response to a pulse excitation also called Green function. We have already discussed this case qualitatively and illustrated it in Fig. 17.

We like to start with a simple form of an energy modulation of the beam and consider a harmonic dependence of the energy deviation around the ring which is closed in itself. At the time $t = 0$ we give each particle an energy change of the form

$$\delta E(\theta) = \delta E_0 \cos(n\theta).$$

This modulation is closed in itself around the ring after $n$ periods. In a real machine one does not have cavities distributed around the ring and such a modulation can be obtained approximately by applying a harmonic voltage to a single cavity during one turn. With the particle distribution $f_0(\Delta E)$ before the pulse excitation we will have the modified distribution immediately after the excitation at the time $t = 0^+$ assuming a small modulation

$$f(0^+, \theta, \Delta E) = f_0(\Delta E + \delta E_0 \cos(n\theta)) = f_0(\Delta E) - \frac{df_0}{dE} \delta E_0 \cos(n, \theta).$$

We assume that the induced current is observed at the location $\theta_0$ and consider development of the disturbance in time. Each particle moves around the ring with the angular velocity $\omega_r$. The particle which was at the time $t = 0$ at the position $-\theta$ will be at the point of observation $\theta_0$ at the time $t = (\theta - \theta_0)/\omega_0$. The distribution function at the time $t$ is therefore

$$f(\theta_0, \Delta E, t) = f_0(\Delta E) - \frac{df_0}{dE} \delta E_0 \cos(n\theta_0 - n\omega_r t).$$

To simplify the expression we assume that the beam is observed at the origin $\theta_0 = 0$

$$f(\Delta E, t) = f_0(\Delta E) - \frac{df_0}{dE} \delta E_0 \cos(n\omega_r t).$$

The resulting current $I(t)$ is

$$I(t) = Ne \int \omega_r f(\Delta E, t) dE \approx e\omega_0 \int f(\Delta E, t) dE = I_0 + I_1(t).$$

We assume now that the relative energy spread is small and remember that the revolution frequency $\omega_r$ of the individual particles is a function of their deviation $\Delta E$

$$\omega_r = \omega_0 + \Delta \omega_r, \quad \Delta \omega_r = -\omega_0 \eta \frac{\Delta E}{E}.$$
With this we get for the perturbed current.

\[ I_1(t) = \frac{Ne \omega_0 \eta \delta E}{\beta^2 E_0} \int \frac{dF_0}{d\omega_r} \cos(n\omega_0 t + n\Delta \omega_r t) d\omega_r \]  

Making use of the assumed symmetry

\[ F_0(\Delta \omega_r) = F_0(-\omega_r) \rightarrow \frac{dF_0}{d\omega_r}(\Delta \omega_r) = -\frac{dF_0}{d\omega_r}(\Delta \omega_r) \]
we find that the first term of the above integration vanish and the perturbed current becomes

\[ I_1(t) = -\frac{Ne \omega_0 \eta \delta E}{\beta^2 E_0} \sin(n\omega_0 t) \int \frac{dF_0}{d\omega_r} \sin(n\Delta \omega_r t) d\omega_r = -\frac{Ne \omega_0 \eta \delta E}{\beta^2 E_0} \sin(n\omega_0 t) F(t). \]

The factor in front of the integral of the above equation represents an oscillation with frequency \(n\omega_0\). The integral \( F(t) \) is slowly varying and represents the envelope of the oscillation. It is often called Green function.

Comparing the pulse response of the beam (Green function) with the response to a harmonic excitation (transfer function) (17) we find that it is proportional to the inverse sine Fourier transform of the latter. We would at first expect that they are related by the cosine Fourier transform. However the excitation with a pulse leads first to a modulation in energy which transforms into a current modulation only later. It can be shown that the complex transfer function is the Fourier transform of the Green function.

### 5.7 Example of a Gaussian distribution

We take the case of a Gaussian particle distribution in energy and revolution frequency deviation

\[ F_0(\Delta \omega_r) = \frac{1}{\sqrt{2\pi} \sigma_{\omega}} e^{-\frac{1}{2} \left( \frac{\Delta \omega_r}{\sigma_{\omega}} \right)^2}, \]

where \(\sigma_{\omega}\) is the RMS width of the distribution in \(\omega_r\). The response to a harmonic excitation is according to (17)

\[ \tilde{I}_1(\omega) = \frac{Ne^2 \omega_0^3 \eta U_0}{2\pi \beta^2 E_0} \left( \frac{dF_0}{d\omega_r} \right)(\Delta \omega) + \text{PV} \int \frac{dF_0(\omega_r)}{d\omega_r} \frac{\Delta \omega_r}{\omega - n\omega_r} d\omega_r \]

\[ = \frac{e^2 \omega_0^3 \eta U_0}{2\pi \beta^2 E_0 n} \frac{N}{\sqrt{2\pi} \sigma_{\omega}^2} \left( -\frac{\Delta \omega}{\sigma_{\omega}} e^{-\frac{1}{2} \left( \frac{\Delta \omega}{\sigma_{\omega}} \right)^2} - \text{PV} \int \frac{\Delta \omega_r}{\sigma_{\omega}} e^{-\frac{1}{2} \left( \frac{\Delta \omega_r}{\sigma_{\omega}} \right)^2} d\omega_r \right). \]

The principle value integral cannot be expressed in elementary functions. For the response to a pulse excitation we get

\[ F(t) = \int \frac{dF_0}{d\omega_r} \sin(n\Delta \omega_r t) d\omega_r = -\frac{N}{2\pi \sqrt{2\pi} \sigma_{\omega}^2} \int \frac{\Delta \omega_r}{\sigma_{\omega}} e^{-\frac{1}{2} \left( \frac{\Delta \omega_r}{\sigma_{\omega}} \right)^2} \sin(n\Delta \omega_r t) \frac{d\omega_r}{\sigma_{\omega}} \]

\[ = -\frac{N}{2\pi \sqrt{2\pi}} e^{-\frac{1}{2} \sigma_{\omega}^2 t^2}. \]

These responses of a Gaussian beam to a harmonic and a pulse excitation are plotted in Fig. 19.
Figure 19: Longitudinal response of a coasting beam with Gaussian distribution

5.8 Longitudinal stability

The response to a longitudinal excitation $U_0 e^{-i\omega t}$ is a current perturbation (17)

$$I_1 = \frac{i N e^2 \omega_0^3 \eta U_0}{2 \pi \beta^2 E_0} \int \frac{dF_0(\omega_r)}{d\omega_r} d\omega_r = \frac{N e^2 \omega_0^3 \eta U_0}{2 \pi \beta^2 E_0} \left( \mp \pi \frac{dF_0}{d\omega_r}(\omega) + iPV \int \frac{dF_0(\omega_r)}{d\omega_r} d\omega_r \right).$$

In a storage ring the beam is surrounded by different components like the vacuum chamber, bellows, kicker magnets, cavities etc. which represent a certain longitudinal impedance $Z(\omega)$. The perturbed current $I_1$ induces longitudinal fields in this impedance which can act back on the beam. Integrating the field component at one frequency $\omega$ over the circumference gives the voltage

$$U_{\text{ind}}(\omega) = -I_1(\omega) Z(\omega).$$

We consider now a situation in which the current perturbation $I_1$ established by the external voltage $U_0$ induces in the impedance a voltage which is equal to the applied voltage $U_0$. In this case the external voltage could be turned off and the current perturbation would continue to exist with constant amplitude. If the impedance is increased the induced voltage becomes larger than the one necessary to establish the original perturbation and the amplitude would grow. The case $U_{\text{ind}} = U_0$ represents therefore a stability limit. A
slightly larger impedance would lead to growth of a small perturbation at the frequency \( \omega \). We get the stability limit by replacing \( U_0 \) by \( U_{\text{ind}} \) in (5.8)

\[
I_1(\omega) = -\frac{i Ne^2 \omega_0^3 I_1(\omega) Z(\omega)}{2 \pi \beta^2 E_0 n} \int \frac{dF_0}{n} = \int \frac{dF_0}{n - \omega_r} d\omega_r = (22)
\]

\[
\frac{Ne^2 \omega_0^3 \eta Z(\omega)}{2 \pi \beta^2 E_0 n} \left[ \pi \frac{dF_0}{n} + i PV \int \frac{dF_0}{n - \omega_r} d\omega_r \right] (23)
\]

or

\[
1 = -\frac{i Ne^2 \omega_0^3 \eta Z(\omega)}{2 \pi \beta^2 E_0 n} \int \frac{dF_0}{n - \omega_r} d\omega_r.
\]

The integral and the impedance \( Z(\omega) \) are complex. The above equation gives the maximum complex impedance for which there is still stability.

To bring the above equation into a form more adapted for practical applications we introduce some normalizations [24]. We use the half-width frequency spread at half-height and relate it to the corresponding spread in momentum

\[
2S = -\eta \omega_0 \frac{\Delta p}{p},
\]

and introduce the reduced frequencies

\[
x = \frac{\omega_r - \omega_0}{S}, \quad x_1 = \frac{\omega - n \omega_0}{nS},
\]

the reduced distribution

\[
f_0(x) = 2 \pi S F_0(\omega_r) \quad \text{with} \quad \int f_0(x) dx = 1
\]

and use the average current

\[
I_0 = \frac{e N \omega_0}{2 \pi}.
\]

This gives for the stability condition

\[
1 = -\frac{e I_0 Z(\omega)/n}{2 \pi \beta^2 E_0 \eta (\Delta p/p)^2} \left[ \pi \frac{dF_0}{dx}(x_1) - i PV \int \frac{dF_0}{dx} dx \right]
\]

which we can write in the form

\[
V' + U' = \frac{e I_0}{2 \pi \beta^2 E_0 \eta (\Delta p/p)^2} \left( \frac{Z_r}{n} + \frac{Z_i}{n} \right) = -\left( \pi \frac{dF_0}{dx}(x_1) - i PV \int \frac{dF_0}{dx} dx \right)^{-1}.
\]

The Factor on the left hand side contains now all the global beam parameters and the impedance while the right hand side depends on the form of the distribution only.

We can now plot the above relation in a complex plane and obtain the stability diagram. This is done by calculating the \( U' \) and \( V' \) as a function of \( x_1 \), i.e. of the frequency \( \omega \). The response of the beam to an excitation is only large when \( \omega \) is close to a harmonic \( n \) of the revolution frequency \( \omega_0 \). The quantity \( U' + iV' \) representing the stability diagram is proportional to the inverse response. If the frequency \( \omega \) approaches the harmonic of the revolution frequency a point on the stability diagram comes down from infinity, moves around the origin (while \( \omega \) goes through \( n \omega_0 \)) and moves then again...
to large values. The area surrounded by the stability diagram contains values of $U'$ and $V'$ for which the beam is stable. For a given beam current, energy, frequency spread and other beam parameters can therefore determine the maximum complex stability which still provides stability. In Fig. 20 the stability diagrams of a beam with a Gaussian distribution is plotted. By taking such diagrams for different distributions one can approximate them by a circle with a radius of about 0.6. This gives a very simple stability condition

$$|Z/n| \leq 0.6 \frac{2\pi \beta^2 E_0 \gamma (\Delta p/p)^2}{eI_0}$$

which is called the Keil-Schnell criterion [25].

### 5.9 Discussion

We started by calculating the response of an un-bunched beam to a longitudinal harmonic excitation with frequency $\omega$. We found that this response has a part for which the velocity is in phase with the external acceleration. It can therefore absorb energy and lead to damping, called Landau damping. This result is not so easy to understand physically. A more detailed analysis indicates (shown for the transverse case) that the excitation leads to growing amplitudes of the few particles for which the natural revolution frequency $\omega_r$ (or a harmonic of it) is close the excitation frequency $\omega$. The absorbed energy goes into the growing amplitudes of these particles.

We later considered the response of the particles in the beam to a pulse excitation which established an energy modulation around the ring. The following free oscillation executed by the particle with different frequencies lead to a decay of the coherent energy or current modulation. However the amplitude of the individual particles will stay constant. This response to a pulse excitation (Green function) is easier to understand than the response to a harmonic excitation (transfer function) but not so easy to check experimentally.

To compare this beam response to a longitudinal excitation with the transverse case we find an important difference. In the transfer or Green function the distribution $f_0$ appears for the transverse response while its derivative $df_0/d\omega$ is the relevant quantity.
for the longitudinal case. In the latter case the excitation affects the quantity $\Delta E$ which determines the spread while in the transverse case the distribution is given by parameters which are not affected by the excitation. The Landau damping of bunched beams can be treated in a similar way as the longitudinal Landau damping of the un-bunched beam [2, 26, 27].

The reduction of the coherent amplitude after a pulse excitation and the resistive response of the beam to a harmonic excitation are the basis of Landau damping, but they represent two extreme cases. In the first case the continuous excitation leads to very large amplitudes of a few particles, in the second case most particles obtain a finite amplitude by the pulse. In reality Landau damping avoids an instability at infinitesimal amplitudes. Let us assume, due to statistical fluctuation of the particles, a harmonic current modulation is established. It will induce a current in the beam surroundings which will in turn apply a voltage to the beam. Taking the picture of a harmonic excitation the energy provided by this voltage is absorbed by some gain in amplitude of a few particles and the modulation does not grow further. From the view of the response to a pulse excitation the frequency spread will smear out established modulation faster than the induced voltage can enhance it. Since all this happens at infinitesimally small amplitudes a beam stabilized by Landau damping does not grow in energy spread but can be kept stable for days as experience shows.

References


Abstract
The Coulomb forces between the charged particles of a high-intensity beam in an accelerator create a self-field which acts on the particles inside the beam like a distributed lens, defocusing in both transverse planes. A beam moving with speed \( v \) is accompanied by a magnetic field which partially cancels the electrostatic defocusing effect, with complete cancellation at \( c \), the speed of light. The effect of this ‘direct space charge’ is evaluated for transport lines and synchrotrons where the number of betatron oscillations per machine turn, \( Q \), is reduced by \( \Delta Q \). In a real accelerator, the beam is also influenced by the environment (beam pipe, magnets, etc.) which generates ‘indirect’ space charge effects. For a smooth and perfectly conducting wall, they can easily be evaluated by introducing image charges and currents. These ‘image effects’ do not cancel when \( v \) approaches \( c \), thus they become dominant for high-energy synchrotrons. Each particle in the beam has its particular incoherent tune \( Q \) and incoherent tune shift \( \Delta Q \). If the beam moves as a whole, so the centre of mass executes a coherent betatron oscillation, image charges and currents caused by the beam pipe move as well, leading to coherent tune shifts which also depend on the beam intensity. For a realistic beam, the incoherent tune of a given particle depends on its betatron amplitude and position in the bunch, leading to a tune spread (rather than a tune shift) which occupies a large area in the tune diagram of low-energy machines. The ‘space-charge limit’ of a synchrotron may be overcome by increasing its injection energy; various systems which have actually been built are presented.

1. INTRODUCTION
Space charge is the simplest and most fundamental of the collective effects whose impact generally is proportional to the beam intensity. The charge and current of the beam create self-fields and image fields which alter its dynamic behaviour and influence the single-particle motion as well as coherent oscillations of the beam as a whole. The paper concentrates on changes of the betatron tune in synchrotrons for simple cases which can readily be calculated; the emphasis is on low-energy (more precisely: low-\( \gamma \)) machines where space charge represents a fundamental intensity limitation. The lecture this paper is based upon is intended for physicists, engineers and senior technicians who are not familiar with collective effects and is meant as an introduction to the subject. The reader is assumed to be familiar with the basic laws of classical electrodynamics; SI units are used throughout. In many respects, the paper is inspired by earlier CAS lectures [1–3]. Reference [2] also covers the particular intricacies of high-energy machines (which in this context include all electron synchrotrons) and more complex cases. Methods for measuring space-charge tune shifts/spreads in synchrotrons are presented in Ref. [3].
2. DIRECT SPACE CHARGE (SELF-FIELDS)

Consider two particles of equal charge +e (Fig. 1). At rest, they experience the repulsion due to the Coulomb force. When travelling with velocity $v = \beta c$, they represent two parallel currents $I = ve$ which attract each other by the effect of their magnetic fields. The diagram shows that the overall effect is still repulsive but decreases with the speed; special relativity implies that the forces become equal at the speed of light and thus cancel.

![Fig. 1: Coulomb repulsion and magnetic attraction between two particles of equal charge, at rest and travelling](image)

Now consider many charged particles travelling in an unbunched beam with circular cross section (Fig. 2). The Coulomb repulsion pushes the test particle away from the beam centre; the overall force is zero in the beam centre and increases towards the edge. This behaviour applies also to the test particle in a travelling beam, represented by parallel currents, except that the magnetic force vector is directed towards the beam centre.

![Fig. 2: Electrostatic and magnetic forces on a test particle within a cylindrical unbunched beam](image)

2.1. The self-field

An unbunched beam of circular cross section (radius $a$) and uniform charge density $\eta$ [Cb/m$^3$] moves with constant velocity $v = \beta c$. It has a line charge density (charge per unit length [Cb/m]) of $\lambda = \pi a^2 \eta$, a current density [A/m$^2$] of $J = \beta c \eta$, and a total current of $I = \beta c \lambda$. Figure 3 represents the beam; in the following, the electric $\vec{E}$ and magnetic $\vec{B}$ fields on the surface of a cylinder with radius $r < a$ are calculated (using polar coordinates $r$, $\phi$). Due to symmetry, the electric field has just a radial component ($E_r$), while the magnetic field lines are just circles around the cylinder ($B_\theta$ component only).

$E_r$ is calculated from the Maxwell equation

$$\text{div} \, \vec{E} = \frac{\eta}{\varepsilon_0}$$

and its integral form, Gauss’ law (integration over volume and surface of cylinder with radius $r$ and length $l$)

$$\iiint \text{div} \, \vec{E} dV = \iint \vec{E} \, d\vec{S}$$

306
where $dV$ is a volume element inside the cylinder, and $d\vec{s}$ an element of its surface (Fig. 3), yielding

$$\pi lr^2 \frac{n}{\varepsilon_0} = 2\pi lr E_r$$

(3)

from which one can derive the radial electric field

$$E_r = \frac{I}{2\pi \varepsilon_0 \beta c a^2} \frac{r}{a^2}.$$  

(4)

Similarly, the azimuthal magnetic field $B_\phi$ is determined from another Maxwell equation

$$\text{curl} \vec{B} = \mu_0 \vec{J}$$

(5)

and its integral form, Stokes’ law

$$\oint \vec{B} d\vec{s} = \iint \text{curl} \vec{B} d\vec{S},$$

(6)

where $d\vec{s}$ is a path element in the cross section along the circle with radius $r$, and $d\vec{S}$ a surface element within this circle (Fig. 3). The integrals (6) over the cylinder of radius $r$ and length $l$ result in

$$2\pi r B_\phi = \mu_0 \pi r^2 \beta \varepsilon \eta$$

(7)

yielding the magnetic field

$$B_\phi = \frac{I}{2\pi \varepsilon_0 \beta^2 a^2} \frac{r}{a^2}.$$  

(8)

Indeed, both electric and magnetic fields vanish at $r = 0$, and both increase linearly with $r$ up to the edge of the cylinder ($r = a$).

2.2. The forces

These fields exert a force $\vec{F}$ on a test particles at radius $r$ (the magnetic force is sketched in Fig. 4) which is now calculated:

$$\vec{F} = e(\vec{E} + [\vec{v} \times \vec{B}])$$

(9)

For the geometry under consideration Eq. (9) simplifies to

$$F_r = e(E_r - v_\beta B_\phi)$$

(10)
indicating that the force vector has a purely radial component $F_r$. Inserting $E_r$ and $B_\phi$ from equations (4) and (8) one gets for the radial force on the test particle at radius $r$

$$F_r = \frac{eI}{2\pi\varepsilon_0\beta c} (1 - \beta^2) \frac{r}{a^2} = -\frac{eI}{2\pi\varepsilon_0\beta c} \frac{1}{\gamma^2 a^2} \frac{r}{a^2}.$$  \hspace{2cm} (11)

In the $(1 - \beta^2)$ term, the ‘1’ represents the electric force, and the $\beta^2$ the magnetic one, and indeed they cancel at $\beta = 1$; it is replaced by $1/\gamma^2$ in Eq. (11). Replacing $r$ by the transverse coordinates $x$, $y$ results in the horizontal ($F_x$) and vertical ($F_y$) forces which are linear in $x$ and $y$, respectively:

$$F_x = \frac{eI}{2\pi\varepsilon_0\gamma^2 a^2} x , \quad F_y = \frac{eI}{2\pi\varepsilon_0\gamma^2 a^2} y .$$  \hspace{2cm} (12)

It is instructive to compare the focusing effects of a quadrupole with the uniform high-intensity beam for which Eq. (12) applies. Figure 5 shows the force $F_x$ vs. $x$ for a (horizontally focusing) quadrupole, and for a space-charge dominated uniform beam. While the quadrupole is focusing in one and defocusing in the other plane, direct space charge leads to defocusing in both planes.

2.3. Beam transport with space charge

A FODO transport line is described by Hill’s equation

$$x'' + K(s) x = 0 ,$$  \hspace{2cm} (13)

where $K(s)$ denotes the normalised gradients (positive if focusing, negative if defocusing) of the quadrupole along the beam path $s$ (here for the horizontal plane). A perturbation term $K_{SC}(s)$ describes the (continuous) defocusing action of space charge:

$$x'' + (K(s) + K_{SC} (s)) x = 0 .$$  \hspace{2cm} (14)
**SPACE CHARGE**

\(K_{SC}\) is derived by expressing \(x''\) in terms of transverse acceleration \(d^2x/dt^2\) and thus of the force \(F_x\) [Eq. (12)]

\[
x'' = \frac{d^2x}{ds^2} = \frac{1}{\beta^2 c^2} \frac{d^2 x}{dt^2} = \frac{\dot{x}}{\beta^2 c^2} = \frac{1}{\beta^2 c^2} \frac{F_x}{\gamma m_0 \gamma c} = \frac{2 r_0 I}{e \alpha^2 \beta^3 \gamma^3 c} x,
\]

where \(r_0 = e^2/(4\pi \varepsilon_0 \gamma m_0 c^2)\) is the classical particle radius, \(1.54 \times 10^{-18}\) m for protons. Hill’s equation with space charge (for an unbunched beam with circular cross section and constant charge density) is then

\[
x'' + \left( K(s) - \frac{2 r_0 I}{e \alpha^2 \beta^3 \gamma^3 c} \right) x = 0.
\]

Note the negative sign of the space charge term, reducing the overall focusing of the FODO sequence.

### 2.4. Direct space charge in a synchrotron: Incoherent tune shift

The direct space charge leads to defocusing in either plane, and therefore one would expect that particles in a high-intensity beam will experience a lowering of their betatron tunes \(Q\) by \(\Delta Q\). The calculation below applies to the simplest (pretty unrealistic) case: unbunched beam, circular cross section everywhere in the accelerator, constant charge density.

Applying Eq. (13) to a synchrotron lattice will yield the unperturbed horizontal \(Q_\phi\), while Eq. (16) introduces a space-charge defocusing \(\Delta Q_s\) which is readily calculated [4] by integrating the weighted gradient errors around the circumference \(2\pi R\):

\[
\Delta Q_s = \frac{1}{4\pi} \int_0^{2\pi R} K_x(s) \beta_x(s) ds = \frac{1}{4\pi} \int_0^{2\pi R} K_{SC}(s) \beta_x(s) ds.
\]

Taking \(K_{SC}(s)\) from Eq. (16) yields

\[
\Delta Q_s = -\frac{1}{4\pi} \int_0^{2\pi R} \frac{2 r_0 I}{e \beta^3 \gamma^3 c} \frac{\beta_x(s)}{a^2} ds = -\frac{r_0 I R J}{e \beta^3 \gamma^3 c} \left( \frac{\beta_x(s)}{a^2(s)} \right).
\]

The term \(\left( \beta_x(s)/a^2(s) \right)\) (remember: \(a\) is the beam size) is just \(1/E_x\), the inverse of the horizontal emittance, and thus an invariant. Replacing \(I\) by \(N e/(2\pi R)\) (with \(N\) the number of particles in the accelerator) and extending to the vertical plane \((y)\) one gets for the direct space-charge tune shift

\[
\Delta Q_{s,y} = -\frac{r_0 N}{2\pi E_{s,y} \beta^3 \gamma^3}.
\]

\(E_{s,y}\) is the transverse emittance in either plane containing 100% of the particles. The main features are:

- the tune shift is proportional to the intensity;
- it scales with \(1/\gamma^3\), so negligible for all electron synchrotrons, and very small for proton synchrotrons beyond \(\sim 10\) GeV;
- it does not depend on the machine radius \(R\).

### 2.5. Direct space charge for a uniform beam with elliptic beam cross section

For this geometry (Fig. 6), Maxwell’s Eqs. (1) and (5) are satisfied by the electric and magnetic fields \(\vec{E}\) and \(\vec{B}\)

\[
\vec{E} = \frac{I}{\pi \varepsilon_0 (a + b) \beta c} \left( \frac{x}{a} , \frac{y}{b} , 0 \right), \quad \vec{B} = \frac{\mu_0 I}{\pi (a + b)} \left( -\frac{y}{b} , \frac{x}{a} , 0 \right).
\]

A particle at position \((x, y)\) inside the beam will experience a force

\[
\vec{F} = e(\vec{E} + [\vec{v} \times \vec{B}]) = \frac{eI}{\pi \varepsilon_0 \beta c \gamma^2 (a + b)} \left( \frac{x}{a} , \frac{y}{b} , 0 \right).
\]
from which, following the reasoning behind Eq. (15), the space-charge defocusing term $K_{SC,Y}$ (here for the vertical plane)

$$K_{SC,Y} = -\frac{F_y}{m\gamma^3/\beta^2 e^2 y} = -\frac{4r_0I}{eb(a+b)\beta^3\gamma^3 e}\quad(22)$$

and the vertical tune shift

$$\Delta Q_y = \frac{1}{4\pi} \int_0^{2\pi R} K_{SC,Y}(s)\beta_y(s) ds = -\frac{2r_0I}{\epsilon_0\beta^3\gamma^3 E_y} \left\langle \frac{b}{a+b} \right\rangle \quad(23)$$

can be determined. One then approximates the averaging of term $\left\langle \frac{b}{(a+b)} \right\rangle$

$$\left\langle \frac{b}{a+b} \right\rangle \approx \frac{1}{\left\langle \frac{b}{a} \right\rangle + 1} = \frac{1}{\sqrt{\left\langle \beta_x \right\rangle E_x/\left\langle \beta_y \right\rangle E_y} + 1} \quad(24)$$

to obtain the vertical direct space charge tune shift of a uniform beam with elliptic cross section

$$\Delta Q_y \approx -\frac{r_0N}{2\pi E_y\beta^2\gamma^3/\left(1 + \sqrt{\beta_x E_x/\beta_y E_y}\right)} \quad(25)$$

Exchanging all $x'$s and $y'$s, one obtains the horizontal shift $\Delta Q_x$. Typical synchrotrons feature small vertical gap sizes in the dipoles and hence $E_y < E_x$, which in turn gives rise to $\Delta Q_y > \Delta Q_x$.

### 2.6. Direct space charge for a non-uniform beam

In the preceding sections only beams with constant charge density $\eta$ are considered. In what follows, a simple — and rather realistic — example for a non-uniform distribution, namely a bi-Gaussian density in the circular beam cross section, is analysed. The distribution is

$$\eta(r) = \frac{I}{2\pi \beta \sigma^2} e^{-r^2/2\beta^2}, \quad r = \sqrt{x^2 + y^2} \quad(26)$$

where $\sigma$ denotes the r.m.s. value of the distribution projected on the $x$- (or $y$-) axis. Maxwell’s equations (1) and (5) are satisfied by the following expressions for $E_r$ and $B_\phi$ (the other field components vanish because of circular symmetry) as can be readily verified:

$$E_r = \frac{I}{2\pi \varepsilon_0 \beta \sigma^2} \frac{1}{r} \left(1 - e^{-r^2/2\beta^2}\right), \quad B_\phi = \frac{I}{2\pi \varepsilon_0 \beta e^2} \frac{1}{r} \left(1 - e^{-r^2/2\beta^2}\right) \quad(27)$$

This yields a radial force

$$F_r(r) = \frac{eI}{2\pi \varepsilon_0 \beta \gamma^2} \frac{1}{r} \left(1 - e^{-r^2/2\beta^2}\right), \quad(28)$$

$$K. S. CHINDL$$

Fig. 6: Cross section through an elliptic beam with constant charge density $\eta$
which is no longer linear in \(r\), thus the defocusing becomes betatron-amplitude dependent (Fig. 5, right). It is instructive to linearise the force for small \(r\) (near the beam centre)

\[
F(r) = \frac{eI}{2\pi \varepsilon_0 \beta \gamma^2 e r} \left(1 - 1 + \frac{r^2}{2\sigma^2} - \ldots\right) \approx \frac{eI}{2\pi \varepsilon_0 \beta \gamma^2 e 2\sigma^2} r,
\]

resulting in a small-amplitude (vertical) tune shift of

\[
\Delta Q_y = \frac{r_0 N \beta_y}{e\varepsilon_0 \beta \gamma^3} \left(\frac{\beta_y}{2\sigma^2}\right) = -\frac{r_0 N}{2\pi \beta \gamma^3} \frac{2}{E_y} \text{ for } r \ll \sigma \quad (30)
\]

with \(E_y = 4\sigma^2/\beta_y\), the 95% emittance.

This is twice the figure of a uniform beam of the same cross-sectional size and intensity. The important conclusions are: (i) for given emittances and intensity, a uniformly distributed beam gives the smallest possible direct space-charge tune shift; (ii) a — more realistic — non-uniform transverse charge distribution features a betatron amplitude-dependent de-tuning, hence a tune spread rather than a shift. In beams with transverse density profiles resembling a Gaussian which are typical for hadron cooler rings, small-amplitude particles suffer the largest tune depression.

3. TUNE SHIFT WITH WALL EFFECTS

In Section 2, the impact of the self-field created by the beam alone was investigated, resulting in ‘direct’ space-charge tune shifts and spreads. A real accelerator, however, consists of a vacuum pipe, accelerator gaps, magnets, beam diagnostics, and a high-intensity beam induces surface charges or currents into this environment that act back on the beam, possibly resulting in an ‘indirect’ space-charge tune shift. The basic effect is demonstrated below for a very simple geometry (Fig. 7) where the beam is best represented by a line charge \(\lambda\) of infinite length.

3.1. ‘Incoherent’ tune shift due to conductive walls

The boundary condition on a perfectly conducting plate \((E_{\parallel} = 0)\) is satisfied by introducing ‘image’ line charges of negative sign as shown in Fig. 7 (right). The beam, whose barycenter is halfway between two parallel conducting plates, gives rise to an infinite number of image line charges of alternating sign at positions \(2h, 4h, 6h, \ldots, -2h, -4h, \ldots\) (see Fig. 7 left). A test line charge positioned vertically off-centred by \(y\), but still inside the beam, will be subject to an infinite sum of electric fields; is it non-zero?

The electric field generated by a line charge \(\lambda\) at distance \(d\) (Fig. 8) is

\[
E_y = \frac{\lambda}{2\pi \varepsilon_0} \frac{1}{d} \quad (31)
\]

Applying Eq. (31) to the first pair of image line charges, positioned at \(2h, -2h\), yields the vertical field at point \(y\) generated by the two images, \(E_{\text{ly}}\); more generally, the \(n^{th}\) pair of line charges, positioned at \(2nh, -2nh\), results in \(E_{\text{ny}}\) :

\[
E_{\text{ly}} = \frac{\lambda}{2\pi \varepsilon_0} \left(\frac{1}{2h - y} - \frac{1}{2h + y}\right), \\
E_{\text{ny}} = (-1)^n \frac{\lambda}{2\pi \varepsilon_0} \left(\frac{1}{2nh + y} - \frac{1}{2nh - y}\right) = (-1)^n \frac{\lambda}{4\pi \varepsilon_0} \frac{y}{n^2h^2} \quad (32)
\]
Fig. 7: A line charge $\lambda$, representing the particle beam, between parallel conducting plates of distance $2h$ (left). Electric field components parallel to the conducting plate have to be zero: this is achieved by introducing negative image line charges (right).

Fig. 8: Electric field created at distance $d$ by a line charge extending to infinity

The total vertical electric field $E_{iy}$ is readily obtained by summing $E_{iny}$ from $n = 1$ to infinity

$$E_{iy} = \sum_{n=1}^{\infty} E_{iny} = \frac{\lambda}{4\pi \varepsilon_0 h^2} \sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} y = \frac{\lambda}{4\pi \varepsilon_0 h^2} \frac{\pi^2}{12} y \quad (33)$$

by making use of the expression

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} = \frac{\pi^2}{12} \quad (34)$$

The vertical electric field due to the images is linear in $y$ (vertical distance of the test particle). Between the parallel plates, there are no image charges, therefore Eq. (1) simplifies to $\text{div} \vec{E}_i = 0$ from which the component $E_{ix}$ as well as the forces in both $y$ and $x$ directions are derived:

$$\text{div} \vec{E}_i = \frac{\partial E_{ix}}{\partial x} + \frac{\partial E_{iy}}{\partial y} = 0 \quad \implies \quad E_{ix} = - \frac{\lambda}{4\pi \varepsilon_0 h^2} \frac{\pi^2}{12} x$$

$$F_{iy} = \frac{e\lambda}{\pi \varepsilon_0 h^2} \frac{\pi^2}{48} y, \quad F_{ix} = -\frac{e\lambda}{\pi \varepsilon_0 h^2} \frac{\pi^2}{48} x \quad (35)$$

Following the procedure outlined in Eqs. (15) and (17) which links the force to the $Q$-shift, one obtains — together with the direct space-charge effect [Eq. (18)] — the total incoherent tune shift of a round beam between parallel conducting walls

$$\Delta Q_x = -\frac{2r_0 I R(\beta_y)}{e\varepsilon_0 \beta^3 \gamma} \left( \frac{1}{2(a^2)\gamma^2} - \frac{\pi^2}{48h^2} \right) \text{ direct} - \text{ image} \quad (36)$$

$$\Delta Q_y = -\frac{2r_0 I R(\beta_y)}{e\varepsilon_0 \beta^3 \gamma} \left( \frac{1}{2(a^2)\gamma^2} + \frac{\pi^2}{48h^2} \right).$$
Inspection of Eq. (36) reveals a few salient features:

- the electric image field is vertically defocusing, but horizontally focusing (sign of image term changes), which by the way is not just a feature of this particular geometry, but is typical for most synchrotrons with their rather flattish vacuum pipes;
- the field is larger for small chamber height \( h \);
- Image effects decrease with \( 1/\gamma \), much slower than the direct space-charge term \( (1/\gamma^3) \), and thus are of some concern for electron and high-energy proton machines.

The conductive beam pipe leads to mirror (line) charges as presented above. But the beam travels and represents a d.c. beam current accompanied by an azimuthal d.c. magnetic field which is not shielded by the beam pipe. The magnetic field lines are influenced by ferromagnetic boundaries such as magnets; their effect may be represented by mirror currents, from which an incoherent tune shift due to magnetic images can be calculated (not done here, but the magnetic image coefficients are compiled in Table 1, see later).

### 3.2. ‘Coherent’ tune shift

The term ‘incoherent’ has sneakied into the last paragraph, and now it says: ‘coherent’. What do these terms mean?

- **Incoherent** motion: the beam consists of many particles, each of which moves inside the beam with its individual betatron amplitude, phase, and even tune Q (under the influence of direct space charge). Amplitude and phase are distributed at random over all particles. Except for low-intensity beams, an outside observer (using a position monitor at some azimuth in the synchrotron) does not see any of this random betatron motion. The beam and its centre of gravity — and thus the source of the direct space-charge field — do not move (static beam, Fig. 9).

- **Coherent** motion: A static beam is given a transverse fast deflection (< 1 turn) and starts to perform betatron oscillations as a whole (Fig. 10). This is readily observed by a position monitor. Note that the source of the direct space charge is now moving: individual particles still continue their incoherent motion around the common coherent trajectory and still experience their incoherent tune shifts as well.

![Fig. 9: Incoherent betatron motion of a test particle inside a static beam with its centre of mass at rest.](image)

![Fig. 10: Coherent betatron motion of the whole beam after having experienced a transverse kick.](image)

In the presence of the accelerator environment, the coherently oscillating beam induces image charges/currents which are oscillating as well. This leads to a **coherent tune shift** as is demonstrated in what follows for a simple geometry: a round, perfectly-conducting beam pipe. Figure 11 shows a beam with line charge \( \lambda \) and radius \( a \) performing coherent oscillations of its centre of mass \( \bar{x} \) inside the round beam pipe with radius \( \rho \) \((a \ll \rho)\). The displaced line charge \( \lambda \) induces surface charges on the inside of the beam pipe which can be represented by an image line charge \(-\lambda\) at distance \( b \), where

\[
b = \frac{\ell^2}{\bar{x}} \quad (37)
\]
The image charge pulls the beam away from the centre of the beam pipe: its effect is defocusing. The horizontal electric image field $E_{ix}$ and the horizontal force $F_{ix}$ are

$$E_{ix}(\vec{x}) = \frac{\lambda}{2\pi \varepsilon_0} \frac{1}{b - \vec{x}} \approx \frac{\lambda}{2\pi \varepsilon_0} \frac{1}{b} = \frac{\lambda}{2\pi \varepsilon_0} \frac{1}{\rho^2},$$

$$F_{ix}(\vec{x}) = \frac{e\lambda}{2\pi \varepsilon_0} \frac{1}{\rho^2}.$$  \hspace{1cm} (38)

For symmetry reasons with this particular geometry, the vertical field and force, $E_{iy}$ and $F_{iy}$, are the same as the horizontal ones, and one obtains a coherent tune shift of

$$\Delta Q_{x,y,coh.} = -\frac{r_0 R(\beta_{x,y}) I}{\varepsilon_0 c \beta^2 \gamma p^2} = -\frac{r_0 (\beta_{x,y})}{2\pi \beta^2 \gamma p^2} N.$$ \hspace{1cm} (39)

A few salient features of the coherent tune shift:

- The force is linear in $\vec{x}$, so there is a coherent tune shift (for the unbunched beam under scrutiny).
- The $1/\gamma$ dependence of the tune shift stems from the fact that the charged particles induce the electrostatic field and thus generate a force proportional to their number, but independent of their mass, whereas the deflection of the beam by this force is inversely proportional to their mass $m_0 \gamma$.
- The coherent tune shift is never positive.
- Note that a perfectly conducting beam pipe has been assumed here, for simplicity. The effects of a thin vacuum chamber with finite conductivity are more subtle [2].

### 3.3. The ‘Laslett’ coefficients

A rather realistic accelerator scenario is shown in Fig. 12: a beam with elliptic cross section travels with speed $\beta c$ through an elliptic vacuum chamber (ideal conductor) and between the ferromagnetic yokes of bending magnets. It is unbunched and has a uniform charge density (which is a less realistic assumption). For this and even simpler geometries — some of which are covered here — the incoherent and coherent tune shifts can be expressed in terms of the ‘Laslett’ coefficients [5] ($\varepsilon$ for incoherent, $\xi$ for coherent). The formulae below [Eq. (40)] are given for the vertical plane $y$; just replace $y$ by $x$ to get the horizontal results. The Laslett coefficients for simple geometries are compiled in Table 1.

$$\Delta Q_{y,inc.} = -\frac{N r_0 (\beta_{y})}{\pi \beta^2 \gamma} \left( \frac{\varepsilon_{y,0}^\gamma}{\beta^2 \gamma^2} + \frac{\varepsilon_{y,0}^\gamma}{h^2} \right) + \left( \beta^2 \frac{\xi_{y}^\gamma}{g^2} \right),$$ \hspace{1cm} (40)

$$\Delta Q_{y,coh.} = -\frac{N r_0 (\beta_{y})}{\pi \beta^2 \gamma} \left( \frac{\xi_{y}^\gamma}{h^2} + \beta^2 \frac{\xi_{y}^\gamma}{g^2} \right).$$
Table 1: Laslett coefficients for incoherent (\(\varepsilon\)) and coherent (\(\xi\)) tune shifts, for simple geometries

<table>
<thead>
<tr>
<th>Laslett coefficients</th>
<th>Circular ((a = b, w = h))</th>
<th>Elliptical ((e.g., w = 2h))</th>
<th>Parallel plates ((h/w = 0))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\varepsilon^x_0)</td>
<td>1/2</td>
<td>(\frac{b}{a(a+b)})</td>
<td>0</td>
</tr>
<tr>
<td>(\varepsilon^y_0)</td>
<td>1/2</td>
<td>(\frac{b}{a+b})</td>
<td>0</td>
</tr>
<tr>
<td>(\varepsilon^x_1)</td>
<td>0</td>
<td>(-0.172)</td>
<td>(-0.206)</td>
</tr>
<tr>
<td>(\varepsilon^y_1)</td>
<td>0</td>
<td>0.172</td>
<td>0.206</td>
</tr>
<tr>
<td>(\xi^x_1)</td>
<td>1/2</td>
<td>0.083</td>
<td>0</td>
</tr>
<tr>
<td>(\xi^y_1)</td>
<td>1/2</td>
<td>0.55</td>
<td>0.617((\pi^2/16))</td>
</tr>
<tr>
<td>(\varepsilon^x_2)</td>
<td>(-0.411(-\pi^2/24))</td>
<td>(-0.411)</td>
<td>(-0.411)</td>
</tr>
<tr>
<td>(\varepsilon^y_2)</td>
<td>0.411((\pi^2/24))</td>
<td>0.411</td>
<td>0.411</td>
</tr>
<tr>
<td>(\xi^x_2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(\xi^y_2)</td>
<td>0.617((\pi^2/16))</td>
<td>0.617</td>
<td>0.617</td>
</tr>
</tbody>
</table>

Fig. 12: A beam of elliptic cross section in an elliptic beam pipe and inside a ferromagnetic boundary (dipole magnet); the symbols are used in Table 1

For the elliptic geometry (see Fig. 12) with the vertical beam and vacuum chamber dimensions smaller than horizontally, which is quite common in realistic synchrotrons, most coefficients are larger vertically. Note also that the coherent coefficients \(\xi\) are all positive or zero, that is the coherent tune shifts are negative (defocusing) or zero for all geometries, but never focusing.

4. TUNE SHIFT/SPREAD IN BUNCHEO BEAMS

The arguments and formulae presented in the preceding sections apply to unbunched beams only. There is a class of accelerators where this exceptional situation indeed prevails: hadron (mostly ions or antiprotons) cooler rings or accumulators, where not too small intensities are cooled down to very small transverse emittances, mostly by electron cooling. As the direct space charge tune shift is proportional to the ratio intensity/emittance [the term \(N/E\) in Eq. (19)], values of \(\Delta Q \sim 0.1\) to 0.2 can be reached in these cooling rings.

4.1. Comparison between unbunched and bunched beams

The majority of low-energy machines were built to accelerate particles, thus feature bunched beams. The fact that \(N\) particles are no longer evenly distributed around the ring but are lumped in bunches aggravates space charge effects. This is illustrated in Fig. 13 where some basic features of unbunched beams (left column) are compared with bunched beams (right column).

The coasting beam is described by a constant line density \(\lambda_0\), a band in the longitudinal phase plane of height \(\Delta p\), and (for a 3-D uniform particle density) a \(Q\)-shift \((-0.25\) in this example). If this situation were realistic, one could readily correct this \(Q\)-shift by increasing the external betatron tune by 0.25, and space charge would not be an issue.

For a bunched beam (right-hand column in Fig. 13), the line density varies between 0 and \(\lambda\).
with the average value $\bar{\lambda}$. Note that in low energy hadron synchrotrons, the bunch length is a few to many meters, whereas the transverse beam size is a few cm, so to a good approximation, a bunch consists of many longitudinal slices which behave independently of each other: the space charge tune shift of each slice depends on the local line density $\lambda(s)$. This causes a tune spread along the bunch, for both incoherent (direct space charge + images) and coherent tune shifts. The latter effect has been put forward to explain [6] the rather fast decoherence of a coherently oscillating high-intensity proton beam at injection into the CERN SPS at 26 GeV/c.

What happens to a single particle in the bunch? Each of them performs synchrotron oscillations which periodically vary its relative position $s$ and its momentum deviation $\Delta p$ in the longitudinal phase plane, with a period of typically 100 $\mu$s to a ms, lasting many machine revolutions. By this motion, the particle is driven back and forth in the bunch and periodically ‘feels’ different local line densities. The incoherent tune shift during a synchrotron oscillation period is shown in Fig. 13 for particles a,b,c with increasing synchrotron oscillation amplitudes: particle a stays near the bunch centre with its large line density and experiences large $Q$-shifts but small $Q$-variations during the oscillation, while particle c with its large amplitude synchrotron oscillation undergoes strong variations in $Q$ (from almost zero to −0.5 in this example). It is no longer possible to compensate space charge tune ‘shifts’ by adjusting the external focusing.

![Diagram of beam with tune shift](image)

Fig. 13: Comparison between a coasting (unbunched) (left column) and a bunched beam (right column): particles in the beam pipe, line density $\lambda(s)$, longitudinal phase plane, $Q$-‘shift’ (from top to bottom).
4.2. Transverse space charge with bunched beams: a practical example [7]

The repercussions of bunched beams are illustrated with the CERN PS Booster, a space-charge dominated machine built for a tenfold increase of the CERN PS beam intensity. Figure 14 shows the tune diagram with all the low-order betatron resonances which are deemed to cause undesirable beam losses and beam emittance blow-up. The big ‘necktie’-shaped area depicts the tune spread at 50 MeV (injection) with a total $\Delta Q_y$ of about $0.6$, which covers several dangerous non-linear resonances. Note that individual particles do not stay at constant tunes $Q_x, Q_y$ but are moved up and down in the necktie area, so they repetitively cross the low-order stopbands, and some of them will be lost. Fortunately $\Delta Q$ shrinks as the beam is accelerated by virtue of the $1/\beta^2 \gamma^2$ term in Eq. (19); to profit from this feature, the machine tune is moved into an area clear of these stop-bands as soon as possible. Note that resonances above order 4 ($m Q_x + n Q_y = p$ with $|m| + |n| > 4$) can be ignored in this case because of the short acceleration time of 0.4 sec.

![Fig. 14: Example for a space-charge limited synchrotron: betatron tune diagram and areas covered by direct space-charge tune spreads at injection, intermediate, and extraction energies for the CERN Proton Synchrotron Booster. During acceleration, space charge gets weaker and the 'necktie' area shrinks, enabling the external machine tunes to move the 'necktie' to an area clear of betatron resonances.](image)

4.3. Incoherent tune shift: a practical formula

Starting with Eq. (25) for the vertical direct space charge tune shift of a uniformly charged beam with an elliptic cross section, one develops a more realistic formula for synchrotrons dominated by direct space charge by introducing various ‘form factors’. Replacing $E_x$ by $a^2/\beta_x$, $E_y$ by $b^2/\beta_y$, one gets

$$\Delta Q_y = -\frac{\tau_0}{\pi} \left( \frac{q^2}{A} \right) \frac{N F_y}{B_y} \left\langle \frac{\beta_y}{B(a + b)} \right\rangle . \quad (41)$$

Evaluation of the average term and using $\langle \beta \rangle = R/Q$, the ‘smooth approximation’, yields

$$\left\langle \frac{\beta_y}{B(a + b)} \right\rangle = \left\langle \frac{\beta_y}{b^2(1 + \frac{1}{\gamma})} \right\rangle \approx \frac{1}{E_y \left(1 + \frac{E_x Q_x}{E_y Q_y} \right)} \quad (42)$$
and finally a formula for the incoherent tune shifts in either plane,

\[
\Delta Q_{x,y} = -\frac{r_0}{\pi} \left( \frac{q^2}{A} \right) \frac{N}{\beta\gamma^3} \frac{F_{x,y}G_{x,y}}{B_1} \left( 1 + \sqrt{\frac{E_{x,y}^2}{E_{x,y}^2 + q^2\beta^2\gamma^3}} \right)
\]

by inserting the following form factors:

- \( F_{x,y} \): form factor derived from Laslett’s image coefficients for incoherent tune shifts, \( \varepsilon^{x,y} \) (Table 1)
- \( G_{x,y} \): form factor depending on the particle distribution in the transverse \( x-y \) plane:
  - \( G=1 \) uniform (for \( E_{x,y} \) containing 100% of the beam);
  - \( G=2 \) Gaussian (for \( E_{x,y} \) containing 95% of the beam) (see Eq. 30);
- \( B_1 \): Bunching factor, \( B_1 = \tilde{\lambda}/\hat{\lambda} \), mean/peak line density, 1 for unbunched, < 1 for bunched beams.

The term \( (q^2/A) \), where \( q \) is the charge state and \( A \) the mass of ions, witnesses the strong space-charge forces prevailing in beams of high-charge-state ions at even low intensity \( N \) (number of ions in beam).

4.4. How to overcome the ‘space-charge limit’?

In Fig. 14, an example of a machine working in a heavy space-charge regime is presented. Incoherent tune shifts of beyond 0.5 can barely be tolerated without excessive particle loss, so for a given machine transverse acceptance, there is a hard intensity limit, the space-charge limit. What are the ways to overcome it? The essential parameters determining the tune shift are shown in the formula

\[
\Delta Q \sim \frac{N}{E_y \beta^2 \gamma^3}.
\]

Assume now that a large proton synchrotron is limited in \( N \) because \( \Delta Q \) reaches values around 0.5 when filling the (vertical) acceptance \( E_y \) which cannot be changed without rebuilding the machine. One sees that the only way to increase \( N \) is then to raise \( \beta\gamma^3 \), that is the injection energy. This has indeed been done in the past in several laboratories, in the following ways:

- Make the linear accelerator (Linac) longer (FNAL): By adding more tanks, the proton energy was raised from 200 to 400 MeV, yielding a potential gain factor of 2.6 in \( \beta\gamma^3 \) and thus in the limit intensity the downstream synchrotron (in fact, the FNAL Booster) can digest: almost a factor 2 was achieved (Fig. 15).
- Add a small Booster synchrotron to raise the injection energy of the main ring. This was the strategy adopted by CERN (adding a four-ring Booster to the PS), and by BNL where a fast-cycling Booster was added to fill the AGS. Note that the circumference of the Booster has to be much smaller (factor 4 in both cases) to enable filling of the main ring. In both cases, the limiting intensity indeed was boosted dramatically, albeit less than the calculated factor (Fig. 16, Table 2).

<table>
<thead>
<tr>
<th>Linac (MeV)</th>
<th>Booster (GeV)</th>
<th>( N = R/r )</th>
<th>Potential gain in ( N ) (by ( \beta^2 \gamma^3 ))</th>
<th>Gain achieved</th>
</tr>
</thead>
<tbody>
<tr>
<td>CERN PS</td>
<td>50</td>
<td>1</td>
<td>4 (rings)</td>
<td>59</td>
</tr>
<tr>
<td>BNL AGS</td>
<td>200</td>
<td>1.5</td>
<td>4 (batches)</td>
<td>26</td>
</tr>
</tbody>
</table>
Space charge

Fig. 15: Increasing the Linac energy at FNAL raises the limit intensity of the Booster synchrotron by almost a factor 2.

Fig. 16: Adding a smaller Booster synchrotron (1/4 of main ring circumference) boosts the limiting proton intensity: At CERN a four-ring, slow-cycling Booster fills the PS in one batch, whereas a fast-cycling Booster in BNL fills the AGS in four batches.
References


Abstract

What are the mechanisms leading to beam instabilities in circular accelerators? This lecture, intended for physicists and engineers with little or no knowledge on the subject, tries to answer this question. The physical concepts are presented for a few typical examples rather than attempting an exhaustive overview of the many types of coherent collective effects, impossible anyway in a one-hour lecture. As they are somewhat simpler to describe, the lecture concentrates on longitudinal instabilities.

1. BASIC MECHANISM DRIVING AN INSTABILITY

Assume a bunched beam circulating in a synchrotron. The bunches will induce electromagnetic fields in the beam pipe and give rise to image or wall currents since the beam pipe itself is a conductor. In turn, these currents generate an electromagnetic field that acts back on the beam. Under certain conditions – the discussion of which is the very aim of this lecture – the beam may become unstable.

A circulating bunch, corresponding to an a.c. beam current \( I_B \), induces image charges on the inner surface of the vacuum chamber that circulate with the bunch (Fig. 1). Thus they form an a.c. image or wall current \(-I_W\) of the same magnitude but opposite sign to the beam current.

The vacuum chamber has a finite conductivity and, moreover, changes its shape, cross-section, etc. along the beam path, and therefore presents an impedance to this wall current. The impedance \( Z = Z_r + i Z_i \) can be resistive (real), capacitive or inductive (imaginary). Thus the wall current induces a voltage \( V \sim I_W Z \) which gives rise to a longitudinal electric field, which may act back (accelerating or decelerating) on the bunch. While \( Z \) depends on the geometry and material of the vacuum enclosure and on the exciting frequency, \( V \) is proportional to the wall current and thus the beam current: Instabilities are intensity dependent, and in general stronger at higher beam currents.

Instabilities are investigated with the following scheme:

**Step 1**: Start with a nominal, unperturbed particle distribution (i.e. longitudinal position, energy, density, etc.).

**Step 2**: Apply a small perturbation which has a simple form – called ‘mode’ by the experts – and determine forces acting back on the beam.
Step 3: Calculate how the pattern would change under the forces. If it disappears, the beam is stable; if it is self-sustaining or even increases, the beam is unstable.

2. NEGATIVE MASS INSTABILITY

2.1. Qualitative treatment
This instability is no longer a problem but can be used as a starting point to understand collective effects and how they grow. Consider an unbunched beam circulating in a synchrotron. Consider furthermore a perturbation of the line density $\lambda(s)$ which describes a very simple azimuthal modulation pattern, namely eight humps around the ring (Fig. 2). Will the humps increase or erode?

![Fig. 2: Azimuthal line density modulation (left); zooming in one modulation (right)](image)

Particle B in Fig. 2 has a larger charge density behind it, pushing it forward. Conversely, particle A will be slowed down by the hump of charge in front of it. Thus the self-force $F$ (which is proportional to $-d\lambda/ds$) increases the energy of particle B, and decreases the energy of particle A. The stability of this ‘mode’ now depends on the transition gamma $\gamma$:

Below transition energy ($\gamma < \gamma_t$): an increase in energy leads to an increase in revolution frequency $\omega_0$, so A and B are moving away from the hump which is eroding.

Above transition energy ($\gamma > \gamma_t$): when the energy increases, the revolution frequency decreases, so A and B move towards the summit enhancing the mountain.

So far, not a single formula was used. However, in what follows, the ‘negative mass’ instability is treated by a more quantitative and rigorous method, which is a complementary way to illustrate some of the concepts used in analysing instabilities.

2.2. Fields generated by the beam
An unbunched beam with modulated line density $\lambda(s)$ travels through a vacuum pipe as depicted in Fig. 3.

What is the longitudinal electrical field $E_z$ generated by the beam? This may be evaluated by applying Stokes’ law along the rectangular path in Fig. 3. The radial electrical field $E_r$ and the azimuthal magnetic field $B_\theta$ are evaluated in Eq. (1), featuring the well-known linear increase proportional to radius $r$ inside the beam, and being proportional to $1/r$ outside the beam.
Fig. 3: Electrical ($E$) and magnetic ($B$) fields induced into a vacuum pipe (radius $b$) by a beam (radius $a$) with longitudinal density modulation $\lambda(s)$

\[ E_r = \frac{e\lambda}{2\pi\varepsilon_0} \frac{1}{r} \quad B_\phi = \frac{\mu_0 e\lambda\beta c}{2\pi} \frac{1}{r} \quad r \geq a , \]  
\[ E_r = \frac{e\lambda}{2\pi\varepsilon_0} \frac{r}{a^2} \quad B_\phi = \frac{\mu_0 e\lambda\beta c}{2\pi} \frac{r}{a^2} \quad r \leq a . \]  

Applying Stokes’ law which relates a line integral to an integral over the surface enclosed by the line

\[ \int_{\text{LINE}} E \, dl = \frac{\partial}{\partial t} \int_{\text{SURFACE}} \vec{B} \, d\vec{a} = -\frac{\partial}{\partial t} \Delta s \int_0^b B_\phi \, dr \]  
\[ (2) \]

and using

\[ \frac{\partial \lambda}{\partial t} = \frac{\partial \lambda}{\partial s} \frac{ds}{dt} = \beta c \frac{\partial \lambda}{\partial s} \]  
\[ (3) \]

one gets for the longitudinal electrical field $E_s$ on the beam axis

\[ E_s = -\frac{e\gamma_0}{4\pi\varepsilon_0} (1 - \beta^2) \frac{\partial \lambda}{\partial s} + E_w = -\frac{e\gamma_0}{4\pi\varepsilon_0} \frac{1}{\gamma^2} \frac{\partial \lambda}{\partial s} + E_w , \]  
\[ (4) \]

where $E_w$ is the electrical field parallel to the vacuum pipe surface.

For a perfectly conducting wall, $E_w$ vanishes, and one is left with

\[ E_s = -\frac{e\gamma_0}{4\pi\varepsilon_0} \frac{1}{\gamma^2} \frac{\partial \lambda}{\partial s} , \]  
\[ (5) \]
However, the vacuum chamber generally has a finite resistance and, in particular, an inductance due to cross-section changes, small cavity-like structures, and the like. One may lump these in an inductance per unit length $L/2\pi R$ ($R$ radius of the synchrotron) from which one calculates $E_w$:

$$E_w = -\frac{L}{2\pi R} \frac{dI_w}{dt} = \frac{L}{2\pi R} e^\beta c \frac{\partial \lambda}{\partial t} = \frac{L}{2\pi R} e^\beta c^2 \frac{\partial \lambda}{\partial s} \tag{6}$$

and the voltage per turn $U_s$ along the beam path due to the inductive wall

$$U_s = e^\beta c R \omega_0 L \frac{\partial \lambda}{\partial s} \tag{7}$$

### 2.3. Fields acting back on the beam

The beam line density $\lambda(\theta)$ (with the machine azimuth $\theta = s/R$) and the corresponding instantaneous current $I(\theta)$ are basically d.c. ($\lambda_0$, $I_0$) with a small a.c. component of amplitude $\lambda_1$, $I_1$, as given below:

$$\begin{align*}
\lambda &= \lambda_0 + \lambda_1 e^{i(n \theta - \Omega t)} \quad \text{instantaneous density}\\
I &= I_0 + I_1 e^{i(n \theta - \Omega t)} \quad \text{instantaneous current} \tag{8}
\end{align*}$$

$\lambda(\theta)$ features $n$ humps ($n = 8$ in the previous example). An antenna in the wall at azimuthal position $\theta$ would pick up the pattern rotating with angular frequency $\Omega = n\omega_0$, and it would stay like that for a low-intensity beam. However, a non-vanishing a.c. component induces a voltage per turn of

$$U_s = -I_1 e^{i(n \theta - \Omega t)} Z(\Omega) \tag{9}$$

voltage per turn (small) a.c. component longitudinal impedance
due to the longitudinal impedance $Z$ which in general depends on $\Omega$.

The trick to computing the effect of the fields acting back on the beam is to postulate that this leads to a (small) complex frequency shift $\Delta \Omega$ which modifies the motion of the pattern:

$$\begin{align*}
\Omega &= n\omega_0 + \Delta \Omega \\
\uparrow & \quad \uparrow & \quad \uparrow \\
\text{perturbed frequency} & \quad \text{unperturbed} & \quad \text{(complex) frequency shift} \tag{10}
\end{align*}$$

While this idea appears simple, the rigorous treatment is rather involved. Instead, it is better to follow a shortcut using the well-known analysis of the motion of a particle in the longitudinal phase plane due to an external accelerating RF voltage with amplitude $V_0$ and frequency $f_0$. The phase $\phi$ of a particle with respect to the RF wave satisfies the differential equation for synchrotron oscillations

$$\frac{d}{dt} \left[ \frac{E_0 \beta^2 \gamma \phi}{2\pi \eta h f_0^2 e} \right] + V_0 (\sin \phi - \sin \phi_s) = 0, \tag{11}$$

where $\eta = 1/\gamma^2 - 1/\gamma_i^2$ is the slip factor, $E_0$ the particle rest energy, $h$ the RF harmonic number, and $f_0 = \omega_0/2\pi$ the revolution frequency around the synchrotron. Assuming a stable phase angle of $\phi = 0$ (stationary bucket), small-amplitude phase oscillations ($\sin \phi \approx \phi$), and neglecting the slow variation of $\beta, \gamma$, this equation becomes linear and features constant coefficients:

$$\left[ \frac{E_0 \beta^2 \gamma}{2\pi \eta h f_0^2 e} \right] \phi + V_0 \phi = 0. \tag{12}$$

The small-amplitude synchrotron oscillation frequency is then
\[
\omega_s^2 = \frac{e \eta h V_0 \omega_0^2}{2 \pi E_0 \beta^2 \gamma^2}.
\]  
(13)

Here one takes an imaginative leap by replacing in Eq. (13) the term \( V_0 h \) (external voltage \( \times \) harmonic number) by the beam-induced voltage

\[
\begin{array}{cccc}
\text{\( i \)} & \text{\( n \)} & \text{\( Z \)} & \text{\( I_0 \)} \\
\text{\( \downarrow \)} & \text{\( \downarrow \)} & \text{\( \downarrow \)} & \\
\text{difference in harmonic phase definition} & \text{complex impedance} & \text{\( Z = Z_r + iZ_i \)} & \\
\end{array}
\]

yielding a frequency shift of

\[
(\Delta \Omega)^2 = (\Omega - n\omega_0)^2 = -\frac{e \eta \omega_0^2 n I_0}{2 \pi \beta^2 E_0 \gamma} \left( Z_r + i Z_i \right),
\]  
(14)

which happens to be the correct result! This is the (complex) frequency shift required to sustain a ‘self-consistent’ density modulation.

With the definition of the instantaneous current as given in Eq. (8)

\[
I(t, \Theta) = I_0 + I_1 e^{i(n\theta - \Omega t)} = I_0 + I_1 e^{i(n\theta - (n\omega_0 + \Delta \Omega) t)},
\]  
(15)

where \( \Omega \) is replaced by the slightly modified value \( n\omega_0 + \Delta \Omega \), one gets

\[
I(t, \Theta) = I_0 + I_1 e^{\Delta \Omega t} e^{i(n\theta - (n\omega_0 + \Delta \Omega) t)}
\]

\[
\text{growth or damping } \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow \\
\text{of modulation pattern } \text{real frequency shift},
\]

where \( \Delta \Omega_r \) is the real frequency shift of the rotating pattern, and \( \Delta \Omega_i \) describes the growth or damping of the mode.

Some lengthy algebra based on Eqs. (6), (7) and (9), and use of the formula for the impedance of free space, \( Z_0 = 1/\varepsilon_0 c = 377 \, \Omega \), yield simplified expressions for the (capacitive) space-charge impedance

\[
Z_i = \frac{ng_0 Z_0}{2 \beta \gamma^2}
\]  
(17)

and the inductive impedance

\[
Z_i = -n\omega_0 L.
\]  
(18)

The ‘negative mass’ instability is fully covered by Eq. (14): if the right-hand side is real and positive, \( \Delta \Omega \) is real as well and there is just a real frequency shift of the rotating pattern which then remains stable. However, if the right-hand side is negative or complex, \( \Delta \Omega \) has an imaginary part and is therefore always unstable.

325
In the qualitative treatment, the special case of a perfectly conducting wall (resistive impedance $Z_r = 0$) was considered. Table 1 below is based on Eq. (14) and shows in which situations the beam becomes unstable.

Table 1:

<table>
<thead>
<tr>
<th>$Z_r$</th>
<th>$\gamma &lt; \gamma_l$ ($\eta &gt; 0$)</th>
<th>$\gamma &gt; \gamma_l$ ($\eta &lt; 0$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z_i &gt; 0$ (capacitive)</td>
<td>$\Delta \Omega_i = 0 \Rightarrow$ stable</td>
<td>$\Delta \Omega_i \neq 0 \Rightarrow$ unstable</td>
</tr>
<tr>
<td>$Z_i &lt; 0$ (inductive)</td>
<td>$\Delta \Omega_i \neq 0 \Rightarrow$ unstable</td>
<td>$\Delta \Omega_i = 0 \Rightarrow$ stable</td>
</tr>
</tbody>
</table>

Not surprisingly, this corresponds to the results of the qualitative treatment. A historical note: the term ‘negative mass’ was coined by inspecting Eq. (12) where the term in front of $\mathcal{C}d^2\phi/dt^2$ becomes negative above transition energy ($\eta < 0$) drawing an analogy to the equation of an oscillating spring, $m\mathcal{C}d^2x/dt^2 + k^2x = 0$, this term may be defined as ‘mass’.

In a real synchrotron, the vacuum pipe always has some small resistance, so in general $Z > 0$. Then, from Eq. (13), $\Delta \Omega_i$ features two solutions, with one of them featuring a non-vanishing positive imaginary part, and thus – at this stage of the analysis – the line density modulation will grow under all conditions.

2.4. Stability diagram

The stability diagram is a widely used tool whose significance will become obvious in Section 3. It relates the imaginary part of $\Delta \Omega_i$, the growth (or damping) rate, to the complex impedance $Z_i$ by plotting contours of constant growth rates ($\Delta \Omega_i = \text{const}$) in the $Z_r, Z_i$ plane. Simplifying Eq. (14) by lumping the expression which contains all beam parameters into a single constant $\xi$

\[(\Delta \Omega_i)^2 = -i\xi(Z_i + iZ_r) = \xi(Z_r - iZ_i) = (\Delta \Omega + i\Delta \Omega_i)^2,
\]

and equating real and imaginary parts

\[
\Delta \Omega实在^2 - \Delta \Omega_i^2 = \xi Z_i \Rightarrow \Delta \Omega_i = \sqrt{\xi Z_i + \Delta \Omega实在^2}
\]

\[
i2\Delta \Omega_r \Delta \Omega_i = -iZ_r \xi
\]

results in parabolic contours for $\Delta \Omega_i = \text{const}$:

\[
Z_r = 2\Delta \Omega_i \sqrt{Z_i/\xi + \Delta \Omega实在^2/\xi^2}.
\]

Typical contours are shown in Fig. 4. The area where there is no growth ($\Delta \Omega_i = 0$) is infinitely small and concentrated on the $Z_i$ – axis, that is for $Z_r = 0$.

In summary, at this stage of the analysis, the negative mass instability drives any unbunched beam mode unstable due to the non-vanishing resistivity of the vacuum chamber. Fortunately, this is not true in most real machines, by virtue of a curious phenomenon called Landau damping.

3. LANDAU DAMPING

3.1. Basic idea

In real machines, not all particles in the beam move with the same frequency. Staying with the longitudinal motion as in the example of the density modulation pattern rotating around the synchrotron, the natural energy spread in the beam causes a spread in revolution frequencies (albeit for $\eta \neq 0$). The
coherent motion imposed by the instability becomes confused and may even collapse faster than the rise time of the instability, rendering the beam stable. This is illustrated by the de-coherence of two particles, sketched in Fig. 5.

![De-coherence of two particles](image)

Fig. 5: De-coherence of the longitudinal motion (phase) of two particles after a longitudinal kick at $t = 0$

### 3.2. Landau damping – some analysis

While the basic idea appears simple, one would like to know how much of this frequency spread is actually needed to tame the instability. In what follows, an outline of the complete analysis, which is rather involved, is presented.

Consider a beam consisting of $N$ particles, each oscillating at slightly different frequency $\Omega$, but all of them between $\Omega_1$ and $\Omega_2$ with a density function $g(\Omega)$ normalized by

$$\int_{\Omega_1}^{\Omega_2} g(\Omega) d\Omega = 1$$  \hspace{1cm} (22)

and shown in Fig. 6.

How will they react on an external coherent excitation $\hat{e}^{\omega t}$ (which in fact is generated, for example, by the drive field of a longitudinal coherent instability)? The response of a single particle (with its particular resonant frequency $\Omega$ in general different from the external excitation frequency $\omega$) is
To evaluate $S$, the overall coherent response to the excitation, one integrates the single-particle responses over the $N$ oscillators

$$S = \frac{N}{2\Omega_0} \int_{\Omega_1}^{\Omega_2} \frac{d\Omega}{\Omega - \omega} \cdot \delta(\omega).$$

A more thorough analysis, involving Liouville’s theorem, suggests that the term $\frac{d\delta(\Omega)}{d\Omega}$, rather than $\delta(\Omega)$, has to be employed in Eq. (24).

If the excitation frequency $\omega$ falls outside the beam frequency range (Fig. 7a), there will be a finite coherent response $S$. However, things become less obvious if the excitation frequency is inside the range of the $N$ oscillators ($\Omega_1 < \omega < \Omega_2$) (Fig. 7b), as now the integrand in Eq. (24) features a pole at $\Omega = \omega$.

As described in any textbook on analysis, integral (24) can be evaluated by assuming that $\Omega$ is
complex rather than real and integrating along a half-circle around the pole as depicted in Fig. 8. This procedure is not convergent in all cases, but is so in the present one.

![Diagram](image)

**Fig. 8:** Integral path in the complex $\Omega$ plane enabling the evaluation of the integral with a pole at $\Omega = \omega$ by integrating along a half-circle around the pole

Applying this recipe (Eq. 26), the integral $S$ features a ‘principal value’ and a ‘residuum’ which can be interpreted as a ‘resistive’ term (absorbing energy and thus damping) and a ‘reactive’ term (which does not absorb energy):

\[
S = i \frac{N}{2\pi \sigma_0} \int_{\Omega_1}^{\Omega_2} \frac{d\Omega}{\Omega - \omega} d\Omega - i\Pi = \frac{N}{2\pi \sigma_0} \left[ -\Pi + i \int_{PV} \frac{d\Omega}{\Omega - \omega} d\Omega \right] e^{i\omega t}.
\]

(25)

### 3.3. Stability diagram with Landau damping

The preceding analysis is the most difficult step in the evaluation of beam stability with Landau damping. Note that the amplitude $A$ of the external excitation $A e^{i\omega t}$ was set to 1 in the above analysis, but obviously it depends on beam current, impedance, etc. [see Eq. (14)]. Note also that there is no, or little, Landau damping whenever the beam particles are subjected to an excitation outside their frequency range; **Landau damping only works if the excitation frequency falls inside the frequency range of the beam**.

A complete analysis yields a stability diagram (Fig. 9) which dramatically differs from Fig. 4 where the beam was unstable under all conditions. Plotting the contours of equal growth rates in a $Z$, $Z_t$ diagram ($Z_r = \text{Re}(Z)$, $Z_t = \text{Im}(Z)$), one now disposes of a finite area of stability, allowing in particular for some limited resistive impedance $\text{Re}(Z)$.

The contour with growth rate 0 is called the stability limit, inside of which the growth rate is negative (damping). The form of the stable ‘bottle’ depends on the distribution $g(\Omega)$, while its size scales with beam parameters like the current. For realistic (‘reasonable’) distribution functions, a circle can be inscribed which pretty well approximates the shape. By this technique one can derive a handy approximation for the longitudinal stability limit of unbunched beams, coined as the ‘Keil–Schnell criterion’:

\[
|Z|_n \leq F \frac{m_0 \sigma^2 \beta^2 \gamma |n| (\Delta p/p)^2}{e I_0}.
\]

(26)

Here $Z$ is the complex longitudinal impedance evaluated at $\omega = n\omega_0$, with $\omega_0$ the average revolution frequency, $m_0$ the particle rest mass, $\Delta p/p$ the relative momentum spread, $I_0$ the beam current, and $F$ a form factor of the order of 1, depending on the distribution.
The criterion imposes approximate threshold values for \( \Delta p/p \) (minimum) and the beam intensity \( I_0 \) (maximum) beyond which, for a given impedance \( Z \) and slip factor \( \eta \), the beam becomes unstable.

This longitudinal instability manifests itself in incoherent energy oscillations which are difficult to visualize; the end result is an increase in the beam’s energy spread.

A final remark on Landau damping: its effect is not just limited to the negative mass instability. In fact, providing a frequency spread proves to counteract all types of beam instabilities, be they longitudinal, transverse, for unbunched or bunched beams. Thus Landau damping represents, along with active feedback systems, the most powerful tool at the disposal of accelerator physicists to tame coherent beam instabilities.

3.4. Negative mass instability – an example

This is by no means just of academic interest and may present a prohibitive performance limitation as shown in the following example. The CERN 25 GeV Proton Synchrotron (PS) is part of the LHC injector chain. Before transferring the beam to the 450 GeV SPS (the LHC injector synchrotron), short bunches (total length < 4 ns) with 25 ns spacing have to be prepared. To this end, the 16 beam bunches arriving at the PS ejection flat top are adiabatically de-bunched, and then re-bunched at \( h = 84 \) (25 ns spacing). The beam stays de-bunched during \( \sim 100 \) ms and is not allowed to suffer an increase in \( \Delta p/p \) so as to avoid bunch lengths of more than 4 ns after re-bunching. However, this goal could initially not be achieved for the nominal LHC intensity of \( 8 \times 10^2 \) p as is illustrated in Fig. 10 which shows the beam frequency spread (horizontal axis) during the whole process (time goes downwards). The longitudinal machine impedance of the CERN PS is here dominated by a 114 MHz cavity (left picture) whose impedance \( Z \) at the resonance frequency is difficult to reduce (right picture). The technique used to visualize the frequency spread measures the Schottky noise. In both diagrams one starts with a large energy (frequency) spread (beam bunched on \( h = 16 \)), which is gradually reduced by adiabatic de-bunching, until the RF is cut off and the beam is left coasting. A strong blow-up is then observed in the high-impedance case (left), which apparently does not satisfy the Keil–Schnell criterion, whereas no blow-up is observed for the lower impedance case (114 MHz cavity partially short-circuited, right).
Instabilities

Fig. 10: Increase in momentum spread due to negative mass instability in the CERN PS during de-bunching of $8 \times 10^{12}$ protons. Horizontal axis: frequency spread (proportional to momentum spread); vertical axis: time, circa 200 ms, moving downwards. Left spectrogram: high longitudinal impedance $Z$, beam strongly blown up. Right: low $Z$, now blow-up observed. See text for more details.

4. IMPEDANCE OF A RESONATOR

The machine impedance $Z$ is a function of $\omega$ and typically takes on its largest values at the resonance frequencies of cavity-like objects which thus represent the most critical machine components, prone to drive instabilities. Therefore a closer look at these objects is worth while. The characteristic features of such objects are quite similar to a parallel RLC circuit (Fig. 11) where the a.c. current $I$ represents the a.c. beam current, $R$ the ‘shunt impedance’, $\omega_r$ the resonance angular frequency, $Q$ the quality factor [Eq. (27)].

$$\omega_r = \frac{1}{\sqrt{LC}} \quad \text{resonance frequency}$$

$$Q = R\sqrt{\frac{L}{C}} = \frac{R}{\frac{1}{\omega L}} \quad \text{quality factor}$$
The differential equation of the circuit is

$$\dot{V} + \frac{\omega_r}{Q} V + \omega_r^2 V = \omega_r \frac{R}{Q} i$$  \hspace{1cm} (28)$$

with the solution

$$V(t) = V_0 e^{-\alpha t} \cos \left[ \omega_r \sqrt{1 - 1/4Q^2} t + \phi \right].$$  \hspace{1cm} (29)$$

This represents a damped oscillation with the damping rate $\alpha = 1/\tau = \omega_r / 2Q$.

One calculates the impedance $Z(\omega)$ by exciting the circuit with the current $I = I_0 e^{i\omega t}$, $(-\infty < \omega < \infty)$ and looks for solutions of the form $V = V_0 e^{i\omega t}$. Note that the range of $\omega$ includes negative frequencies simply because it allows doing all calculations with the more handy $e^{i\omega t}$ rather than the clumsy trigonometric functions. By virtue of Eq. 29, one then gets

$$-\omega^2 V_0 e^{i\omega t} + i \frac{\omega \omega_r}{Q} V_0 e^{i\omega t} + \omega_r^2 V_0 e^{i\omega t} = i \frac{\omega \omega_r R}{Q} I_0 e^{i\omega t}$$  \hspace{1cm} (30)$$

from which the impedance $Z(\omega) = V_0/I_0$ is derived.

$$Z(\omega) = \frac{V_0}{I_0} = \frac{R - \frac{1}{1 + iQ \frac{\omega^2 - \omega_r^2}{\omega \omega_r}}}{1 + \left( \frac{\omega^2 - \omega_r^2}{\omega \omega_r} \right)^2}.$$  \hspace{1cm} (31)$$

$Z(\omega)$ is complex because in general $V_0$ is not in phase with the excitation $I_0$. Equation (31) may be re-written to yield the real and imaginary parts of the impedance of a resonator

$$Z(\omega) = Z_r(\omega) + i Z_i(\omega) = R - \frac{iQ \frac{\omega^2 - \omega_r^2}{\omega \omega_r}}{1 + \left( \frac{\omega^2 - \omega_r^2}{\omega \omega_r} \right)^2}$$  \hspace{1cm} (32)$$

which are shown in Fig. 12.

![Impedance Diagram](image-url)

Fig. 12: Real ($Z_R$) and imaginary ($Z_I$) parts of the impedance $Z$ of a resonator as a function of $\omega$. Notice that $Z_R$ is an even function of $\omega$ whereas $Z_I$ is an odd function of $\omega$.

The expression for the impedance of a narrow-band resonator (‘high-$Q$’ cavity) can be simplified near the resonance frequency $\omega_r$ (with $\Delta \omega = \omega - \omega_r$) to

$$Z(\omega) \approx R - \frac{1 - i2Q \frac{\Delta \omega}{\omega_r}}{1 + \left( 2Q \frac{\Delta \omega}{\omega_r} \right)^2}.$$  \hspace{1cm} (33)$$
A narrow-band impedance like this one features a high quality factor $Q$ and thus a low damping rate $\alpha$: once the beam has induced a signal into this object, it will oscillate during many machine turns, memorizing the fields induced during many passages of all bunches (multi-bunch effects). The converse is true for a broad-band cavity: $Q$ is low, the damping rate is large, the induced fields collapse rapidly and are not memorized long enough to have repercussions on subsequent bunches, but only on the bunch itself (single-bunch effects).

5. ONE BUNCH AND A NARROW-BAND CAVITY: ‘ROBINSON’ INSTABILITY

Turning now to bunched beams, the most simple case is the dynamics of one bunch in a narrow-band cavity (historically, this instability was generated by the accelerating cavities). Imagine a bunch performing coherent synchrotron oscillations with the synchrotron frequency $\omega_s$: the bunch rotates in the longitudinal ($\Delta E - \phi$) phase plane. Projecting this motion on the $\phi$-axis, a synchronous detector sees the charge density oscillating around the synchronous phase (see Fig. 13).

![Diagram of a bunch performing coherent synchrotron oscillations in the phase plane and as seen by a synchronous phase detector](image)

Fig. 13: A bunch performing coherent synchrotron oscillations in the phase plane and as seen by a synchronous phase detector

The energy and phase of the bunch vary periodically with $\omega_s$: in instability jargon, this is called ‘rigid bunch’ or ‘dipole’ mode.

Assume that the cavity is tuned at $\omega_r$, which is near to, but not exactly at, a multiple $n$ of the bunch’s revolution angular frequency $\omega_c (\omega_r \approx \omega = n\omega_c)$. A close look at Fig. 14 reveals the behaviour of the beam under various conditions:

- For $\gamma < \gamma_t$ and beam frequency $\omega$ below resonance frequency $\omega_r$: whenever $\Delta E > 0$ during the phase oscillation, $\omega$ increases (because below transition), sees a larger real impedance $R_\pm$ which takes more energy from the beam than at $R_0$, a process which stabilizes the dipole oscillation. Conversely,
- for $\gamma < \gamma_t$ and $\omega > \omega_r$ the beam is unstable.

- For $\gamma > \gamma_t$ and $\omega < \omega_r$: whenever $\Delta E > 0$ during the phase oscillation, $\omega$ decreases (because above transition), sees a lower real impedance $R_\pm$ which absorbs less energy from the beam than at $R_0$, thus rendering the motion unstable. Conversely,
- for $\gamma > \gamma_t$ and $\omega > \omega_r$ the beam is stable.

The instability is removed by fine-tuning of the cavity’s resonance frequency $\omega_r$ slightly away from the beam frequency $\omega = n\omega_c$. However, nowadays, with the ever increasing performance requirements, an active feedback on the cavity tune proves more efficient for removing this instability.
6. INSTABILITIES WITH MANY BUNCHES

6.1. Coupled-bunch modes and their origin

Modern synchrotrons often work with many bunches (e.g. large colliders and their injectors, beauty factories, synchrotron radiation sources, etc.). Moreover, high-\(Q\) cavities, sometimes superconducting, are employed to ensure efficient acceleration. Often these cavities have (undesirable) higher-order modes where the impedance at certain multiples of the revolution frequency takes on high values. As outlined in Section 4, a combination of many bunches and narrow-band resonant impedances enables the latter to memorize the fields for many bunch passages: the induced field due to the first bunch drives the motion of the second bunch, which in turn excites the third bunch, and so on until the first bunch appears back at the cavity for a second time, eventually leading to coupled-bunch instabilities. Taking as an example a machine with \(M = 4\) bunches performing synchrotron oscillations, all four bunches may oscillate in phase. However, in addition to this in-phase oscillation mode (mode number \(n = 0\)), three other ‘modes’ of coupling are possible with synchrotron oscillation phase shifts between consecutive bunches of \(\pi / 2, \pi, 3\pi / 2\) \((n = 1, 2, 3)\). These four modes are depicted in Fig. 15 which shows the motion of the four bunches in the longitudinal phase plane for the four modes.

For a machine with \(M\) bunches, there are \(M\) modes, with bunch-to-bunch phase shifts \(2\pi n / M\), \(0 \leq n \leq M - 1\). All these modes are possible, but may grow unstable only in the presence of a high impedance at the mode’s frequency.

6.2. Stability of coupled-bunch mode: qualitative analysis

Staying with the example discussed above, there are four bunches in a machine, and a (perturbing) resonator tuned at the revolution frequency. Under what conditions do the bunches become unstable? Three different cases are discussed below, all of them above transition energy where particles move clockwise in the longitudinal \((\Delta E, \phi)\) plane.

Case 1: the four bunches do not perform coherent synchrotron oscillations. Figure 16 shows the voltage induced into the resonator by each of the four bunches during one machine revolution time \(T_i\), and the motion of each bunch in the longitudinal phase plane – there is no motion in this case as there are no coherent synchrotron oscillations.

The voltage induced by bunch 2 cancels with the one induced by bunch 4, and likewise the voltages generated by bunches 1 and 3 cancel. Hence there is no net induced voltage, the beam is stable.
Fig. 15: The four modes of coupled bunch oscillations for a synchrotron with four bunches.

Fig. 16: Voltages induced by each of four bunches in a cavity tuned at the revolution frequency. Bunches do not perform synchrotron oscillations in this case.
Case 2: the four bunches do perform synchrotron oscillation with a bunch-to-bunch phase shift of $\pi/2$ (mode $n = 1$). Figure 17 explains what happens to these bunches.

While the voltage of bunches 2 and 4 cancel as in Fig. 16, bunches 1 and 3 induce a net voltage. This voltage in turn affects bunch 2 which gets accelerated while bunch 4 is decelerated, thus increasing their synchrotron oscillation amplitude, and the bunches are unstable. It is easy to figure out that one-quarter of a synchrotron oscillation later (usually many machine turns), bunches 2 and 4 will drive 1 and 3 unstable.

Case 3: The four bunches perform synchrotron oscillations with bunch-to-bunch phase shift of $3\pi/2$ (mode $n = 3$) (Fig. 18).

Here again, the voltages induced by bunches 2 and 4 cancel, whereas 3 and 1 induce a net voltage. Bunch 2 is accelerated by this voltage, and bunch 4 decelerated, exactly as for mode $n = 1$ (Fig. 17). However, the synchrotron oscillation amplitudes of bunches 2 and 4 are both shrinking, so in this case the mode $n = 3$ is stable! Notice that a fixed observer, looking at the oscillating bunches with a synchronous detector, cannot tell the difference between modes $n = 1$ and $n = 3$. Yet, one of them is unstable, the other is stable.

In this qualitative approach, modes $n = 0$ (all bunches oscillating in phase) as well as $n = 2$ (phase shift $\pi$ between adjacent bunches) are both stable. The cases discussed above apply above transition energy; below transition, bunches rotate anti-clockwise in the $(\Delta E - \phi)$ plane, thus mode $n = 1$ becomes stable, and $n = 3$ unstable.

### 6.3. Higher-order coupled-bunch modes

So far the most important type of longitudinal oscillation mode, the dipole mode where the bunches rigidly rotate in the phase plane, has been discussed. Under certain conditions, bunches may change their shape, rather than their position, periodically, and this coherent motion may be driven by an impedance and become unstable as well. Defining a further mode number for the bunch shape, $m$, with $m = 1$ for the dipole mode, one often observes $m = 2$ (quadrupole mode), $m = 3$ (sextupole mode), and sometimes even $m = 4$ (octupole mode). What do these modes look like?
Various oscilloscope signals as well as the bunch’s shapes in the longitudinal phase plane are compiled in Fig. 19 for a machine with five proton bunches (CERN PS Booster). Two adjacent bunches are shown.

What can one learn from Fig. 19?

- A ‘mountain range’ display is a record of the bunch shape over many turns, where, for example, each tenth turn is visualized and vertically displaced.
- The mountain range displays in Fig. 19a feature a shift in the oscillation phase between the two consecutive bunches.
- The higher the mode, the more often the pattern repeats in vertical direction. While the dipole mode is visualized during about one oscillation period, the higher-order modes feature several periods. In fact, while the dipole mode oscillates with the angular synchrotron frequency \( \gamma \) (\( \sim 2 \pi \) 2 kHz in this case), the higher modes oscillate with \( m \omega_0 \). This can readily be understood by inspecting the sketches in Fig. 19c (longitudinal phase plane) where the patterns rotate with the synchrotron frequency.

Notice that ‘quadrupole’ oscillations in the (\( \Delta E - \phi \)) plane are often observed after injecting a beam from a lower energy synchrotron. This is due to insufficient adaptation between the respective RF voltages and is called ‘longitudinal mismatch’. However, these oscillation do not necessarily give rise to an instability.

### 7. LONGITUDINAL MICROWAVE INSTABILITY

This instability is observed in synchrotrons with high-intensity bunches. It does not matter whether there are many bunches or just a few, because it is a single-bunch effect.

#### 7.1. Signature

High-frequency density modulations along the bunch are observed, as sketched in Fig. 20. The frequency typically ranges from several 100 MHz to several GHz, thus their wavelength is much shorter than the bunch length. The coherent signal is difficult to observe (in general with spectrum analysers), because it becomes rapidly blurred and leads to a very fast and uncontrolled blow-up of the bunch area which may then largely exceed the acceptance of the RF system.
Fig. 19: Dipole and higher-order coupled-bunch oscillation modes as observed for a proton synchrotron with five bunches (CERN PS Booster). The first row (a) shows ‘mountain range’ displays of two adjacent oscillating bunches, the second row (b) the superimposed signals as observed by a synchronous phase detector, and the third row (c) sketches the bunch shape in the phase plane. The first column covers the $m = 1$ (dipole) mode, followed by $m = 2$ (quadrupole), $m = 3$ (sextupole), $m = 4$ (octupole) modes.

Fig. 20: Signature of longitudinal microwave instability: a high-frequency density modulation is superimposed on the bunch shape $\lambda(\phi)$.
Typically, the e-folding time of the growth is shorter than a synchrotron oscillation period. Even lepton machines which normally are less prone to instabilities because of the damping due to synchrotron radiation suffer from this effect because the growth rate is faster than the damping rate. For lepton machines it is also called 'turbulent bunch lengthening' simply because for a given RF voltage, the blow-up in bunch area manifests itself in a longer bunch.

7.2. The broad-band impedance

In order to drive a density modulation of, for example, 1 GHz unstable, there must be an impedance at $\sim 1$ GHz. As highlighted in Section 2, the vacuum chambers of synchrotrons, and in particular of older machines which were not built for high intensities in the first place, feature many abrupt cross-section changes, cavity-like elements, etc. All these elements of a synchrotron, together with objects like septum magnets and beam observation devices, may be lumped into one low-Q 'resonator', also called broad-band impedance. Recalling the impedance of a resonator (Eq. 33), one calculates this broad-band impedance

$$Z(\omega) = R_s \frac{1 - i Q \frac{\omega^2 - \omega_0^2}{\omega\omega_r}}{1 + \left(\frac{Q \frac{\omega^2 - \omega_0^2}{\omega\omega_r}}{\omega\omega_r}\right)^2}$$

(34)

with $Q \approx 1$ and the resonance frequency $\omega_r \approx 1$ GHz. This impedance $Z(\omega)$ is sketched in Fig. 21.

![Fig. 21: Broad-band impedance with resonance frequency about 1 GHz and $Q \approx 1$. This impedance approximates the sum of individual impedances of most elements of a synchrotron.](image)

In contrast with a narrow-band impedance, a wide-band impedance has a very short 'memory' for fields induced by a bunch, thus there is barely any coupling between adjacent bunches. Therefore, this is a single-bunch effect where the head of the bunch induces a high-frequency field which increases along the bunch. Obviously this model, using the approximate broad-band impedance, lacks precision, but has the merit of explaining the observed effect rather well.

In order to estimate this impedance, one measures its value for small $\omega$, where according to Eq. (34) $Z(\omega)$ may be approximated by

$$Z(\omega) \approx i \frac{R_s \omega}{Q \omega_r} = i \frac{R_s}{Q} \frac{\omega}{\omega_0} \frac{\omega_0}{\omega_r}.$$ 

(35)

Replacing the ratio $\omega/\omega_0$ by $n$, and with $Q = R_s / (\omega_r L)$ (Eq. 28) one finds for the low-frequency impedance

$$\left| \frac{Z}{n} \right|_{\omega_0} = L \omega_0.$$ 

(36)
This is an inductive impedance, normalized to $\frac{n}{Z}$, the multiple of the revolution frequency $\omega_0$. While $|Z/n|$ is constant for small $\omega_0$, $|Z|$ rises linearly with $\omega$ as shown in Fig. 21 (dotted line).

The value of $|Z/n|$ determined in this way is conveniently called the impedance of the synchrotron and is given in $\Omega$. For older machines, this value is in the ballpark of 20–50 $\Omega$ whereas for more modern machines, often designed on purpose to minimize this figure, it is of the order of 1 $\Omega$.

### 7.3. Stability limit

As for the coasting beam instability, experience has shown that, also in this case, a stability criterion provides a rule of thumb to decide whether or not a bunch will suffer from microwave instabilities. One applies the Keil–Schnell coasting beam criterion (26) to the instantaneous current and momentum spread in the bunch

$$
\left| \frac{Z}{n} \right| \leq F \frac{m_0 c^2 \beta^2 \gamma |\eta|}{e} \left[ \frac{(\Delta p/p)^2}{I} \right] \text{ instant.}
$$

(37)

where $\Delta p/p$ and $I$ are the instantaneous relative momentum spread and beam current, respectively, and $F$ a form factor of the order of unity for protons, and about 8 for leptons (‘Keil–Schnell–Boussard’ criterion).

It is interesting to note that a bunch with a given longitudinal emittance and given bunch population features a largely different stability behaviour depending on its shape in the phase plane: as sketched in Fig. 22, short bunches with large $\Delta p/p$ are more stable than long bunches with low $\Delta p/p$.

This may be explained by inspecting relation (37) where the ratio $(\Delta p/p)^2/I$ is larger in the short bunch case, thus allowing for a larger limit impedance $|Z/n|$ or a larger current $I$.

---

8. **FURTHER READING**

This paper is a write-up of a one-hour CAS Lecture. It is very much inspired by the lecture notes ‘Instabilities’ by E.J.N. Wilson, edited in 1997. There are quite a few other lecture notes which may guide the reader towards further insight in the subject.
• A. Hofmann’s classical 1976 write-up of his lecture on longitudinal phenomena in the first CAS [1].
• J.L. Laclare’s very complete analysis on both coasting and bunched beam instabilities (transverse and longitudinal) [2], [3].
• A more pragmatic description of instabilities which includes their observation and correction by J. Gareyte [4].
• A special treatise on multi-bunch instabilities and their cure by F. Pedersen [5].
• Rather recent lecture notes on instabilities and beam intensity limitations in synchrotrons, with special emphasis on the findings in lepton machines, by S. Myers [6].
• Of course, A.W. Chao’s book, which covers all collective beam instabilities, will enlighten any interested reader [7].
References


Abstract
A charged particle beam travelling across perfectly conducting structures whose boundaries do not have constant cross section, such as an RF cavity or bellows, generates longitudinal and transverse wake fields. We discuss in this lecture the general features of wake fields, and show a few simple examples in cylindrical geometry: perfectly conducting pipe and the resonant modes of an RF cavity. We then study the effect of wake fields on the dynamics of a beam in a linac, such as beam break-up instabilities and how to cure them.

1 INTRODUCTION
Parasitic forces, also called wake fields, are generated by a charged particle beam interacting with the vacuum chamber components. These components may have a complex geometry: kickers, bellows, RF cavity, diagnostics components, special devices, etc. To solve Maxwell’s equations in a given structure with the beam current as the source of fields, a study of the field is required. For this complicated task, dedicated computer codes were developed to solve the electromagnetic problem in the frequency or in the time domain. There are several useful codes for the design of accelerator devices, like MAFIA, ABCI, URME, etc., as reported in Ref. [1].

In this lecture we discuss the general features of the parasitic fields [2-10], and then show a few simple examples of them in cylindrical geometry: a perfectly conducting pipe and the resonant modes of an RF cavity. Although the space charge forces have been studied separately [11], they can be seen as a particular case of wake fields, see Appendix B for a simple example [12].

We then study the effect of the wake fields on the dynamics of a beam in a linac such as beam break-up instabilities, assuming a high-energy beam where the motion is ‘frozen’ in longitudinal space, and the way to cure it [13].

2 WAKE FIELDS AND POTENTIALS

2.1 Wake potentials
Parasitic fields depend on the particular charge distribution of the beam. It is therefore desirable to know what is the effect of a single charge (i.e. find the Green function) in order to reconstruct the fields produced by any charge distribution.

The electromagnetic fields created by a point charge act back on the charge itself and on any other charge of the beam. We therefore focus our attention on the source charge $q_0$, and on the test charge $q$, assuming that both are moving with the same constant velocity $v = \beta c$ on trajectories parallel to the axis.

Let $E$ and $B$ be the fields generated by $q_0$ inside the structure, $(s_0 = vt, \, r_0)$ be the position of the source charge, and $(s = s_0 + z, \, r)$ be the position of the test charge $q$.

Since the velocity of both charges is along $z$, the Lorentz force has the following components:
Thus, there can be two effects on the test charge: a longitudinal force which changes its energy, and a transverse force which deflects its trajectory. If we consider a device of length \( L \), the energy gain \( J \) is:

\[
U = \int_0^L F_\parallel ds
\]

and the transverse deflecting kick \([\text{Nm}]\) is:

\[
M = \int_0^L F_\perp ds.
\]

Note that the integration is performed over a given path of the trajectory. These quantities, normalized to the charges, are called *wake potentials* (volt/coulomb) and are both functions of the distance \( z \):

- **Longitudinal wake potential \([\text{V/C}]\)**: 
  \[
  w_\parallel = -\frac{U}{q_0 q}.
  \]

- **Transverse wake potential \([\text{V/Cm}]\)**: 
  \[
  w_\perp = \frac{1}{r_0 q_0 q} M.
  \]

The minus sign in the longitudinal wake-potential means that the test charge loses energy when the wake is positive. Positive transverse wake means that the transverse force is defocusing.

As a first example, consider the longitudinal wake-potential of 'space charge'. The longitudinal force inside a relativistic cylinder of radius \( a \) travelling inside a cylindrical pipe of radius \( b \) is given by [11]:

\[
F_z(r,z) = \frac{-q}{4\pi \varepsilon_0 \gamma} \left(1 - \frac{r^2}{a^2} + 2 \ln \frac{b}{a}\right) \frac{\partial \lambda(z)}{\partial z}.
\]

Note that since the space charge forces move together with the beam, and the electric field is constant vs. \( z \), we can derive the wake potential per unit length (volt/coulomb metre). To get the wake potential of a piece of pipe, we just multiply by the pipe length. Assuming \( r \to 0 \) and a charge line density given by \( \lambda(z) = q_0 \delta(z) \) we obtain:

\[
\frac{dw_\parallel(z)}{ds} = \frac{1}{4\pi \varepsilon_0 \gamma} \left(1 + 2 \ln \frac{b}{a}\right) \frac{\partial}{\partial z} \delta(z).
\]

Another interesting case is the longitudinal wake-potential of a resonant higher order mode (HOM) in an RF cavity. When a charge crosses a resonant structure, it excites the fundamental and higher order modes. Each mode can be treated as an electric RLC circuit loaded by an impulsive current, as shown in Fig. 1.
Fig. 1: RF cavity and the equivalent RLC parallel circuit model driven by a current generator

Just after the charge passage, the capacitor is charged with a voltage $V_0 = Cq_0$ and the electric field is $E_{00} = V_0/l_0$. The time evolution of the electric field is governed by the same differential equation of the voltage:

$$\dot{V} + \frac{1}{RC} \dot{V} + \frac{1}{LC} V = \frac{1}{C} \dot{I}. \quad (8)$$

The passage of the impulsive current charges only the capacitor, which changes its potential by an amount $V_c(0)$. This potential will oscillate and decay producing a current flow in the resistor and inductance. For $t > 0$ the potential satisfies the following equation and boundary conditions:

$$\dot{V} + \frac{1}{RC} \dot{V} + \frac{1}{LC} V = 0$$
$$V(t = 0^+) = \frac{q}{C} \equiv V_0$$
$$\dot{V}(t = 0^+) = \frac{\dot{q}}{C} = \frac{I(0^+)}{C} = \frac{V_0}{RC} \quad (9)$$

which has the following solution:

$$V(t) = V_0 e^{-\Gamma t} \left[ \cos(\omega t) - \frac{\Gamma}{\omega} \sin(\omega t) \right] \quad (10)$$

where $\omega^2 = 1/LC$ and $\Gamma = 1/2RC$.

Putting $z = ct$ ($z$ is positive behind the charge) we obtain (see Fig 2.):

$$w(z) = \frac{-V(z)}{q_0} = w_0 e^{-\Gamma z/c} \left[ \cos(\omega z/c) - \frac{\Gamma}{\omega} \sin(\omega z/c) \right]. \quad (11)$$
2.2 Loss factor

It is also useful to define the loss factor as the normalized energy lost by the source charge $q_0$:

$$k = - \frac{U(z = 0)}{q_0}.$$  \hfill (12)

Although in general the loss factor is given by the longitudinal wake at $z = 0$, for charges travelling with the velocity of light, the longitudinal wake potential is discontinuous at $z = 0$, see Fig. 3.

The exact relationship between $k$ and $w(z \to 0)$ is given by the beam loading theorem [14].

2.3 Beam loading theorem

When the source charge travels with the velocity of light $v = c$, the electromagnetic fields are left behind and called ‘wake fields’. Any electromagnetic perturbation produced by the charge cannot overtake the charge itself. This means that the longitudinal wake-potential vanishes in the region $z < 0$. This property is a consequence of the ‘causality principle’. Causality requires that the longitudinal wake-potential of a charge travelling with the velocity of light is discontinuous at the origin.

The beam loading theorem states that:

$$k = \frac{w_\parallel(z \to 0)}{2}.$$  \hfill (13)

For example, the beam loading theorem is fulfilled by the wake potential of the resonant mode. In fact the energy lost by the charge $q_0$ loading the capacitor is:

$$U = \frac{CV_0^2}{2} = \frac{q_0^2}{2C}.$$
giving $k = 1/2C$, to compare with:

$$w_\parallel(z \to 0) = \frac{1}{C}.$$ 

### 2.4 Relationship between transverse and longitudinal forces

Another important feature worth mentioning is the differential relationship existing between longitudinal and transverse forces:

$$\nabla \perp F_\parallel = \frac{\partial}{\partial z} F_\perp$$

$$\nabla \perp w_\parallel = \frac{\partial}{\partial z} w_\perp$$  \hspace{1cm} (14)

The above relations are known as the ‘Panofsky–Wenzel theorem’ [15].

### 2.5 Coupling impedance

The wake potentials are used to study the beam dynamics in the time domain ($s = vt$). If we take the equation of motion in the frequency domain, we need the Fourier transform of the wake potentials. Since these quantities are in ohms they are called coupling impedances:

**Longitudinal impedance** [$\Omega$]:

$$Z_\parallel(\omega) = \frac{1}{\nu} \int_{-\infty}^{\infty} w_\parallel(z) e^{-}\frac{\omega}{\nu} dz.$$  \hspace{1cm} (15)

**Transverse impedance** [$\Omega/m$]:

$$Z_\perp(\omega) = \frac{i}{\nu} \int_{-\infty}^{\infty} w_\perp(z) e^{-}\frac{\omega}{\nu} dz.$$  \hspace{1cm} (16)

The coupling impedances ($\Omega/m$) of the ‘space charge’ wake is:

$$\frac{\partial Z_\parallel(\omega)}{\partial s} = \frac{1}{\nu} \int_{-\infty}^{\infty} \frac{\partial w_\parallel(z)}{\partial s} e^{-}\frac{\omega}{\nu} dz = \frac{1 + 2\ln(b/a)}{\nu 4\pi \epsilon_0 \gamma^2} \int_{-\infty}^{\infty} d\zeta e^{-}\frac{\omega}{\nu} \zeta dz.$$  \hspace{1cm} (17)

where: $\int_{-\infty}^{\infty} \delta(z)f(z)dz = f'(0)$, so that:

$$\frac{\partial Z_\parallel(\omega)}{\partial s} = \frac{-i\omega Z_0}{4\pi c \beta^2 \gamma^2} \left(1 + 2\ln\frac{b}{a}\right).$$  \hspace{1cm} (18)

The longitudinal coupling impedance of a resonant HOM is given by:

$$Z_\parallel(\omega) = \frac{R_s}{1 - iQ_r \left(\frac{\omega_r - \omega}{\omega_r}\right)}$$  \hspace{1cm} (19)

where $R_s = w_\parallel / 2\Gamma$ is the shunt impedance and $Q_r = \omega_r / 2\Gamma$ is the quality factor, quantities that we can obtain with the computer codes. Note that the loss factor is also determined: $k = \omega R_s / 2Q_r$.

The transverse impedance is given by:
\[ Z_{\perp}(\omega) = \frac{c}{\omega} \frac{R_{s\perp}}{1 - iQ_r \left( \frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right)} , \quad (20) \]

with \( R_{s\perp} = R_s / b^2 \) being the transverse shunt impedance.

### 2.6 Wake potential and energy loss of a bunched distribution

When we have a bunch with density \( \lambda(z) \), we may wonder what is the amount of energy lost or gained by a single charge \( e \) in the beam. To this end we calculate the effect on the charge from the whole bunch by means of the convolution integral, see Fig. 4:

\[
U(z) = -e \int_{-\infty}^{\infty} w_{||}(z' - z) \lambda(z') dz' . \quad (21)
\]

![Fig. 4: Convolution integral](image)

This permits the definition of the wake potential of a distribution:

\[
W_{||}(z) = -\frac{U(z)}{qq_0} . \quad (22)
\]

The total energy lost by the bunch is computed summing up the loss of all particles:

\[
U_{\text{bunch}} = -\frac{1}{e} \int_{-\infty}^{\infty} U(z') \lambda(z') dz' . \quad (23)
\]

### 3 WAKE-FIELD EFFECTS IN LINEAR ACCELERATORS

#### 3.1 Energy spread

The longitudinal wake forces change the energy of individual particles depending on their position in the beam. As a consequence, the wake can induce an energy spread in the beam.

For example, the energy spread induced by the space charge force in a Gaussian bunch is given by:

\[
\frac{dU(z)}{ds} = -e \int_{-\infty}^{\infty} \frac{dw_{||}(z' - z)}{ds} \lambda(z') dz' = -\frac{eq}{4\pi \varepsilon_0 \gamma^2 \sqrt{2\pi} \sigma_z} \left( 1 + 2 \ln \frac{b}{a} \right) z e^{-z^2/(2\sigma_z^2)} . \quad (24)
\]

The bunch head gains energy, while the tail loses energy.
In a similar way one can show that the energy loss induced by a resonant HOM on the charges inside a rectangular, uniform bunch is given by:

\[
U(z) = \frac{-eq_0w^2}{2} \sin \left[ \frac{\omega_z}{c} \left( \frac{l_0}{2} + z \right) \right].
\]

(25)

3.2 Beam break-up

A beam injected off-centre in a linac, because, for example, of focusing quadrupole misalignment, executes betatron oscillations. The displacement produces a transverse wake field in all the devices crossed during the flight, which deflects the trailing charges, see Fig. 5.

In order to understand the effect, consider a simple model with only two charges \( q_1 = Ne/2 \) (leading = half bunch) and \( q_2 = e \) (trailing = single charge). The leading charge executes free betatron oscillations:

\[
y_1(s) = \hat{y}_1 \cos \left( \frac{\omega_y}{c} s \right).
\]

(26)

The trailing charge, at a distance \( z \) behind, over a length \( L_w \) experiences a deflecting force proportional to the displacement \( y_1 \), and dependent on the distance \( z \):

\[
\langle F_y^{\text{self}}(z, y_1) \rangle = \frac{Ne^2}{2L_w} w_1(z) y_1(s).
\]

(27)

Notice that \( L_w \) is the length of the device for which the transverse wake has been computed. For example, in the case of a cavity cell \( L_w \) is the length of the cell. This force drives the motion of the trailing charge:

\[
y_2^\prime + \left( \frac{\omega_y}{c} \right)^2 y_2 = \frac{Ne^2 w_1(z)}{2E_0L_w} \hat{y}_1 \cos \left( \frac{\omega_y}{c} s \right).
\]

(28)

This is the typical equation of a resonator driven at the resonant frequency.

The solution is given by the superposition of the ‘free’ oscillation and a ‘forced’ oscillation which, being driven at the resonant frequency, grows linearly with \( s \), as shown in Fig. 6.

\[
y_2(s) = \hat{y}_2 \cos \left( \frac{\omega_y}{c} s \right) + y_2^{\text{forced}}
\]

(29)
At the end of the linac of length $L$, the oscillation amplitude is grown by ($\hat{y}_1 = \hat{y}_2$):

$$\begin{bmatrix} \Delta \hat{y}_2 \\ \hat{y}_2 \end{bmatrix}_{\text{max}} = \frac{cN_e w(z)L_y}{4\omega_y E_0 L_w} \hat{y}_1 \sin\left(\frac{\omega_y}{c} s\right).$$

(31)

If the transverse wake is given per cell, the relative displacement of the tail with respect to the head of the bunch depends on the number of cells. It depends, of course, also on the focusing strength through the frequency $\omega_y$.

To extend the analysis to a particle distribution, we write the transverse equation of motion of a single particle with the inclusion of the transverse wake field effects as [7]:

$$\frac{\partial}{\partial s} \left[ \gamma(s) \frac{\partial y(z,s)}{\partial s} \right] + k_y^2(s)\gamma(s)y(z,s) = -\frac{e^2 N_p}{m_0 c^2 L_w} \int_z^w y(s,z')w(z'-z)\lambda(z')dz'.$$

(32)

where $\gamma(s)$ is the relativistic parameter, $k_y(s)$ the beta function, $N_p$ the number of particles of the bunch, and $\lambda(z)$ the longitudinal bunch distribution.

The solution of the equation in the general case is unknown. We can, however, apply a perturbation method to obtain the solution at any order in the wake-field intensity. Indeed we write:

$$y(z,s) = \sum_n y^{(n)}(z,s).$$

(33)

The first-order solution is found from the equation

$$\frac{\partial}{\partial s} \left[ \gamma(s) \frac{\partial y^{(0)}(z,s)}{\partial s} \right] + k_y^2(s)\gamma(s)y^{(0)}(z,s) = 0.$$
It is important to note that the above equation does not depend on \( z \) any more. This means that the bunch distribution remains constant along the structure.

If the \( s \)-dependence of \( \gamma(s) \) and \( k_y^2(s)\gamma(s) \) is moderate, we can use the WKB approximation, and the solution of the above equation with the starting conditions \( y(0) = y_m, y'(0) = 0 \) is [2]

\[
y^{(0)}(s) = \sqrt{\frac{\gamma_0 k_y}{\gamma(s) k_y(s)}} y_m \cos[\psi(s)]
\]  

(35)

where

\[
\psi(s) = \int_0^s k_y(s') ds'
\]  

(36)

represents the unperturbed transverse motion of the bunch.

The second-order differential equation is obtained by substituting the first-order solution in the right-hand side of Eq. (31) giving

\[
\frac{\partial}{\partial s} \left[ \gamma(s) \frac{\partial y^{(1)}(z, s)}{\partial s} \right] + k_y^2(s)\gamma(s)y^{(1)}(z, s) = -\frac{e^2 N_p}{m_0 c^2 L_w} y^{(0)}(s) \int_z^\infty w\_{zz}(z' - z) \lambda(z') dz'.
\]  

(37)

The forced solution of the above equation can be written in the form

\[
y^{(1)}(z, s) = -y_m \frac{e^2 N_p}{m_0 c^2 L_w} \sqrt{\frac{\gamma_0 k_y}{\gamma(s) k_y(s)}} G(s) \int_z^\infty w\_{zz}(z' - z) \lambda(z') dz'
\]  

(38)

where

\[
G(s) = \frac{1}{\gamma(s') k_y(s')} \sin[\psi(s) - \psi(s')] \cos[\psi(s')] ds' = \frac{1}{2} \int_0^s \sin[\psi(s) - 2\psi(s')] ds' + \frac{1}{2} \sin[\psi(s)] \int_0^s \frac{1}{\gamma(s') k_y(s')} ds'.
\]  

(39)

The first integral undergoes several oscillations with \( s \) and, if \( \gamma(s) \) and \( k_y(s) \) do not vary much, we can write

\[
y^{(1)}(z, s) = -y_m \frac{e^2 N_p}{2m_0 c^2 L_w} \sqrt{\frac{\gamma_0 k_y}{k_y(s)}} \sin[\psi(s)] \int_0^s \frac{ds'}{\gamma(s') k_y(s')} \int_z^\infty w\_{zz}(z' - z) \lambda(z') dz'.
\]  

(40)

The last integral represents the transverse wake-potential produced by the whole bunch. This solution can then be substituted again in the right-hand side of Eq. (32) to obtain a second-order equation and so on. If we consider constant \( \gamma(s) \) and \( k_y(s) \), we obtain the same result of the two-particle model when we substitute \( \lambda(z) \) with 1/2 in the particle positions.
For example, by using the wake of LEP superconducting cavities, it is possible to find that for a Gaussian bunch, the wake potential is as given in Fig. 7.

![Fig. 7: Transverse wake potential for LEP superconducting cavities](image)

For a conservative estimation of the BBU effect, one should use the maximum value of the curve in this case, thus eliminating the \( z \) dependence.

If the BBU effect is strong, it is necessary to include higher order terms in the perturbation expansion. Assuming:

- rectangular bunch distribution, \( \lambda(z) = 1/l_0 \), \(-l_0/2 < z < l_0/2\), \( l_0 \) bunch length,
- monoenergetic beam,
- constant acceleration gradient \( dE_0 / ds = \text{const} \),
- constant beta function,
- linear wake function inside the bunch \( w(z) = w_0z/l_0 \),

the sum of Eq. (33) can be written in terms of powers of the dimensionless parameter \( \eta \) also called BBU strength

\[
\eta = \frac{e^2 N_p}{k_y (dE_0 / ds) L_w} \frac{w_{10}}{\ln \left( \frac{\gamma_f}{\gamma_i} \right)}
\]  

(41)

with \( \gamma_i \) and \( \gamma_f \) respectively the initial and final relativistic parameter.

By using the method of stepping descent [5], it is possible to obtain the asymptotic expression of \( \gamma(z, s) \) finding at the end of the linac,

\[
\gamma(L_m) = \gamma_m \sqrt{\frac{\gamma_i}{6\pi\gamma_f}} \frac{1}{\eta^{1/6}} \exp \left[ \frac{3\sqrt{3}}{4} \eta^{1/3} \right] \cos \left[ k_y L_m - \frac{3}{4} \eta^{1/3} + \frac{\pi}{12} \right]
\]

(42)

the two-particle model is different from the first-order solution, and gives a tail displacement growing exponentially with \( \eta \), resulting in better agreement with the simulation in Fig. 6.

### 3.3 BNS damping

The BBU instability is quite harmful and hard to get under control even at high energy with a strong focusing, and after careful injection and steering. A simple method to cure it has been proposed on the basis that the strong oscillation amplitude of the bunch tail is mainly due to ‘resonant’ driving. If the
tail and the head move with a different frequency, this effect can be significantly reduced [13], compare Fig. 8 with Fig. 6.

\[
\begin{align*}
\text{Fig. 8: HOMDYN simulation of a typical BNS damping, 50 μm initial offset, 2% energy spread}
\end{align*}
\]

Let us assume that the tail oscillates with a frequency \( \omega_y + \Delta \omega_y \), the two particle model equation of motion reads:

\[
y''_2 + \left( \frac{\omega_y + \Delta \omega_y}{c} \right)^2 y_2 = \frac{N e^2 w_1(z)}{2 \beta^2 E_0 L_w} \hat{y}_1 \cos \left( \frac{\omega_y}{c} s \right)
\]

the solution of which is

\[
y_2(s) = \frac{c^2 N e^2 w_1(z)}{4 \omega_y \Delta \omega_y E_0 L_w} \hat{y}_1 \left[ \cos \left( \frac{\omega_y + \Delta \omega_y}{c} s \right) - \cos \left( \frac{\omega_y}{c} s \right) \right]
\]

where the amplitude of the oscillation is limited.

Furthermore, by a suitable choice of \( \Delta \omega_y \), it is possible to fully depress the oscillations of the tail. Setting

\[
\hat{y}_2 \cos \left( \frac{\omega_y + \Delta \omega_y}{c} s \right) + y_2(s) = \hat{y}_1 \cos \left( \frac{\omega_y}{c} s \right)
\]

we get

\[
\Delta \omega_y = \frac{c^2 N e^2 w_1(z)}{4 \omega_y E_0 L_w}.
\]

The extra focusing at the tail can be obtained by using a RFQ, where head and tail see a different focusing strength; exploiting the energy spread across the bunch which, because of the chromaticity, induces a spread in the betatron frequency. An energy spread correlated with the position is attainable with the external accelerating voltage, or with the wake fields.
4 LANDAU DAMPING

There is a fortunate stabilizing effect against collective instabilities called ‘Landau Damping’. The basic mechanism relies on the fact that if the particles in the beam have a spread in their natural frequencies (synchrotron or betatron), their motion cannot be coherent for a long time.

4.1 Driven oscillators

In order to understand the physical nature of this effect, we consider a simple harmonic oscillator, at rest for \( t < 0 \), driven by an oscillatory force for \( t > 0 \).

\[
\frac{d^2x}{dt^2} + \omega^2 x = A \cos(\Omega t) .
\]  

(47)

The general solution is given by the superposition of the free and forced solutions, see Appendix 1:

\[
x(t) = \frac{A}{\omega^2 - \Omega^2} \left[ \cos(\Omega t) - \cos(\omega t) \right] .
\]  

(48)

Let us assume that the external force is driving a particle population characterized by a spread of natural frequency of oscillation around a mean value \( \omega \). Furthermore, leave the forcing frequency \( \Omega \) inside the spectrum so that \( \delta \equiv \Omega - \omega \ll \omega \).

The motion of a given particle in the bunch can be approximated by:

\[
x(t) \approx \frac{At}{2\omega_x} \sin \left( \frac{\delta}{2} t \right) .
\]  

(49)

Let us observe two particles in the bunch, one with \( \delta = 0 \), and the other with \( \delta \neq 0 \). Both are at rest, and at \( t = 0 \) they start to oscillate with the same amplitude and phase (coherency). However, while the amplitude of the former charge grows indefinitely (driven at resonance), the latter reaches a maximum amplitude (beating of two close frequencies). The system of the particle has lost coherency at the time when the beating amplitude is maximum, i.e. for \( t = \pi/\delta \).

At any time \( t^* \), only those oscillators inside the bandwidth \( |\delta| < \pi / t^* \), oscillate coherently. The longer we wait, the narrower the coherent bandwidth and therefore the smaller the number of ‘coherent’ particles.

4.2 Amplitude of oscillations

At any instant we can divide the bunch population into two groups: coherent particles, oscillating all together with an amplitude growing linearly with time; and ‘incoherent’ particles, with different phases, and a saturated amplitude of oscillation.

It is interesting that, although the amplitude of the coherent oscillators grows linearly with time, the average amplitude of the whole system remains bounded, as the number of coherent particles decreases inversely with time

\[
\left< x(t) \right>_{\text{max}} = \frac{1}{N} \left[ \sum_{\text{coh}} x(t) + \sum_{\text{incoh}} x(t) \right]_{\text{max}} .
\]  

(50)
Consider the time when the coherent particles have the maximum amplitude. The amplitudes of the incoherent particles, being uncorrelated, have a zero average. For the coherent particles we have

\[
\langle x(t) \rangle_{\text{max}} = \frac{N_{\text{coh}}}{N} x_{\text{max}}(t) = \frac{N_{\text{coh}}}{N} A \frac{t}{2\omega_i} .
\]  

(51)

On the other hand, the number of oscillators keeping coherency decreases with time

\[
N_{\text{coh}} = \frac{N}{\Delta \omega} \frac{\pi}{t} \Rightarrow \langle x(t) \rangle_{\text{max}} = \frac{\pi}{\Delta \omega} \frac{A}{2\omega_i} .
\]  

(52)

4.3 Energy of the system

What happens to the energy of the system? In this case we distinguish the coherent and incoherent particles. The energy of the coherent particles has quadratic growth with time, while the energy of the incoherent particles is bounded. In this case, although the number of coherent oscillators decreases with time, the total energy still grows linearly

\[
E(t) = E_{\text{coh}}(t) + E_{\text{incoh}}(t)
\]

\[
E_{\text{coh}}(t) = N_{\text{coh}} \left[ \frac{1}{2} k x_{\text{coh}}^2(t) \right] = \frac{\pi}{2} \frac{N A^2}{\omega_i^2} \left( \frac{1}{4\Delta \omega} \right)^2 t .
\]

(53)

(54)

To conclude, when a force drives this kind of system, initially the whole system follows the external force. Thereafter, fewer and fewer particles are driven at the resonance. The result is that although the system absorbs energy, the average amplitude remains bounded.

This mechanism also works when the driving force is produced by the bunch itself. To make the coherent instability start, the rise time of the instability has to be shorter than the ‘de-coherency’ time of the bunch: \( \tau_{\text{inst}} < \tau_{\text{decoh}} = 2\pi / \Delta \omega \).
**APPENDIX A – DRIVEN OSCILLATORS**

Consider an harmonic oscillator with natural frequency \( \omega \), with an external excitation at frequency \( \Omega \)

\[
\ddot{x} + \omega^2 x = A \cos(\Omega t) .
\]

The general solution is

\[
x(t) = x^{\text{free}}(t) + x^{\text{driven}}(t)
\]

\[
\cos(\Omega t) \Rightarrow e^{i\Omega t}
\]

\[
x^{\text{free}}(t) = x_m^f e^{i\omega t}
\]

\[
x^{\text{driven}}(t) = \tilde{x}_m^d e^{\Omega t} .
\]

The driven solution (steady state) is found by direct substitution in the differential equation

\[
(\omega^2 - \Omega^2) \tilde{x}_m^d e^{\Omega t} = A e^{i\Omega t} \Rightarrow x^{\text{driven}}(t) = \frac{A}{(\omega^2 - \Omega^2)} e^{i\Omega t} .
\]

The general solution has to satisfy the initial condition at \( t = 0 \). In our case we assume that the oscillator is at rest for \( t = 0 \)

\[
x^{\text{free}}(t = 0) = -x^{\text{driven}}(t = 0)
\]

\[
\tilde{x}_m^f = -\frac{A}{\omega^2 - \Omega^2}
\]

thus we get

\[
x(t) = \frac{A}{\omega^2 - \Omega^2} \left[ e^{i\Omega t} - e^{i\omega t} \right] .
\]

taking only the real part

\[
x(t) = \frac{A}{\omega^2 - \Omega^2} \left[ \cos(\Omega t) - \cos(\omega t) \right] .
\]

This expression is suitable for deriving the response of the oscillator driven at resonance or at very close frequency: \( \omega = \Omega + \delta \), with \( \delta \to 0 \). Defining: \( \tilde{\omega} = (\omega + \Omega) / 2 \), equivalent to \( \omega = \tilde{\omega} + \delta / 2 \) or \( \Omega = \tilde{\omega} - \delta / 2 \) the solution is given by

\[
x(t) = \frac{A}{2 \tilde{\omega} \delta} \left\{ \cos(\tilde{\omega} t) \cos(\delta \omega / 2) + \sin(\tilde{\omega} t) \sin(\delta t / 2) \right\} +
\]

\[
+ \left[ \cos(\tilde{\omega} t) \cos(\delta t / 2) + \sin(\tilde{\omega} t) \sin(\delta t / 2) \right] \}
\]

that is

\[
x(t) = \frac{A}{2 \tilde{\omega}} \sin(\tilde{\omega} t) \sin(\delta t / 2) \equiv \frac{A t}{2 \tilde{\omega}} \sin(\tilde{\omega} t) \frac{\sin(\delta t / 2)}{\delta t / 2}
\]

with the limit

\[
\lim_{\delta \to 0} x(t) = \frac{A t}{2 \tilde{\omega}} \sin(\tilde{\omega} t) .
\]
APPENDIX B – POWER RADIATED BY A BUNCH PASSING THROUGH A TAPER

In the case of a charge distribution, and \( \gamma \rightarrow \infty \), the electric field lines are perpendicular to the direction of motion and travel with the charge [6], as shown in Fig. B.1. In other words, the field-map does not change during the charge-flight, as long as the trajectory is parallel to the pipe axis. Under this condition the transverse field intensity can be computed in the static case, applying the Gauss and Ampere laws:

\[
\oint_S E \cdot dS = \int_V \rho dV, \quad \oint_S B \cdot dl = \mu_0 \oint_S J \cdot ndS \tag{B.1}
\]

Let us consider a cylindrical beam of radius \( a \) of current \( I \), with uniform charge density \( \rho = I / \pi a^2 v \) and current density \( J = I / \pi a^2 \), propagating with relativistic speed \( v = \beta c \) along the \( z \) axis of a cylindrical, perfectly conducting pipe of radius \( b \), as shown in Fig. B.1.

![Fig. B.1: Cylindrical bunch of radius \( a \) propagating inside a cylindrical, perfectly conducting pipe of radius \( b \)](image)

By applying the relations (B.1) one can obtain for the radial component of the electric field

\[
E_r = \frac{I}{2 \pi \varepsilon_0 a^2 v} r \quad \text{for} \quad r \leq a
\]

\[
E_r = \frac{I}{2 \pi \varepsilon_0 v} \quad \text{for} \quad r > a
\]

and the relation \( B_\phi = \frac{\beta}{c} E_r \) holds.

The electrostatic potential satisfying the boundary condition \( \varphi(b) = 0 \) is given by:

\[
\varphi(r,z) = \int_r^b E_r(r',z) dr' = \begin{cases} \frac{I}{4 \pi \varepsilon_0 v} \left( 1 + 2 \ln \frac{b}{a} - \frac{r^2}{a^2} \right) & \text{for} \quad r \leq a \\ \frac{I}{2 \pi \varepsilon_0 v} \ln \frac{b}{r} & \text{for} \quad a \leq r \leq b \end{cases}
\]

How can a perturbation of the boundary conditions affect the beam dynamics? Consider the following example: a smooth transition of length \( L \) (taper) from a beam pipe of radius \( b \) to a larger beam pipe of radius \( d \) is experienced by the beam [6]. To satisfy the boundary condition of a perfectly conducting pipe in the tapered region, the field lines are bent as shown in Fig. B.2. Therefore there must be a longitudinal \( E_z(r,z) \) field component in the transition region.

A test particle running outside the beam charge distribution along the transition of length \( L \) will experience a voltage given by [12]:

357
\[ V = - \int_{z}^{z+L} E_z(r, z') dz' = - \left[ \varphi(r, z + L) - \varphi(r, z) \right] = -\frac{I}{2\pi\varepsilon_0 v} \ln \frac{d}{b} \]

that is decelerating if \( d > b \). The power lost by the beam in order to sustain the induced voltage is given by

\[ P_{\text{lost}} = VI = \frac{I^2}{2\pi\varepsilon_0 v} \ln \frac{d}{b}. \quad (B.2) \]

**Fig. B.2:** Smooth transition of length \( L \) (taper) from a beam pipe of radius \( b \) to a larger beam pipe of radius \( d \)

It means that for \( d > b \) the power is deposited to the energy of the fields: moving from left to right of the transition, the beam induces the fields in the additional room around the bunch bunch (i.e. in the region \( b < r < d, \ 0 < z < l_0 \)) at the expense of the only available energy source, that is the kinetic energy of the beam itself.

**Fig. B.3:** During the beam propagation in the taper additional electromagnetic power flow is required to fill up the new available room

To verify this interpretation, compute the electromagnetic power radiated by the beam to fill up the additional room available, see Fig. B.3. Integrating the Poynting vector through the surface \( \Delta S = \pi \left( d^2 - b^2 \right) \) representing the additional power passing through the right part of the beam pipe [12], one obtains

\[ P_{\text{em}} = \int_{\Delta S} \left( \frac{1}{\mu} \hat{E} \times \hat{B} \right) \cdot n dS = \int_{b}^{d} \frac{E_r B_\theta}{\mu} 2\pi r dr = \frac{I^2}{2\pi\varepsilon_0 v} \ln \frac{d}{b} \]

exactly the same expression of Eq. (B.2). Notice that if \( d < b \) the beam gains energy. If \( d \rightarrow \infty \) the power goes to infinity, an unphysical result like this is nevertheless consistent with the original assumption of an infinite energy beam (\( \gamma \rightarrow \infty \)).
References

Abstract
The performance of particle colliders is usually quantified by the beam energy and the luminosity. We derive the expressions for the luminosity in case of bunched beams in terms of the beam parameters and the geometry. The implications of additional features such as crossing angle, offsets and hourglass effect on the luminosity are calculated. Important operational aspects like integrated luminosity, space and time structure of interactions etc. are discussed.

The measurement of luminosity for $e^+e^-$ as well as hadron colliders and the methods for the calibration of the absolute luminosity are described.

1 Introduction
In particle physics experiments the energy available for the production of new effects is the most important parameter. The required large centre of mass energy can only be provided with colliding beams where little or no energy is lost in the motion of the centre of mass system (cms). Besides the energy the number of useful interactions (events), is important. This is especially true when rare events with a small production cross section $\sigma$ are studied. The quantity that measures the ability of a particle accelerator to produce the required number of interactions is called the luminosity and is the proportionality factor between the number of events per second $dR/dt$ and the cross section $\sigma$:

$$\frac{dR}{dt} = L \cdot \sigma.$$  \hfill (1)

The unit of the luminosity is therefore $\text{cm}^{-2}\text{s}^{-1}$.

In this lecture we shall first give the main arguments which lead to a general expression for the luminosity and derive the formula for basic cases. Additional complications such as crossing angle and offset collisions are added to the calculation. Special effects such as the hourglass effect and the consequences of different beam profiles are estimated from the generalized expression.

Besides the absolute value of the luminosity, other issues are important for physics experiments, such as the integrated luminosity and the space and time structure of the resulting interactions.

In the final section we shall discuss the measurement of luminosity in both, $e^+e^-$ as well as hadron colliders.

2 Why colliding beams?
The kinematics of a particle with mass $m$ can be expressed by its momentum $\vec{p}$ and energy $E$ which form a four-vector $p = (E, \vec{p})$ whose square $p^2$ is (with the appropriate norm):

$$p^2 = E^2 - \vec{p}^2 = m^2.$$  \hfill (2)

In the collision of two particles of masses $m_1$ and $m_2$ the total centre of mass energy can be expressed in the form

$$(p_1 + p_2)^2 = E_{cm}^2 = (E_1 + E_2)^2 - (\vec{p}_1 + \vec{p}_2)^2.$$  \hfill (3)

This is the available energy for physics experiments.
In the case of a collider where the collision point is at rest in the laboratory frame (i.e. \( \vec{p}_1 = -\vec{p}_2 \)), the centre of mass energy becomes:

\[
E_{\text{cm}}^2 = (E_1 + E_2)^2 .
\] (4)

When one particle is at rest, i.e. in the case of so-called fixed target experiments, (i.e. \( \vec{p}_2 = 0 \)), we get:

\[
E_{\text{cm}}^2 = (m_1^2 + m_2^2 + 2m_2 E_{\text{lab}}^1) .
\] (5)

A comparison for different types of collisions is made in Table 1. From this table it is rather obvious why colliding beams are necessary to get the high centre of mass energies required for particle physics experiments.

### Table 1: Centre of mass energy for different types of collisions.

<table>
<thead>
<tr>
<th></th>
<th>( E_{\text{cm}} ) as collider (GeV)</th>
<th>( E_{\text{cm}} ) with fixed target (GeV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p ) on ( p ) (7000 on 7000 GeV)</td>
<td>14000</td>
<td>114.6</td>
</tr>
<tr>
<td>( e ) on ( e ) (100 on 100 GeV)</td>
<td>200</td>
<td>0.32</td>
</tr>
<tr>
<td>( e ) on ( p ) (30 on 920 GeV)</td>
<td>235</td>
<td>7.5</td>
</tr>
</tbody>
</table>

3 Computation of luminosity

3.1 Fixed target luminosity

In order to compute a luminosity for fixed target experiment, we have to take into account the properties of both, the incoming beam and the stationary target. The basic configuration is shown in Fig. 1. The incoming beam is characterized by the flux \( \Phi \), i.e. the number of particles per second. When the target is homogeneous and larger than the incoming beam, the distribution of the latter is not important for the luminosity.

The target is described by its density \( \rho_T \) and its length \( l \). With a definition of the luminosity like:

\[
\mathcal{L}_{FT} = \Phi \rho_T l
\] (6)

Fig. 1: Schematic view of a fixed target collision.
we write the interaction rate
\[ \frac{dR}{dt} = \Phi \rho_T \cdot \sigma_p = L_{FT} \cdot \sigma_p \]  

as desired.

3.2 Colliding beams luminosity

In the case of two colliding beams, both beams serve as target and "incoming" beam at the same time. Obviously the beam density distribution is now very important and the generalization of the above expression leads to the convolution of the 3-D distribution functions. We treat the case of bunched beams,

\[ \frac{dR}{dt} = L \sigma_p \]

but it can easily be extended to unbunched beams or other concepts like very long bunches. A schematic picture is shown in Fig. 2. Since the two beams are not stationary but moving through each other, the overlap integral depends on the longitudinal position of the bunches and therefore on the time as they move towards and through each other. For our integration we use the distance of the two beams to the central collision point \( s_0 = c \cdot t \) as the "time" variable (see Fig. 2). A priori the two beams have different distribution functions and different number of particles in the beams.

The overlap integral which is proportional to the luminosity \( (L) \) we can then write as:

\[ L \propto K \int \int \int \int_{-\infty}^{+\infty} \rho_1(x, y, s, s_0) \rho_2(x, y, s, s_0) dx dy ds ds_0. \]  

(8)

Here \( \rho_1(x, y, s, s_0) \) and \( \rho_2(x, y, s, s_0) \) are the time dependent beam density distribution functions. We assume, that the two bunches meet at \( s_0 = 0 \). Because the beams are moving against each other, we have to multiply this expression with a kinematic factor [1]:

\[ K = \sqrt{(v_1^2 - v_2^2)^2 - (v_1^* \times v_2^*)^2/c^2}. \]  

(9)

In the next step we assume head-on collisions \( (v_1^* = -v_2^*) \) and that all densities are uncorrelated in all planes. In that case we can factorize the density distributions and get for the overlap integral:

\[ L = 2N_1 N_2 f N_b \int \int \int \int_{-\infty}^{+\infty} \rho_{1x}(x) \rho_{1y}(y) \rho_{1s}(s - s_0) \rho_{2x}(x) \rho_{2y}(y) \rho_{2s}(s + s_0) dx dy ds ds_0. \]  

(10)

We have completed the formula with the beam properties necessary to calculate the value of the luminosity: \( N_1 \) and \( N_2 \) are the intensities of two colliding bunches, \( f \) is the revolution frequency and \( N_b \) is the number of bunches in one beam.

To evaluate this integral one should know all distributions. An analytical calculation is not always possible and a numerical integration may be required. However in many cases the beams follow "reasonable" profiles and we can obtain closed solutions.
4 Luminosity of Gaussian beams colliding head-on

Often it is fully justified to assume Gaussian distributions. The luminosity is determined by the overlap of the core of the distributions and the tails give practically no contribution to the luminosity. We shall come back to this point in a later section.

For the first calculation we assume Gaussian profiles in all dimensions of the form:

\[ \rho_{i\pm}(z) = \frac{1}{\sigma_{i\pm} \sqrt{2\pi}} \exp\left( -\frac{z^2}{2\sigma_{i\pm}^2} \right) \text{ where } i = 1, 2, \quad z = x, y, \]

\[ \rho_s(s \pm s_0) = \frac{1}{\sigma_{s \pm} \sqrt{2\pi}} \exp\left( -\frac{(s \pm s_0)^2}{2\sigma_{s \pm}^2} \right). \]

Furthermore we assume equal beams, i.e.: \( \sigma_{1x} = \sigma_{2x}, \sigma_{1y} = \sigma_{2y}, \sigma_{1s} = \sigma_{2s} \).

Next we assume the number of particles per bunch \( N_1 \) and \( N_2 \), a revolution frequency of \( f \) and the number of bunches we call \( N_b \). In the case of exactly head-on collisions of bunches travelling almost at the speed of light, the kinematic factor becomes 2.

Using this in equation (10) we get the first integral:

\[
\mathcal{L} = \frac{2 \cdot N_1 N_2 f N_b}{(\sqrt{2\pi})^6 \sigma_{x}^2 \sigma_{y}^2 \sigma_{s}^2} \int \int \int \int e^{-\frac{x^2}{\sigma_{x}^2}} e^{-\frac{y^2}{\sigma_{y}^2}} e^{-\frac{z^2}{\sigma_{s}^2}} dxdydzdsds_0
\]

integrating over \( s \) and \( s_0 \), using the well known formula:

\[ \int_{-\infty}^{\infty} e^{-at^2} dt = \sqrt{\pi/a} \]

we get a first intermediate result:

\[
\mathcal{L} = \frac{2 \cdot N_1 N_2 f N_b}{8(\sqrt{\pi})^4 \sigma_{x}^2 \sigma_{y}^2} \int e^{-\frac{x^2}{\sigma_{x}^2}} -\frac{y^2}{\sigma_{y}^2} dxdy.
\]

Finally, after integration over \( x \) and \( y \):

\[ \Rightarrow \mathcal{L} = \frac{N_1 N_2 f N_b}{4\pi \sigma_{x} \sigma_{y}}. \]

This is the well-known expression for the luminosity of two Gaussian beams colliding head-on. It shows how the luminosity depends on the number of particles per bunch and the beam sizes. This reflects the 2-dimensional target charge density we have seen in the evaluation of the fixed target luminosity.

For the more general case of: \( \sigma_{1x} \neq \sigma_{2x}, \sigma_{1y} \neq \sigma_{2y} \), but still assuming approximately equal bunch lengths \( \sigma_{1s} \approx \sigma_{2s} \) we get a modified formula:

\[
\mathcal{L} = \frac{N_1 N_2 f N_b}{2\pi \sqrt{\sigma_{1x}^2 + \sigma_{2x}^2} \sqrt{\sigma_{1y}^2 + \sigma_{2y}^2}}.
\]

This formula is easy to verify and also straightforward to extend to other cases. Here it is worth to mention that the luminosity does not depend on the bunch length \( \sigma_{s} \). This is due to the assumption of uncorrelated density distributions.

5 Examples

In Table 2 we give some examples of different colliders and their luminosity and other relevant parameters. One may notice the very different interaction rate, in particular between hadron colliders and high energy lepton colliders. This is due to the small total cross section of \( e^+ e^- \) interactions. Furthermore, since so-called B-factories such as PEP and KEKB operate near or on resonances, the interaction rate varies very strongly with the precise energy. Therefore we write the term NA in the table.
Table 2: Example of different colliders. We show the energy, luminosity, beam sizes and interaction rate for a comparison.

<table>
<thead>
<tr>
<th></th>
<th>Energy</th>
<th>$\mathcal{L}$</th>
<th>rate</th>
<th>$\sigma_x/\sigma_y$</th>
<th>Particles per bunch</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPS (p$\bar{p}$)</td>
<td>315x315</td>
<td>$6 \times 10^{30}$</td>
<td>$4 \times 10^6$</td>
<td>60/30</td>
<td>$\approx 10 \times 10^{10}$</td>
</tr>
<tr>
<td>Tevatron (p$\bar{p}$)</td>
<td>1000x1000</td>
<td>$50 \times 10^{30}$</td>
<td>$4 \times 10^6$</td>
<td>30/30</td>
<td>$\approx 30/8 \times 10^{10}$</td>
</tr>
<tr>
<td>HERA (e$^+$ p)</td>
<td>30x920</td>
<td>$40 \times 10^{30}$</td>
<td>40</td>
<td>250/50</td>
<td>$\approx 3/7 \times 10^{10}$</td>
</tr>
<tr>
<td>LHC (pp)</td>
<td>7000x7000</td>
<td>$10000 \times 10^{30}$</td>
<td>$10^9$</td>
<td>17/17</td>
<td>11 $10^{10}$</td>
</tr>
<tr>
<td>LEP (e$^+$ e$^-$)</td>
<td>105x105</td>
<td>$100 \times 10^{30}$</td>
<td>$\leq 1$</td>
<td>200/2</td>
<td>$\approx 5 \times 10^{11}$</td>
</tr>
<tr>
<td>PEP (e$^+$ e$^-$)</td>
<td>9x3</td>
<td>$3000 \times 10^{30}$</td>
<td>NA</td>
<td>150/5</td>
<td>$\approx 2/6 \times 10^{10}$</td>
</tr>
<tr>
<td>KEKB (e$^+$ e$^-$)</td>
<td>8x3.5</td>
<td>$10000 \times 10^{30}$</td>
<td>NA</td>
<td>77/2</td>
<td>$\approx 1.3/1.6 \times 10^{10}$</td>
</tr>
</tbody>
</table>

6 Additional complications in real machines

So far we have assumed ideal head-on collisions of bunches where the particle densities in the three dimensions are uncorrelated. In practice, we have to include additional effects in our computations, some of the most important are:

- Crossing angle
- Collision offset (wanted or unwanted)
- Hour glass effect
- Non-Gaussian beam profiles
- Non-zero dispersion at collision point
  - $\delta \beta^*/\delta s = \alpha^* \neq 0$

Crossing angles are often used to avoid unwanted collisions in machines with many bunches (e.g. LHC, CESR, KEKB). Such crossing angles can have important consequences for beam-beam effects [2] but also affect the luminosity. When beams do not collide exactly head-on but with a small offset, the luminosity is reduced. Such an offset can be wanted (e.g. to reduce luminosity or during measurements) or unwanted, for example as a result of beam-beam effects [2]. The so-called hourglass effect is a geometrical effect which includes a dependence of the transverse beam sizes on the longitudinal position and therefore violates our previous assumption of uncorrelated particles densities. When the beam profiles deviate from a Gaussian function, we may have to apply some correction factors and when the dispersion at the interaction point is not zero, the effective beam sizes are increased, leading to a smaller luminosity. In case of optical imperfections the collision point may not be at the minimum of the betatron function $\beta^*$, i.e. at the waist, but slightly displaced with implications for the effective beam sizes.

Some of the most important of these additional effects we shall investigate in the following sections.

6.1 Crossing angles

A very prominent collider with a crossing angle is the LHC presently under construction at CERN. In the LHC one has almost 3000 closely spaced bunches and to avoid numerous unwanted interactions, the two beams collide at a total crossing angle of around $\approx 300 \mu$rad. The Fig.3 shows a schematic illustration of the collision region. Colliders with unbunched, i.e. coasting beams such as the ISR need a sizeable crossing angle to confine the interaction region (e.g. $\approx 18^\circ$ at the ISR).

In the following we shall assume without loss of generality that the crossing angle is in the horizontal plane. The overlap integrals are evaluated in the x and y coordinate system and therefore we have
Fig. 3: Schematic view of two bunches colliding at a finite crossing angle.

Fig. 4: Rotated reference system for collisions at a finite crossing angle.

to transform our bunches into the proper system. The geometry of a collision at a crossing angle $\Phi$ is shown in Fig.4. To make the treatment more symmetric, we have assumed that the total crossing angle is made up by two rotations $\Phi/2$ and $-\Phi/2$ each of the two beams in the x-s plane (see Fig.4).

To compute the integral we have to transform $x$ and $s$ to new, rotated coordinates which are now different for the two beams:

$$x_1 = x \cos \frac{\phi}{2} - s \sin \frac{\phi}{2}, \quad s_1 = s \cos \frac{\phi}{2} + x \sin \frac{\phi}{2}$$

$$x_2 = x \cos \frac{\phi}{2} + s \sin \frac{\phi}{2}, \quad s_2 = s \cos \frac{\phi}{2} - x \sin \frac{\phi}{2}$$

(18)

(19)

The overlap integral becomes:

$$L = 2 \cos^2 \frac{\phi}{2} N_1 N_2 f N_b \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \rho_{1x}(x_1)\rho_{1y}(y_1)\rho_{1s}(s_1 - s_0)$$

$$\rho_{2x}(x_2)\rho_{2y}(y_2)\rho_{2s}(s_2 + s_0) dx dy ds ds_0 .$$

(20)

The factor $2 \cos^2 \frac{\phi}{2}$ is the kinematic factor when the two velocities of the bunches are not collinear (from Eq. (9)).
After the integration over $y$ and $s_0$, using the formula:
\[
\int_{-\infty}^{+\infty} e^{-(at^2+bt+c)} \, dt = \sqrt{\pi/a} \cdot e^{\frac{b^2}{4a}}
\]
we get:
\[
\mathcal{L} = \frac{N_1 N_2 f N_b}{8 \pi^2 \sigma_x \sigma_y} 2 \cos^2 \frac{\phi}{2} \int e^{-\frac{2 \cos^2\left(\phi/2\right)+2 \sin^2\left(\phi/2\right)}{\sigma_x^2}} e^{-\frac{2 \sin^2\left(\phi/2\right)+2 \cos^2\left(\phi/2\right)}{\sigma_y^2}} \, dx \, ds.
\]
We make the following approximations: since both $x$ and $\sin(\phi/2)$ are small, we drop all terms of the type $\sigma_x^2 \sin^4(\phi/2)$ or $x^k \sin^4(\phi/2)$ for all $k+1 \geq 4$ and approximate $\sin(\phi/2) \approx \tan(\phi/2)$ by $\phi/2$. After the final integrations we get for the luminosity an expression of the form:
\[
\mathcal{L} = \frac{N_1 N_2 f N_b}{4 \pi \sigma_x \sigma_y} \cdot S.
\]
This looks exactly like the well known formula we have derived already, except for the additional factor $S$, the so-called luminosity reduction factor which can be written as:
\[
\frac{1}{\sqrt{1 + \left(\frac{\sigma_x}{\sigma_y} \tan \frac{\phi}{2}\right)^2}} \cdot \frac{1}{\sqrt{1 + \left(\frac{\sigma_x}{\sigma_y} \frac{\sigma_y}{\sigma_x}\phi\right)^2}}.
\]
For small crossing angles and $\sigma_x \gg \sigma_{x,y}$ we can simplify the formula to:
\[
S = \frac{1}{\sqrt{1 + \left(\frac{\sigma_x}{\sigma_y} \tan \frac{\phi}{2}\right)^2}} \approx \frac{1}{\sqrt{1 + \left(\frac{\sigma_x}{\sigma_y} \frac{\sigma_y}{\sigma_x}\phi\right)^2}}.
\]
A popular interpretation of this result is to consider it a correction to the beam size and to introduce an ”effective beam size” like:
\[
\sigma_{x,eff} = \sigma \cdot \sqrt{1 + \left(\frac{\sigma_x}{\sigma_y}\frac{\sigma_y}{\sigma_x}\phi\right)^2}.
\]
The effective beam size can then be used in the standard formula for the beam size in the crossing plane. This concept of an effective beam size is interesting because it also applies to the calculation of beam-beam effects of bunched beams with a crossing angle [5].

As an example we use the parameters of the LHC. The number of particles per bunch is $1.15 \times 10^{11}$, the beam sizes in the two planes $\approx 16.7 \ \mu m$, the bunch length $\sigma_x = 7.7 \ cm$ and the total crossing angle $\Phi = 285 \ \mu rad$. With the revolution frequency of $11.245 \ kHz$ and 2808 bunches, we get for the head-on luminosity $1.2 \times 10^{34} \ cm^{-2} s^{-1}$. For the luminosity reduction factor $S$ we get 0.835 and the final LHC luminosity with a crossing angle becomes $\approx 1.0 \times 10^{34} \ cm^{-2} s^{-1}$.

### 6.2 Crossing angles and offset beams

A modification of the previous scheme is needed when the beams do not collide head-on, but with a small transverse offset. In order to be general, we shall treat the case with crossing angle and offsets. The Fig. 5 shows the modified geometry when we have the same crossing angle as before, but beam 1 is displaced by $d_1$ and beam 2 is displaced by $d_2$ with respect to their reference orbits.

The coordinate transformations are now:
\[
x_1 = d_1 + x \cos \frac{\phi}{2} - s \sin \frac{\phi}{2}, \quad \quad s_1 = s \cos \frac{\phi}{2} + x \sin \frac{\phi}{2},
\]
\[
x_2 = d_2 + x \cos \frac{\phi}{2} + s \sin \frac{\phi}{2}, \quad \quad s_2 = s \cos \frac{\phi}{2} - x \sin \frac{\phi}{2}.
\]
Following the previous strategy and approximations for the integration, we get after integrating $y$ and $s_0$:

\[
\mathcal{L} = \frac{N_1 N_2 f N_b}{8\pi^2 \sigma_x \sigma_y^2} 2 \cos^2 \frac{\phi}{2} \int e^{\frac{x^2 \cos^2(\phi/2) + x^2 \sin^2(\phi/2)}{\sigma_x^2}} e^{\frac{x^2 \sin^2(\phi/2) + x^2 \cos^2(\phi/2)}{\sigma_y^2}} \times e^{\frac{d_1^2 + d_2^2 + 2(d_1 + d_2) s \cos(\phi/2) - 2(d_2 - d_1) s \sin(\phi/2)}{2\sigma_x^2}} dx ds .
\]  

(29)

After the integration over $x$ we obtain:

\[
\mathcal{L} = \frac{N_1 N_2 f N_b}{8\pi \sigma_x \sigma_y} 2 \cos^2 \frac{\phi}{2} \int_{-\infty}^{+\infty} W e^{-\frac{(As^2 + 2Bs)}{\sigma_x \sigma_y}} ds
\]

(30)

with:

\[
A = \frac{\sin^2 \frac{\phi}{2}}{\sigma_x^2} + \frac{\cos^2 \frac{\phi}{2}}{\sigma_y^2}, \quad B = \frac{(d_2 - d_1) \sin(\phi/2)}{2\sigma_x^2}
\]

(31)

and

\[
W = e^{-\frac{1}{4\sigma_x^2}(d_2-d_1)^2}.
\]

(32)

We can re-write the luminosity with three correction factors:

\[
\mathcal{L} = \frac{N_1 N_2 f N_b}{4\pi \sigma_x \sigma_y} \cdot W \cdot e^{\frac{\mu^2}{\pi}} \cdot S .
\]

(33)

This factorization enlightens the different contributions and allows straightforward calculations. The last factor $S$ is the already calculated luminosity reduction factor for a crossing angle. One factor $W$ reduces the luminosity in the presence of beam offsets and the factor $e^{\frac{\mu^2}{\pi}}$ is only present when we have a crossing angle and offsets simultaneously.

### 6.3 Hourglass effect

So far we have assumed uncorrelated beam density functions in the transverse and longitudinal planes. In particular, we have assumed that the transverse beam sizes are constant over the whole collision regions. However, since the $\beta$-functions have their minima at the collision point and increase with the distance this is not always a good approximation. In a low-$\beta$ region the $\beta$-function varies with the distance $s$ to
In our formulae we have to replace $\sigma$ by $\sigma(s)$ and get a more general expression for the luminosity:

$$L = \frac{N_1 N_2 f N_b}{8\pi \sigma_x^{\ast} \sigma_y^{\ast}} \frac{2 \cos \frac{\phi}{2}}{\sqrt{\pi \sigma_s}} \int_{-\infty}^{+\infty} \frac{e^{-s^2 A}}{1 + \left( \frac{s}{\beta^*} \right)^2} ds$$

(35)

with

$$A = \frac{\sin^2 \frac{\phi}{2}}{(\sigma_x^{\ast})^2[1 + \left( \frac{s}{\beta^*} \right)^2]} + \frac{\cos^2 \frac{\phi}{2}}{(\sigma_y^{\ast})^2}.$$  

(36)

Usually it is difficult to compute this integral analytically and it has to be evaluated by numerical integration.

To estimate the importance and relevance of this effect, we shall use the parameters of the LHC, i.e., $N_1 = N_2 = 1.15 \times 10^{11}$ particles/bunch, 2808 bunches per beam, a revolution frequency of $f = 11.2455$ kHz, and a crossing angle of $\phi = 285 \mu$rad. The nominal $\beta$-functions at the interaction point are $\beta_x^{\ast} = \beta_y^{\ast} = 0.55$ m, leading to beam sizes of $\sigma_x^{\ast} = \sigma_y^{\ast} = 16.7 \mu$m, and we use a r.m.s. bunch length of $\sigma_s = 7.7$ cm.

In the simplest case of a head on collision we get for the luminosity $L = 1.200 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$.

**Fig. 6:** Schematic illustration of the hourglass effect. $\beta(s)$ is plotted for two different values of $\beta^{\ast}$.
The effect of the crossing angle we can estimate by the evaluation of the factor $S$ and get:

$$L = 1.000 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}.$$  

When we further include the hourglass effect we get:

$$L = 0.993 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}.$$  

While the effect of the crossing angle is very sizeable ($S = 0.835$), the further reduction by the hourglass effect is small, at least for the nominal LHC parameters. For smaller $\beta$-functions at the interaction point this may not be the case.

### 6.4 Sensitivity to beam profiles

So far we have assumed Gaussian distribution functions in all dimensions. However this is not always the case, in particular for hadron beams, and it is necessary to evaluate the validity of this assumption, i.e., the importance for the derived results. Since it is mainly the core of the beam distribution which contributes to the luminosity, we can hope that the r.m.s. as a measure of the beam size (and therefore implying a Gaussian profile) is a good approximation.

To make a quantitative study, we can compute the luminosity for a flat beam using the complete overlap integral and compare to the simplified calculation, i.e. compute the r.m.s. and use them in the standard formula.

We assume flat distributions of the form:

$$\rho_1(x, y) = \rho_2(x, y) = \frac{1}{2a}, \quad \text{for} \quad -a \leq z \leq a, \quad z = x, y$$  \hspace{1cm} (37)

and calculate the r.m.s. in $x$ and $y$:

$$< (x, y)^2 > = \int_{-\infty}^{+\infty} (x, y)^2 \cdot \rho(x, y) \ dx dy .$$  \hspace{1cm} (38)

We compute the correct luminosity (without the constants which are equal for all cases) from the integral:

$$L = \int_{-\infty}^{+\infty} \rho_1(x, y) \rho_2(x, y) \ dx dy .$$  \hspace{1cm} (39)

The quantity $L \sqrt{< x^2 > < y^2 >}$ gives a measure for the quality of the approximation when it is compared to the same expression for a Gaussian beam. The astonishing result is that the error one makes with this approximation is only a few % [3]. The same result holds for other "reasonable" distributions such as parabolic or cosine-like [3].

### 7 Other luminosity issues

There are further issues related to the luminosity which are important for the experiments, such as:

- Integrated luminosity
- Time structure of interactions
- Space structure of interactions

The geometry of the interaction regions as well as some basic parameters entering the standard luminosity formulae are very important for the above issues and may need reconsideration to trade off between the different requirements.
7.1 Integrated luminosity

7.1.1 Definition of integrated luminosity

The maximum luminosity, and therefore the instantaneous number of interactions per second, is very important, but the final figure of merit is the so-called integrated luminosity:

\[ \mathcal{L}_{\text{int}} = \int_0^T \mathcal{L}(t') \, dt' \]  

because it directly relates to the number of observed events:

\[ \mathcal{L}_{\text{int}} \cdot \sigma_p = \text{number of events of interest} \, . \]  

The integral is taken over the sensitive time, i.e., excluding possible dead time. For an evaluation one needs a realistic model for the decay of the luminosity with time. Different possibilities exist and usually one assumes some behaviour (e.g., exponential) with a given lifetime \( \tau \):

\[ \mathcal{L}(t) \rightarrow \mathcal{L}_0 \exp \left( -\frac{t}{\tau} \right) \, . \]  

Contributions to this life time we have from the decay of beam intensity with time, the growth of the transverse emittance, increase of the bunch length etc. The advantage of assuming an exponential decay is that the contributions from different processes can be easily added. The differences between the different models is very small in practice.

7.1.2 Optimization of integrated luminosity

The aim of the operation of a collider must be to optimize the integrated luminosity. Two parts of the operation must be distinguished: the luminosity run with a lifetime \( \tau \) and the preparation time between two luminosity runs \( t_p \). The optimization problem is very similar to the challenge of a formula 1 racing team: the length of running with decreasing performance (slowing down with ageing tires) and the time needed to restore the performance (changing tires). The best strategy should minimize the overall time needed.

In principle, the knowledge of the preparation time allows an optimization of \( \mathcal{L}_{\text{int}} \).

If we assume an exponential decay of the luminosity \( \mathcal{L}(t) = \mathcal{L}_0 \cdot e^{t/\tau} \) we want to maximize the average luminosity \( \langle \mathcal{L} \rangle \):

\[ \langle \mathcal{L} \rangle = \frac{\int_{t_r}^{t_r + t_p} \mathcal{L}(t) \, dt}{t_r + t_p} = \mathcal{L}_0 \cdot \tau \cdot \frac{1 - e^{-t_r/\tau}}{t_r + t_p} \, . \]  

Here \( t_r \) is the length of a luminosity run and \( t_p \) the preparation time between two runs. Since \( t_r \) is a "free" parameter, i.e. can be chosen by the operation crew, we can optimize this expression and get a (theoretical) maximum for:

\[ t_r \approx \tau \cdot \ln(1 + \sqrt{2t_p/\tau + t_p/\tau}) \, . \]  

Assuming some parameters for the LHC [4]: \( t_p \approx 10h, \tau \approx 15h \), we get: \( \Rightarrow t_r \approx 15h \).

7.2 Luminous region and space structure of luminosity

In addition to the number of events, the space structure is important for the design and running of a particle physics experiment. The questions we asked are therefore:

- What is the density distribution of interaction vertices ?
- Which fraction of collisions occur \( \pm s \) from the interaction point ?
Fig. 7: Schematic illustration of luminous regions.

The answers depend on beam properties such as $\sigma_x$, $\sigma_y$, and $\sigma_s$ but also on the crossing angle $\phi$. This is very schematically indicated by the overlap regions in Fig. 7. Depending of the beam and machine parameters, this region can be very different, with important consequences for e.g., trigger system or pattern recognition. We evaluate:

\begin{align}
L_0 &= \int_{-\infty}^{+\infty} L(s')ds' \rightarrow L(s) = \int_{-s}^{+s} L(s')ds' \tag{45} \\
L(s) &= \left( \frac{N_1 N_2 f N_b}{8\pi \sigma^* x \sigma^* y} \right) \frac{2 \cos \frac{\phi}{2}}{\sqrt{\pi \sigma_s}} \sqrt{\frac{\pi}{A}} \text{erf} \left( \sqrt{A} s \right) \tag{46} \\
\text{For the integrated luminosity this becomes:} \\
L_{\text{int}}(s) &= \int_0^T \int_{-s}^{+s} L(s', t)ds' dt. \tag{47}
\end{align}

In order to evaluate this numerically, we use again LHC nominal parameters as above. The results of our calculations are shown in Tab.2. While practically all luminosity is seen at a distance of $\pm 12$ cm from the interaction point, about 20% is lost when only a region of $\pm 5.5$ cm is covered by the detector or the software. This does strongly depend on the crossing angle. A detailed examination of this property was done in [6].

7.3 Time structure of luminosity

In addition to the space structure, the time structure of the interactions is an important input for the setup of an experiment and even on the possible physics that can be studied.
### Table 3: Percentage of visible luminosity as a function of distance to interaction point.

<table>
<thead>
<tr>
<th>Integration Range</th>
<th>Percentage of Luminosity</th>
</tr>
</thead>
<tbody>
<tr>
<td>(s = \pm 12) cm</td>
<td>1.000</td>
</tr>
<tr>
<td>(s = \pm 8) cm</td>
<td>0.950</td>
</tr>
<tr>
<td>(s = \pm 7) cm</td>
<td>0.900</td>
</tr>
<tr>
<td>(s = \pm 6) cm</td>
<td>0.850</td>
</tr>
<tr>
<td>(s = \pm 5.5) cm</td>
<td>0.800</td>
</tr>
</tbody>
</table>

In the LHC the bunches cross every 25 ns and it can be calculated easily that for proton-proton collisions one has to expect \(\approx 20\) simultaneous interactions per bunch crossing. They must be digested by the detectors before the next bunch crossing occurs, a non trivial task for the experiments. Some physics studies that cannot be done with such an event pile up may require to run at a lower luminosity.

## 8 Luminosity Measurement

The knowledge of the luminosity is of vital interest for both, the experiments and the accelerator physicists. The most important issues are:

- Determine the cross section from counting rates
- Optimize the luminosity
- Measure machine parameters from luminosity

This list could easily be extended.

### 8.1 Relative Luminosity Measurement

Since the luminosity is directly proportional to the interaction rate, luminosity measurement usually consists of fast counting devices which provide such a signal. However, some of the challenges for such an instrument are:

- Must cover a large dynamic range: \(10^{27} \text{ cm}^{-2}\text{s}^{-1}\) to \(10^{34} \text{ cm}^{-2}\text{s}^{-1}\)
- Very fast, if possible for individual bunches
- Reproducible from run to run
- Should run with different machine conditions (e.g. with and without crossing angle, different optics \((\beta^*)\), etc.)
- Works for different types of particles (p, ions)
- Can be used in a feedback system to optimize the luminosity

When these requirements are fulfilled, the instrument will give a reliable signal that is proportional to the luminosity.
8.2 Absolute luminosity measurement
The relative signal has to be calibrated to deliver the absolute luminosity. We have already seen some effects that affect the absolute luminosity and therefore to a large extent the luminosity measurement. In particular the crossing angle and the luminous region are of importance since they have immediate implications for the geometrical acceptance of the instruments.

In principle one can determine the absolute luminosity when all relevant beam parameters are known, i.e. the bunch intensities, beam sizes (r.m.s. in case of unknown beam profiles) and the exact geometry. However the precise measurement of beam sizes is a challenge, in particular for hadron colliders when a non-destructive measurement is required. When the energy spread in the beams is large (e.g. some e⁺e⁻ colliders, a residual dispersion at the interaction point increases significantly the beam size and must be included.

There exist other methods which relate the counting rate to well known processes which can be used for calibration. We shall discuss several methods for both, lepton and hadron colliders.

8.3 Absolute luminosity - lepton colliders
Once the relative luminosity is known, a very precise method is to compare the counting rate to well known and calculable processes. In case of e⁺ e⁻ colliders these are electromagnetic processes such as elastic scattering (Bhabha scattering). The principle is shown in Fig. 8. Particle detectors are used to measure the trajectories at very small angles and with a coincidence of particles on both sides of the interaction point. For a precise measurement one has to go to very small angles since the elastic cross section σ_{el} has a strong dependence on the scattering angle (σ_{el} ∝ Θ^{-3}).

Furthermore, the cross section diminishes rapidly with increasing energy (σ_{el} ∝ 1/E^2) and the result may be small counting rates. At LEP energies with \mathcal{L} = 10^{30} \text{cm}^{-2} \text{s}^{-1} one can expect only about 25 Hz for the counting rate. Background from other processes can become problematic when the signal is small.

8.4 Absolute luminosity - hadron colliders
For hadron colliders two types of calibration have become part of regular operation, the measurement of the beam size by scanning the beam and the calibration with the cross section for small angle scattering. The determination of the bunch intensities is usually easier, although non-trivial in the case of a collider with several thousand bunches.

Fig. 8: Principle of luminosity measurement using Bhabha scattering for e⁺ e⁻ colliders.
8.4.1 Measurement by profile monitors and beam displacement

Typical profile measurement devices are wire scanners where a thin wire is moved through the beam and the interaction of the beam with the wire gives the signal. For high intensity hadron beams this has however limitations. Non-destructive devices such as synchrotron light monitors are available but the emitted light from hadrons is often not sufficient for a precise measurement.

An alternative is to measure the beam size by displacing the two beams against each other. The relative luminosity reduction due to this offset can be measured and is described by the formula (32) developed earlier:

$$ W = e^{-\frac{d^2}{4\sigma^2}} $$

where $d$ is the separation between the beams and the measurement of the luminosity ratio

$$ \frac{\mathcal{L}(d)}{\mathcal{L}_0} $$

is a direct measurement of $W$.

This method was already used in the CERN Intersection Storage Rings (ISR) and known as "van der Meer scan".

The expected counting rate of such a scan is shown in Fig.9. A fit to the above formula gives

![Fig. 9: Principle of luminosity measurement using transverse beam displacement.](image)

the beam size. A drawback of this method is the distortion of the beam optics in case of very strong beam-beam interactions [2]. This effect has to be evaluated carefully.

A further alternative is a so-called "beam-beam deflection scan" where instead of the change of counting rate the effect on the closed orbit is measured. This method was largely used at LEP and is explained in [2].
8.4.2 Absolute measurement with optical theorem

This method is similar to the measurement of Bhabha scattering for $e^+ e^-$ colliders but requires dedicated experiments and often special machine conditions.

The total elastic and inelastic counting rate is related to the luminosity and the total cross section (elastic and inelastic) by the expression:

$$\sigma_{\text{tot}} \cdot \mathcal{L} = N_{\text{inel}} + N_{\text{el}} \quad \text{(Total counting rate)} \ .$$  \hspace{1cm} (50)

The key to this method is that the total cross section is related to the elastic cross section for small values of the momentum transfer $t$ by the so-called optical theorem [7]:

$$\lim_{t \to 0} \frac{d\sigma_{\text{el}}}{dt} = (1 + \rho^2) \frac{\sigma^2_{\text{tot}}}{10\pi} = \frac{1}{\mathcal{L}} \frac{dN_{\text{el}}}{dt} \big|_{t=0} \ .$$  \hspace{1cm} (51)

Therefore the luminosity can in principle be calculated directly from experimental rates through:

$$\mathcal{L} = \frac{(1 + \rho^2) (N_{\text{inel}} + N_{\text{el}})^2}{16\pi} \left(\frac{dN_{\text{el}}}{dt}\right)_{t=0} \ .$$  \hspace{1cm} (52)

Both counting rates, the total number of events $N_{\text{inel}} + N_{\text{el}}$ and the differential elastic counting rate $dN_{\text{el}}/dt$ at small $t$ have to be measured with high precision. This requires a very good detector coverage of the whole space ($4\pi$) for the inelastic rate and the possibility to measure to very small values of $t$.

A slightly modified version of the above uses the Coulomb scattering amplitude which can be precisely calculated. The elastic scattering amplitude is a superposition of the strong ($f_s$) and Coulomb ($f_c$) amplitudes, the latter dominates at small $t$. We can re-write the differential elastic cross section $\frac{d\sigma_{\text{el}}}{dt}$:

$$\lim_{t \to 0} \frac{d\sigma_{\text{el}}}{dt} = \frac{1}{\mathcal{L}} \frac{dN_{\text{el}}}{dt} \big|_{t=0} = \frac{1}{\mathcal{L}} \frac{N_{\text{el}}}{t} \approx \pi \left| f_c + f_s \right|^2 \approx \pi \left| 2f_{\text{em}} + \frac{\sigma_{\text{tot}}}{4\pi} (\rho + i) e^{-B_{\text{em}}^2} \right|^2 \approx \frac{4\pi \sigma_{\text{em}}^2}{t^2} \big|_{t \to 0} \ .$$  \hspace{1cm} (53)

If the differential cross section is measured over a large enough range, the unknown parameters $\sigma_{\text{tot}}, \rho, B$ and $\mathcal{L}$ can be determined by a fit. A measurement [8–10] together with some crude fits is shown in Fig. 10 to demonstrate the principle. The advantage of this method is that it can be performed measuring only elastic scattering without the need of a full coverage to measure $N_{\text{inel}}$. It is therefore a good way to measure the luminosity (and total cross section $\sigma_{\text{tot}}$ and interference parameter $\rho$!) although the previous method is of more practical importance for regular use.

The measurement of the Coulomb amplitude usually requires dedicated experiments with detectors very close to the beam (e.g., with so-called Roman Pots) and therefore special parameters such as reduced intensity and zero crossing angle. Furthermore, in order to measure very small angle scattering, one has to reduce the divergence in the beam itself ($\sigma' = \sqrt{t/\beta}$). For that purpose special running conditions with a high $\beta^*$ at the collision point are often needed ($\beta^* > 1000 \text{ m}$) [9]. The precision of such a measurement is however as good as a few percent.

9 Not mentioned

At the end we would like to mention different types of colliders that were not treated explicitly, but the results can be extended easily. Examples of such machines are:

- Coasting beams (e.g., ISR).
- Asymmetric colliders (e.g., PEP, HERA).
- Linear colliders (SLC, TESLA etc.).

In this lecture we have tried to include some of the problems encountered in present and foreseen colliders which must be considered when a machine is built and the luminosity is optimized (not necessarily maximized!) To design a machine with the optimum luminosity, other limiting effects such as beam-beam effects [2] must be taken into account, but these are beyond the scope of this lecture.
Fig. 10: Principle of luminosity measurement using optical theorem in proton-proton (antiproton) collisions.

References
BEAM-BEAM INTERACTIONS

Werner Herr, AB Division
CERN, Geneva, Switzerland

Abstract
One of the most severe limitations in high intensity particle colliders is the beam-beam interaction, i.e. the perturbation of the beams as they cross the opposing beam. This introduction to beam-beam effects concentrates on a description of the phenomena that are present in modern colliding beams facilities.

1 INTRODUCTION
The problem of the beam-beam interaction is the subject of many studies since the introduction of the first particle colliders. It has been and will be one of the most important limits to the performance and therefore attracts the interest at the design stage of a new colliding beams facility. A particle beam is a collection of a large number of charges and represents an electromagnetic potential for other charges. It will therefore exert forces on itself and other beams. In the case of particle colliders these forces are experienced as localised periodic distortions when the two beams cross each other. The forces are most important for high density beams, i.e. high intensity and small beam sizes, which are the key to high luminosity. The beam-beam interaction is therefore often the limiting factor for the luminosity.

The electromagnetic forces from particle beams are very non-linear and result in a wide spectrum of consequences for the beam dynamics. Furthermore, as a result of the interaction, the charge distribution can change which has strong implications for the fields and forces. The subject of the beam-beam interaction is very rich of different effects and problems. Although we have a good qualitative understanding of the various phenomena, a complete theory does not exist and exact predictions are still difficult. Numerical techniques such as computer simulations have been used with great success to improve the picture on some aspects of the beam-beam interaction while for other problems the available models are not fully satisfactory in their predictive power. This lecture cannot be a complete review of the subject, but it is worthwhile to try to give an overview of the most important effects and to convey a physical picture or their nature. The more quantitative attempts and numerical methods can be found in the reference list. I shall concentrate on those problems that are important for present, recent and future machines. Other reviews and lectures can be found in [1] to [5].

2 Beam-beam force
The ensemble of charges in a beam produces an electromagnetic field and its knowledge is vital to understand the beam-beam dynamics. In the rest frame of a beam we have only electrostatic fields and to find the forces on other moving charges, we have to transform the fields into the moving frame and to calculate the Lorentz forces.

The fields are obtained by integrating over the charge distributions. The forces can be defocusing or focusing since the test particle can have the same or opposite charge with respect to the beam producing the forces. For the force on the bunch itself, i.e. the direct space charge, the kinematic factor is 

\[ \frac{1}{\gamma^2} \]

This results in the well known \(1/\gamma^2\) behaviour of direct space charge effects, which are always defocusing.

The distribution of particles producing the fields can follow various possible functions, leading to different fields (forces). It is not always possible to integrate the distribution to arrive at an analytical expression for the forces in which case either an approximation or numerical methods have to be used.
This is in particular true for hadron beams, which usually do not experience significant synchrotron radiation and damping. For e\(^-\) e\(^+\) colliders the distribution functions are most likely Gaussian with truncated tails.

In the two-dimensional case of a beam with bi-Gaussian beam density distributions, \(\rho(x,y) = \rho(x) \cdot \rho(y)\) in the transverse planes and with r.m.s. of \(\sigma_x\) and \(\sigma_y\):

\[
\rho_u(u) = \frac{1}{\sigma_u \sqrt{2\pi}} \exp\left(-\frac{u^2}{2\sigma_u^2}\right) \text{ where } u = x, y
\]

we can give the two-dimensional potential \(U(x, y, \sigma_x, \sigma_y)\) as a closed expression:

\[
U(x, y, \sigma_x, \sigma_y) = \frac{ne}{4\pi \epsilon_0} \int_0^{\infty} \frac{\exp\left(-\frac{x^2}{2\sigma_x^2 + q} - \frac{y^2}{2\sigma_y^2 + q}\right)}{\sqrt{(2\sigma_x^2 + q)(2\sigma_y^2 + q)}} dq
\]

where \(n\) is the line density of particles in the beam, \(e\) is the elementary charge and \(\epsilon_0\) the permittivity of free space. Its derivation is given in Appendix I. From the potential we can derive the transverse fields \(\vec{E}\) by taking the gradient.

To demonstrate the application of (2) and for practical purpose it is useful to evaluate (2) for two cases: for elliptical beams with \(\sigma_x \neq \sigma_y\) and for round beams with \(\sigma_x = \sigma_y\).

### 2.1 Elliptical beams

For the above case of bi-Gaussian distributions (i.e. elliptical beams with \(\sigma_x \neq \sigma_y\)) the fields can be derived and for the case of \(\sigma_x > \sigma_y\) we have [8]:

\[
E_x = \frac{ne}{2\epsilon_0 \sqrt{2\pi(\sigma_x^2 - \sigma_y^2)}} \text{Im} \left[ \exp\left(\frac{x + iy}{\sqrt{2(\sigma_x^2 - \sigma_y^2)}}\right) - e^{-\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2}} \exp\left(\frac{x \frac{\sigma_y}{\sigma_x} + iy \frac{\sigma_x}{\sigma_y}}{\sqrt{2(\sigma_x^2 - \sigma_y^2)}}\right) \right]
\]

\[
E_y = \frac{ne}{2\epsilon_0 \sqrt{2\pi(\sigma_x^2 - \sigma_y^2)}} \text{Re} \left[ \exp\left(\frac{x + iy}{\sqrt{2(\sigma_x^2 - \sigma_y^2)}}\right) - e^{-\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2}} \exp\left(\frac{x \frac{\sigma_y}{\sigma_x} + iy \frac{\sigma_x}{\sigma_y}}{\sqrt{2(\sigma_x^2 - \sigma_y^2)}}\right) \right]
\]

The function \(\text{erf}(t)\) is the complex error function

\[
\text{erf}(t) = e^{-t^2} \left[ 1 + \frac{2i}{\sqrt{\pi}} \int_0^t e^{z^2} dz \right]
\]

For the evaluation of the complex error function fast algorithms exist and the fields \(E_x\) and \(E_y\) can be computed more efficiently. It is therefore advantageous to express the fields in terms of the complex error function for computational reasons.

The magnetic field components follow from:

\[
B_y = -\beta_r E_x/c \quad \text{and} \quad B_x = \beta_r E_y/c
\]

The Lorentz force acting on a particle with charge \(q\) is finally:

\[
\vec{F} = q(\vec{E} + \vec{v} \times \vec{B})
\]
2.2 Round beams

We now make the simplifying assumption of round beams \((\sigma_x = \sigma_y = \sigma)\). We can re-write (7) in cylindrical coordinates:

\[
\vec{F} = q(E_r + \beta c B_\Phi) \times \vec{r} \tag{8}
\]

From (2) and with \(r^2 = x^2 + y^2\) we can immediately write the fields from (8) as

\[
E_r = -\frac{n e}{4\pi \epsilon_0} \cdot \frac{\delta}{\delta r} \int_0^\infty \exp\left(-\frac{r^2}{2(\sigma^2 + q)}\right) \frac{dq}{(2\sigma^2 + q)} \tag{9}
\]

and

\[
B_\Phi = -\frac{n e \beta c \mu_0}{4\pi} \cdot \frac{\delta}{\delta r} \int_0^\infty \exp\left(-\frac{r^2}{2(\sigma^2 + q)}\right) \frac{dq}{(2\sigma^2 + q)} \tag{10}
\]

We find that (9) and (10) and therefore the force (8) have only a radial component. The expressions (9) and (10) can easily be evaluated when the derivative is done first and \(\frac{1}{2\sigma^2 + q}\) is used as integration variable. We can now express the radial force in a closed form (using \(\epsilon_0\mu_0 = c^{-2}\)):

\[
F_r(r) = -\frac{n e^2 (1 + \beta^2)}{2\pi \epsilon_0} \cdot \frac{1}{r} \cdot \left[1 - \exp\left(-\frac{r^2}{2\sigma^2}\right)\right] \tag{11}
\]

and for the Cartesian components in the two transverse planes we get:

\[
F_x(r) = -\frac{n e^2 (1 + \beta^2)}{2\pi \epsilon_0} \cdot \frac{x}{r^2} \cdot \left[1 - \exp\left(-\frac{r^2}{2\sigma^2}\right)\right] \tag{12}
\]

and

\[
F_y(r) = -\frac{n e^2 (1 + \beta^2)}{2\pi \epsilon_0} \cdot \frac{y}{r^2} \cdot \left[1 - \exp\left(-\frac{r^2}{2\sigma^2}\right)\right] \tag{13}
\]

The shape of the force as a function of the amplitude is given in Fig.1. For small amplitudes the force is approximately linear and a particle crossing a beam at small amplitudes will experience a linear field. This results in a change of the tune like in a quadrupole. At larger amplitudes (i.e. above \(\approx 1 \sigma\)) the
force deviates strongly from this linear behaviour. Particles at larger amplitudes will also experience a
tune change, however this tune change will depend on the amplitude.
Already from the analytical form (13) we can see that the beam-beam force includes higher multipoles.
In the simplest case of a head-on collision all even multipoles are present in the force. One therefore has
to expect that a large number of resonances can be driven by the beam-beam force.

3 Incoherent effects - single particle effects
The force we have derived is the force of a beam on a single test particle. It can be used to study single
particle or incoherent effects. For that we treat a particle crossing a beam like it was moving through a
static electromagnetic lens. We have to expect all effects that are known from resonance and non-linear
theory such as:

- Unstable and/or irregular motion
- Beam blow up or bad lifetime

3.1 Beam-beam parameter
At first I shall derive the linear tune shift of a small amplitude particle crossing a round beam of a finite
length. We use the force to calculate the kick it receives from the opposing beam, i.e. the change of
the slope of the particle trajectory. Starting from the two-dimensional force and multiplying with the
longitudinal distribution which depends on both, position \( s \) and time \( t \) and I assume has a Gaussian
shape with a width of \( \sigma_s \):

\[
F_r(r, s, t) = N e^{2(1 + \beta^2)} \cdot \frac{1}{\sqrt{(2\pi)^3}} \cdot \frac{2}{r} \cdot \left[ 1 - \exp\left(-\frac{r^2}{2\sigma_s^2}\right) \right] \cdot \left[ \exp\left(-\frac{(s + vt)^2}{2\sigma_s^2}\right) \right]
\]

Now \( N \) is the total number of particles. We make use of Newton’s law and integrate over the collision to
get the radial deflection:

\[
\Delta r' = \frac{1}{mc\beta\gamma} \int_{-\infty}^{\infty} F_r(r, s, t) dt
\]

The radial kick \( \Delta r' \) a particle with a radial distance \( r \) from the opposing beam centre receives is then:

\[
\Delta r' = -\frac{2Nr_0}{\gamma} \cdot \frac{1}{r} \cdot \left[ 1 - \exp\left(-\frac{r^2}{2\sigma_s^2}\right) \right]
\]

(14)

where I have re-written the constants and use the classical particle radius:

\[
r_0 = \frac{e^2}{4\pi\varepsilon_0 mc^2}
\]

(15)

where \( m \) is the mass of the particle. After the integration along the bunch length, \( N \) is the total number
of particles. In the two transverse planes we get the expressions for the kick:

\[
\Delta x' = -\frac{2Nr_0}{\gamma} \cdot \frac{x}{r^2} \cdot \left[ 1 - \exp\left(-\frac{r^2}{2\sigma_s^2}\right) \right]
\]

\[
\Delta y' = -\frac{2Nr_0}{\gamma} \cdot \frac{y}{r^2} \cdot \left[ 1 - \exp\left(-\frac{r^2}{2\sigma_s^2}\right) \right]
\]

For small amplitudes \( r \) we can derive the asymptotic limit:

\[
\Delta r' \bigg|_{r \to 0} = -\frac{Nr_0 r}{\gamma\sigma_s^2} = -r \cdot f
\]

(16)
This limit is the slope of the force at $r = 0$ and the force becomes linear with a focal length as the proportionality factor.

It is well known how the focal length relates to a tune change and we can derive a quantity $\xi$ which is known as the **linear beam-beam parameter**:

$$\xi = \frac{Nr_0\beta^*}{4\pi\gamma\sigma^2}$$

(17)

$r_0$ is the classical particle radius, (e.g.: $r_e, r_p$) and $\beta^*$ is the optical amplitude function ($\beta$-function) at the interaction point.

For small values of $\xi$ and a tune far enough away from linear resonances this parameter is equal to the linear tune shift $\Delta Q$.

The beam-beam parameter can be generalized for the case of non-round beams and becomes:

$$\xi_{x,y} = \frac{Nr_0\beta^*_{x,y}}{2\pi\gamma\sigma_{x,y}(\sigma_x + \sigma_y)}$$

(18)

The beam-beam parameter is often used to quantify the strength of the beam-beam interaction, however it does not reflect the non-linear nature. Nevertheless, it can be used for comparison and as a scaling parameter. In Tab.1 I have summarized beam parameters for LEP and the LHC. Some of the differences are striking: while the beams in the LHC are round at the interaction point, they are very flat in LEP. This is due to the excitation of the beam in the horizontal plane by the strong synchrotron radiation. Another observation is the much larger beam-beam parameter in LEP. The reason for this enormous difference will be discussed in a later section.

**Table 1:** Comparison of parameters for a lepton and hadron collider.

<table>
<thead>
<tr>
<th></th>
<th>LEP</th>
<th>LHC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beam sizes</td>
<td>$160\mu m \cdot 4\mu m$</td>
<td>$16.6\mu m \cdot 16.6\mu m$</td>
</tr>
<tr>
<td>Intensity N</td>
<td>$4.0 \cdot 10^{11}/\text{bunch}$</td>
<td>$1.15 \cdot 10^{11}/\text{bunch}$</td>
</tr>
<tr>
<td>Energy</td>
<td>100 GeV</td>
<td>7000 GeV</td>
</tr>
<tr>
<td>$\beta^<em>_x \cdot \beta^</em>_y$</td>
<td>$1.25 \text{ m} \cdot 0.05 \text{ m}$</td>
<td>$0.55 \text{ m} \cdot 0.55 \text{ m}$</td>
</tr>
<tr>
<td>Crossing angle</td>
<td>0.0</td>
<td>285 $\mu$rad</td>
</tr>
<tr>
<td>Beam-beam parameter($\xi$)</td>
<td>0.0700</td>
<td>0.0034</td>
</tr>
</tbody>
</table>
3.2 Non-linear effects

Before we enter the observations and more subtle beam-beam effects, we should briefly discuss the immediate effect of the non-linearity of the beam-beam force. It manifests as a amplitude dependent tune shift and for a beam with many particles as a tune spread. I shall use an intuitive approach for the calculation of the tune shifts.

The instantaneous tune shift of a particle when it crosses the other beam is related to the derivative of the force with respect to the amplitude (see above for the derivation of the linear tune shift) $\delta F/\delta x$. For a particle performing an oscillation with a given amplitude the tune shift is calculated by averaging the slopes of the force over the range (i.e. the phases) of the particle’s oscillation amplitudes. For the one-dimensional case this is illustrated schematically in Fig.2. The derivative of the beam-beam force from Fig.1 is plotted together with the oscillation range of two particles, one with small ($< 0.5\sigma$) and another with large amplitudes ($> 3.0\sigma$). One could now try to do the averaging by eye and comes to the following observations:

- The particles with small amplitudes have the largest tune shift.
- Particles with larger amplitudes have a smaller tune shift.
- At very large amplitudes the tune shift eventually becomes zero.

An elegant calculation can be done using the Hamiltonian formalism [2,3,9,53] developed for non-linear dynamics. A derivation using phase averaging according to the above recipe is given in the Appendix II. The result of the calculation is shown in Fig.3 (left) where the tune change is plotted as a function of the particle amplitude.

In the 2-dimensional case, the tune shifts ($\Delta Q_x$, $\Delta Q_y$) of a particle with amplitudes $x$ and $y$ depend on both, horizontal and vertical amplitudes. The detuning must be computed and presented in a 2-dimensional form, i.e. the amplitude $(x, y)$ is mapped into the tune space $(Q_x, Q_y)$ or alternatively to the 2-dimensional tune change $(\Delta Q_x, \Delta Q_y)$. Such a presentation is usually called a "tune footprint" and an example is shown in Fig.3 (right). It maps the amplitudes into the tune space and each "knot" of the mesh corresponds to a pair of amplitudes. Some of these pairs are indicated in the figure and correspond to the amplitudes in units of $\sigma$. Amplitudes between 0 and 6 $\sigma$ in both planes are used. The cross indicates the original, unperturbed tunes without the beam-beam interaction. An entire beam with many particles at
different amplitudes occupies this tune spread.
Inspecting the figure, we find an important result: The maximum tune spread for a single head-on collision is equal to the tune shift of a particle with small amplitudes and for small tune shifts equal to the beam-beam parameter $\xi$. In the simple case of a single head-on collision the parameter $\xi$ is therefore a measure for the tune spread in the beam.
When the tune is spread out it requires a finite space in the tune diagram and if it is too large it can cross resonances. In principle, the tune spread must be small enough that the required area can be fitted into a resonance free region of the working diagram. However, the beam-beam force can excite resonances of high order and a priori it is not clear up to which order resonances have to be avoided. For the LHC a total tune spread of about 0.010 to 0.015 is assumed to be the acceptable limit.

3.3 Beam stability
When the beam-beam interaction becomes too strong, the beam can become unstable or the beam dynamics is strongly distorted. One can distinguish different types of distortions and a few examples are:

- Non-linear motion can become stochastic
  - reduction of "dynamic aperture"
  - particle loss and bad lifetime

- Distortion of beam optics: dynamic beta (LEP)
- Vertical blow-up above the so-called beam-beam limit

Since the beam-beam force is very non-linear, the motion can become "chaotic". This often leads to a reduction of the available dynamic aperture. The dynamic aperture is the phase space region in which the beam remains stable. Particles outside the dynamic aperture can eventually get lost. The dynamic aperture is usually evaluated by tracking particles with a computer program through the machine where they experience the fields from the machine elements and other effects such as wake fields or the beam-beam interaction. Particles are usually lost when they have reached large amplitudes. In the real machine this is noticed by the measurement of particle losses and a reduced beam life time.
Since the beam-beam interaction is basically a very non-linear lens in the machine, it distorts the optical properties. In particular it may create a noticeable beating of the $\beta$-function around the whole machine and at the location of the beam-beam interaction itself. This can be approximated by inserting a quadrupole which produces the same tuneshift at the position of the beam-beam interaction. The r.m.s. beam size at the collision point is now proportional to $\sqrt{\beta'_p}$, where $\beta'_p$ is the perturbed $\beta$-function which can be significantly different from the unperturbed $\beta$-function $\beta^*$. This in turn changes the strength of the beam-beam interaction and the parameters have to be found in a self-consistent form. This is called the dynamic beta effect. This is a first deviation from our assumption that the beams are static non-linear lenses. A strong dynamic beta effect was found in LEP [14] due to its very large tune shift parameters. Another effect that can be observed in particular in $e^+e^-$ colliders is the blow up of the emittance which naturally limits the reachable beam-beam tune shifts.

### 3.4 Beam-beam limit

In $e^+e^-$ colliders the beam sizes are usually an equilibrium between the damping due to the synchrotron radiation and heating mechanisms such as quantum excitation, intra-beam scattering and very importantly, the beam-beam effect. This leads to a behaviour that is not observed in a hadron collider. When the luminosity is plotted as a function of the beam intensity, it should increase approximately as the current squared [18], in agreement with:

$$\mathcal{L} = \frac{N^2 \cdot k f}{4\pi\sigma_x\sigma_y}$$

(19)

Here $k$ is the number of bunches per beam and $f$ the revolution frequency [18]. At the same time the beam-beam parameter $\xi$ should increase linearly with the beam intensity according to (18):

$$\xi_y = \frac{N \cdot r_e \beta_y}{2\pi\gamma\sigma_y(\sigma_x + \sigma_y)}$$

(20)

In all $e^+e^-$ colliders the observation can be made that above a certain current, the luminosity increases approximately proportional to the current, or at least much less than with the second power [11]. Another observation is that at the same value of the intensity the beam-beam parameter $\xi$ saturates. This is shown for three $e^+e^-$ colliders in Fig.4 and schematically illustrated in Fig.5. This limiting value of $\xi$ is

![Fig. 4: Measurements of luminosity and beam-beam limit in $e^+e^-$ colliders. Logarithmic scale of the axes to demonstrate change of exponent.](image-url)
commonly known as the **beam-beam limit**.

When we re-write the luminosity as:

\[
\mathcal{L} = \frac{N^2 \cdot kf}{4\pi\sigma_x\sigma_y} = \frac{N\cdot kf}{4\pi\sigma_x} \cdot \frac{N}{\sigma_y}
\]

we get an idea of what is happening. In \(e^+e^-\) colliders the horizontal beam size \(\sigma_x\) is usually much larger than the vertical beam size \(\sigma_y\) and changes very little. In order for the luminosity to increase proportionally to the intensity \(N\), the factor \(N/\sigma_y\) must be constant. This implies that with increasing current the vertical beam size increases in proportion above the beam-beam limit. This has been observed in all \(e^+e^-\) colliders and since the vertical beam size is usually small, this emittance growth can be very substantial before the life time of the beam is affected or beam losses are observed.

The dynamics of machines with high synchrotron radiation is dominated by the damping properties and the beam-beam limit is not a universal constant nor can it be predicted. Simulation of beams with many particles can provide an idea of the order of magnitude [12, 13].

### 4 Crossing angle

To reach the highest luminosity, it is desirable to operate a collider with as many bunches as possible since the luminosity is proportional to their number [18].

In a single ring collider such as the SPS, Tevatron or LEP, the operation with \(k\) bunches leads to \(2k\) collision points. When \(k\) is a large number, most of them are unwanted and must be avoided to reduce the perturbation due to the beam-beam effects. Various schemes have been used to avoid these unwanted "parasitic" interactions. Two prominent examples are shown in Fig.6. In the SPS, Tevatron and in LEP so-called Pretzel schemes were used. When the bunches are equidistant, this is the most promising method. When two beams of opposite charge travel in the same beam pipe, they can be moved onto separate orbits using electrostatic separators. In a well-defined configuration the two beams cross when the beams are separated (Fig.6, left). To avoid a separation around the whole machine, the bunches can be arranged in so-called trains of bunches following each other closely. In that case a separation with electrostatic separators is only needed around the interaction regions. Such a scheme was used in LEP in the second phase [10] and it is schematically illustrated in Fig.6 (right).

Contrary to the majority of the colliders, the LHC collides particles of the same type which therefore must travel in separate beam pipes. At the collision points of the LHC the two beams are brought together and
Fig. 6: Beam separation with a Pretzel scheme (SPS, Tevatron, LEP, left) and with short bunch trains (LEP, right).

Fig. 7: Schematic layout of the LHC collision points and beams.

into collision (Fig.7). An arrangement of separation and recombination magnets is used for the purpose to make the beams cross (Fig.8). During that process it is unavoidable that the beams travel in a common vacuum chamber for more than 120 m. In the LHC the distance between the bunches is only 25 ns and therefore the bunches will meet in this region. In order to avoid the collisions, the bunches collide at a small crossing angle of 285 $\mu$rad. The basic principle is shown in Fig.9: while two bunches collide at a small angle (quasi head-on) at the centre, the other bunches are kept separated by the crossing angle. However, since they travel in a common beam pipe, the bunches still feel the electromagnetic forces...
from the bunches of the opposite beam. When the separation is large enough, these so-called long-range interactions should be weak. From the bunch spacing and the length of the interaction region one can easily calculate that at each of the four LHC interaction points we must expect 30 of these long-range encounters, i.e. in total 120 interactions. The typical separation between the two beams is between 7 and 10 in units of the beam size of the opposing beam.

4.1 Long-range beam-beam effects

Although the long-range interactions distort the beams much less than a head-on interaction, their large number and some particular properties require careful studies:

- They break the symmetry between planes, i.e. also odd resonances are excited.
- While the effect of head-on collisions is strongest for small amplitude particles, they mostly affect particles at large amplitudes.
- The tune shift caused by long-range interactions has opposite sign in the plane of separation compared to the head-on tune shift.
- They cause changes of the closed orbit.
- They largely enhance the so-called PACMAN effects.

4.1.1 Opposite sign tune shift

The opposite sign of the tune shift can easily be understood when we come back to the method for calculating the tune spread, explained with the help of Fig.2. We average again the oscillation of a small amplitude particle as it samples the focusing force of the beam-beam interaction. In Fig.10 I show the range of oscillation for central collisions and for the interaction of separated beams. In both cases for particles with small oscillation amplitudes. When the separation is larger than \( \approx 1.5 \sigma \), the focusing (slope of the force as a function of the amplitude) changes the sign and the resulting tune shift assumes the opposite sign.

To some extend this property could be used to partially compensate long-range interactions when a configuration is used where the beams are separated in the horizontal plane in one interaction region and in the vertical plane in another one.

4.1.2 Strength of long-range interactions

The geometry of a single encounter is shown in Fig.11. The particles in the test bunch receive a kick (change of slope) \( \Delta x' \). Assuming a separation \( d \) in the horizontal plane, the kicks in the two planes can be written as:

\[
\Delta x' = -\frac{2N r_0}{\gamma} \cdot \frac{(x + d)}{r^2} \cdot \left[ 1 - \exp\left(-\frac{r^2}{2\sigma^2}\right) \right]
\]  

(22)
Fig. 10: Derivative of beam-beam force for round beams. Oscillation range of particles of centred and separated beams.

Fig. 11: Long-range interaction, schematic.

with \( r^2 = (x + d)^2 + y^2 \). The equivalent formula for the plane orthogonal to the separation is:

\[
\Delta y' = -\frac{2N \gamma_0}{\gamma} \cdot \frac{y}{y^2} \left[ 1 - \exp\left( -\frac{r^2}{2\sigma^2} \right) \right]
\]  \hspace{1cm} (23)

It is fairly obvious that the effect of long-range interactions must strongly depend on the separation. The calculation shows that the tune spread \( \Delta Q_{lr} \) from long-range interactions alone follows an approximate scaling (for large enough separation, i.e. above \( \approx 6\sigma \)):

\[
\Delta Q_{lr} \propto \frac{N}{d^2}
\]  \hspace{1cm} (24)

where \( N \) is the bunch intensity and \( d \) the separation. Small changes in the separation can therefore result in significant differences.
4.1.3 Footprint for long-range interactions

Contrary to the head-on interaction where the small amplitude particles are mostly affected, now the large amplitude particles experience the strongest long-range beam-beam perturbations. This is rather intuitive since the large amplitude particles are the ones which can come closest to the opposing beam as they perform their oscillations. We must therefore expect a totally different tune footprint. Such a footprint for only long-range interactions is shown in Fig.12. Since the symmetry between the two planes is broken, the resulting footprint shows no symmetry. In fact, the tune shifts have different signs for \( x \) and \( y \), as expected. For large amplitude one observes a "folding" of the footprint. This occurs when large amplitude particles are considered for which the oscillation amplitudes extend across the central maximum in Fig.10, i.e. when the oscillation amplitude is larger than the separation between the beams.

Such large amplitudes are treated in Fig.10 to demonstrate this feature. In practice, these amplitudes are usually not important since in real machines no particles reach these amplitudes. In Fig.13 (left) I show separately the footprints for two head-on proton-proton collisions, and long-range interactions with horizontal separation and vertical separation, respectively. The differences and in particular the different sign of the long-range tune shift are clearly visible. The long-range footprints are shifted away from the original tune (0.28,0.31) in opposite directions. Amplitudes between 0 and 6 \( \sigma \) are shown in the figure. In the same figure on the right I show the combined footprint, i.e. for particles which experience two head-on collisions, long-range interactions in one interaction point with horizontal separation and a second with vertical separation. A partial compensation can be seen and the footprint is again symmetric in \( x \) and \( y \). In particular the linear tune shifts of the central parts are very well compensated.

4.1.4 Dynamic aperture reduction due to long-range interactions

For too small separation, the tune spread induced by long-range interactions can become very large and resonances cannot be avoided any more. The motion can become irregular and as a result particles at large amplitudes can get lost. This is demonstrated in Fig.14. The emittance increase of a large amplitude particle (5 \( \sigma \)) is computed for different values of the crossing angle and hence of the beam separation. For large enough angles the emittance increase is very small, but it increases over several orders of magnitude when the crossing angle drops significantly below 300 \( \mu \)rad. These results are obtained by particle...
tracking [15, 16].

To evaluate the dynamic aperture in the presence of beam-beam interactions, a simulation of the complete machine is necessary and the interplay between the beam-beam perturbation and possible machine imperfections is important [17].

For the present LHC parameters we consider the minimum crossing angle to be 285 µrad.

4.2 Beam-beam induced orbit effects

When two beams do not collide exactly head-on, the force has a constant contribution which can easily be seen when the kick $\Delta x'$ (from (22), for sufficiently large separation) is developed in a series:

$$\Delta x' = \frac{\text{const.}}{d} \cdot \left[ 1 - \frac{x}{d} + O\left(\frac{x^2}{d^2}\right) + \ldots \right]$$  (25)

A constant contribution, i.e. more precisely an amplitude independent contribution, changes the orbit of the bunch as a whole (Fig.15). When the beam-beam effect is strong enough, i.e. for high intensity
and/or small separation, the orbit effects are large enough to be observed. When the orbit of a beam changes, the separation between the beams will change as well, which in turn will lead to a slightly different beam-beam effect and so on. The orbit effects must therefore be computed in a self-consistent way [19], in particular when the effects are sizeable. The closed orbit of an accelerator can usually be corrected, however an additional effect which is present in some form in many colliders, sets a limit to the correction possibilities. A particularly important example is the LHC and I shall therefore use it to illustrate this feature.

4.3 PACMAN bunches

The bunches in the LHC do not form a continuous train of equidistant bunches spaced by 25 ns, but some empty space must be provided to allow for the rise time of kickers. These gaps and the number of bunches per batch are determined by requirements from the LHC injectors (PS, SPS etc.) and the preparation of the LHC beam (bunch splitting). The whole LHC bunch pattern is composed of 39 smaller batches (trains of 72 bunches) separated by gaps of various length followed by a large abort gap for the dump kicker at the end. Fig. 16 shows the actual LHC filling scheme with the various gaps in the train. In the LHC, only 2808 out of 3564 possible bunches are present with the above filling scheme. Due to

72 bunches

\[ \Delta t_1 \Delta t_2 \Delta t_3 \Delta t_4 \]

\[ \Delta t_1 \text{ 8 bunches missing} \]
\[ \Delta t_2 \text{ 38 bunches missing} \]
\[ \Delta t_3 \text{ 39 bunches missing} \]
\[ \Delta t_4 \text{ 119 bunches missing} \]

**Fig. 16:** Bunch filling scheme in the LHC.

the symmetry, bunches normally meet other bunches at the head-on collision point. For the long-range interactions this is no longer the case. This is illustrated in Fig. 17. Bunches at the beginning and at
the end of a small batch will encounter a hole and as a result experience fewer long-range interactions than bunches from the middle of a batch [20]. In the limit, the first bunch of a batch near a large gap encounters no opposing bunch before the central collision and the full number of bunches after. Bunches with fewer long-range interactions have a very different integrated beam-beam effect and a different dynamics must be expected. In particular they will have a different tune and occupy a different area in the working diagram, therefore may be susceptible to resonances which can be avoided for nominal bunches. The overall space needed in the working diagram is therefore largely increased [20, 21]. Another consequence of reduced long-range interactions is the different effect on the closed orbit of the bunches. We have to expect a slightly different orbit from bunch to bunch. This effect is demonstrated in Fig. 18 where I show the horizontal position at one head-on collision point for 432 bunches (out of 2808). The bunches in the middle of a batch have all interactions and therefore the same orbit while the bunches at the beginning and end of a batch show a structure which exhibits the decreasing number of long-range interactions. The orbit spread is approximately 10 - 15% of the beam size. Since the orbits of the two beams are not the same, it is impossible to make all bunches collide exactly head-on. A significant fraction will collide with an offset. Although the immediate effect on the luminosity is small [18], collisions at an offset can potentially affect the dynamics and are undesirable. The LHC design should try to minimize these offsets [20, 21]. A second effect, the different tunes of the bunches, is shown in Fig. 19. For three batches it shows a sizeable spread from bunch to bunch and without compensation effects [21] it may be too large for a safe operation.
4.4 Beam-beam deflection scan

The orbit effect can be useful when one has only a few bunches, i.e. no PACMAN effects. It can be used to optimize luminosity by a so-called beam-beam deflection scan. During this procedure the two beams are displaced in steps against each other and the two beams get an orbit kick, depending on their distance. The resulting orbits are measured around the whole circumference after each step and from this measurement the beam-beam kick can be computed easily. The result of such a measurement is shown in Fig.20 [22]. It shows the computed kick as a function of the beam separation. The solid line is a fit through the measurements and is used to derive beam parameters such as beam sizes and the position when the beams collide exactly head-on. It reflects the force shown in Fig.1, however for very flat beams. This method was used extensively in LEP to centre the two beams [22].
5 Coherent beam-beam effects

So far we have mainly studied how the beam-beam interaction affects the single particle behaviour and treated the beam-beam interaction as a static lens. In the literature this is often called a “weak-strong” model: a “weak” beam (a single particle) is perturbed by a “strong” beam (not affected by the weak beam). When the beam-beam perturbation is important, the model of an unperturbed, strong beam is not valid anymore since its parameters change under the influence of the other beam and vice versa. When this is the case, we talk about so-called “strong-strong” conditions. The first example of such a “strong-strong” situation was the orbit effect where the beams mutually changed their closed orbits. These closed orbits had to be found in a self-consistent way. This represents a static strong-strong effect.

In the next step we investigate dynamic effects under the strong-strong condition.

When we consider the coherent motion of bunches, the collective behaviour of all particles in a bunch is studied. A coherent motion requires an organized behaviour of all particles in a bunch. A typical example are oscillations of the centre of mass of the bunches, so-called dipole oscillations. Such oscillations can be driven by external forces such as impedances and may be unstable. At the collision of two counter-

![Fig. 21: Coherent kick leading to oscillations of the bunch.](image)

rotating bunches not only the individual particles receive a kick from the opposing beam, but the bunch as an entity gets a coherent kick (Fig.21). This coherent kick of separated beams can excite coherent dipole oscillations. Its strength depends on the distance between the bunch centres at the collision point. It can be computed by adding the individual contributions of all particles. For small distances it can be shown [27, 28] that it is just one half of the incoherent kick a single particle would receive at the same distance. For distances large enough the incoherent and coherent kicks become the same.

5.1 Coherent beam-beam modes

To understand the dynamics of dipole oscillations we first study the simplest case with one bunch in each beam. When the bunches meet turn after turn at the collision point, their oscillation can either be exactly in phase (0 degree phase difference) or out of phase (180 degrees or $\pi$ phase difference). Any other oscillation can be constructed from these basic modes. The modes are sketched very schematically in Fig.22. The relative positions of the bunches as observed at the interaction point are shown for two consecutive turns $n$ and $n + 1$. The first mode is called the 0-mode (or sometimes called $\sigma$-mode) and the second the $\pi$-mode. In the first mode the distance between the bunches does not change turn by turn and therefore there is no net force driving an oscillation. This mode must oscillate with the unperturbed frequency (tune) $Q_0$. For the second mode the net force difference between two turns is a maximum and the tune becomes $Q_0 + \Delta Q_{coh}$. The sign of $\Delta Q_{coh}$ depends whether the two beams have equal charge (defocusing case) or opposite charge (focusing case). The calculation of $\Delta Q_{coh}$ is non-trivial: when the bunches are considered as rigid objects, the tune shift can be computed easily using the coherent kick but is underestimated [23]. The correct calculation must allow for changes of the density distribution during the collision and moreover, must allow a deviation from a Gaussian function. The computation requires
to solve the Vlasov-equation of two coupled beams [29–33].
The results are summarized in a schematic picture in Fig.23. The 0-mode is found at the unperturbed tune as expected. The $\pi$-mode is shifted by 1.2-1.3 $\xi$. The precise value depends on the ratio of the horizontal and vertical beam sizes [30]. We have seen before the incoherent tune spread (footprint) the individual particles occupy and we know that it spans the interval $[0.0,1.0] \xi$, starting at the 0-mode (Fig.23).

Here we can make an important observation: under the strong-strong condition the $\pi$-mode is a discrete mode outside the incoherent spectrum [31, 32]. This has dramatic consequences for the stability of the beams. A coherent mode that is outside an incoherent frequency spectrum cannot be stabilized by Landau damping [34]. Under these conditions the coherent beam-beam effect could drive the dipole oscillation to large amplitudes and may result in the loss of the beam. Observations of the coherent beam-beam effects have been made at PETRA [23]. The coherent beam-beam modes have also been observed in LEP during a special experiment with only one bunch in each beam. The frequency spectrum of two bunches in LEP is shown in Fig.24 and clearly demonstrated the two modes [24].

The observation in a lepton collider is simplified by the presence of strong damping which reduces the danger of a beam loss. Nevertheless experimental evidence exist for the observation of coherent beam-beam modes in hadron colliders. Beam-beam modes have been observed with high intensity coasting beams in the ISR [25], and recently one has succeeded to measure the beam-beam modes in a bunched hadron collider at RHIC [26]. In Fig.25 the spectra with and without the beam-beam interaction are compared and a significant signal can be observed. Without the beam-beam interaction only the 0-mode ($\sigma$-mode) is observed and in the presence of the other beam, the $\pi$-mode is visible.

Coherent beam-beam modes can be driven by head-on collisions with a small offset or by long-range interactions. In the first case and for small oscillations, the problem can be linearized and the theoretical

---

**Fig. 22:** Basic dipole modes of two bunches. Relative position of the bunches at the interaction point at two consecutive turns.

**Fig. 23:** Frequency of coherent beam-beam modes and incoherent spectra.
treatment is simplified. The forces from long-range interactions are very non-linear but the numerical evaluation is feasible. Since the coherent shift must have the opposite sign for long-range interactions, the situation is very different. In particular the $\pi$-mode from long-range interactions alone would appear on the opposite side of the 0-mode in the frequency spectrum [32, 43]. Both, the incoherent and the coherent spectra include both types of interactions. The interference between the two types is however complicated and beyond the scope of such a school [32, 35].

5.2 Measures against coherent beam-beam modes

The loss of Landau damping for coherent beam-beam modes can be a problem for hadron colliders and possible measures against coherent beam-beam modes may have to be considered. We have argued that the coherent motion requires an ‘organized’ motion of many particles and this will help us to suppress the coherent modes. In order to maintain this organized motion, a high degree of symmetry between the two beams is required and breaking this symmetry can be used to destroy the collective behaviour. Several possible methods have been proposed [32, 36, 38, 44]:

- Different bunch intensities
Different tunes in the two beams
Unequal distribution of interaction points
Phase differences between interaction points
Synchrotron motion

This list is not complete but sufficient to demonstrate the principle.

5.2.1 Unequal bunch intensities

A very quantitative prediction was made in [31, 32] when the strong-strong condition is fulfilled. To lock the two beams into a common motion, the mutual forces on each other (and therefore the tune shift) should be as equal as possible. Different bunch intensities break this condition and it was found in [31, 32] that the intensity ratio between the weaker and the stronger beam must not be less than 0.60. Fig. 26 shows the result of a simulation [43].

Fig. 26: Coherent spectra for different bunch intensities. Left figure for an intensity ratio of 0.65, right figure for a ratio of 0.55.

very well reproduced. Two beams with $10^4$ macro-particles each were tracked through the machine and subjected to the beam-beam interaction in one interaction point. To get a quantitatively correct result the fields must be computed from the actual particle distributions which change from turn to turn and usually deviate from a Gaussian shape. Simulation programs with advanced algorithms have been developed for such strong-strong simulations [33, 43].

In Fig.26 we show the frequency spectra from two simulations: the figure on the left with an intensity ratio of 0.65 and the right figure for a ratio of 0.55. For a ratio of 0.65 the $\pi$-mode is visible as a peak, however the shift away from the 0-mode is smaller than for equal intensities. As predicted by [31, 32], with decreasing ratio the frequency of the $\pi$-mode moves closer and closer towards the incoherent spectrum until it merges exactly at the predicted ratio. In this case Landau damping is restored and the $\pi$-mode has disappeared and only the 0-mode and the incoherent spectrum are visible.

5.2.2 Unequal tunes of the two beams

Another way to break the symmetry between the two beams is to operate them with different betatron frequencies, i.e. different tunes [36]. When the tune difference is large enough, the two beams are decoupled and cannot perform a coherent oscillation [36, 37]. This is demonstrated in Fig.27. The fractional tune of one beam is kept constant at 0.310 while the fractional tune of the other beam is slowly changed. The spectrum of the first beam is then determined from its motion and shown in Fig.27.
Fig. 27: Coherent spectra for different tunes of the two beams. Left figure for a tune difference of 0.002, right figure for a difference of 0.004.

Fig. 27 (left) the $\pi$-mode is still visible when $\Delta Q$ between the two beams was 0.002. After the tune difference $\Delta Q$ between the two beams is increased to 0.004 (Fig. 27, right), the coherent modes are suppressed and only the incoherent spectra of the two beams are left [36, 43].

5.2.3 **Synchrotron motion**

A particular feature of hadron colliders can be very efficient to damp coherent beam-beam modes. In hadron colliders it is a coincidence, that the absolute value of the beam-beam parameter and the magnitude of the synchrotron tune are of the same order of magnitude (LHC: $\xi = 0.0033$, $Q_s = 0.0021$). When the transverse and longitudinal motions are coupled, this leads to so-called synchro-betatron coupling, and the appearance of incoherent synchrotron sidebands, separated by $m \cdot Q_s$ from the tune $Q_0$.

This can lead to a situation where an incoherent synchrotron sideband overlaps with the position of the coherent $\pi$-mode. The result of this overlap is Landau damping! This effect was suggested in [32] and demonstrated by multi-particle simulation in [44].

5.2.4 **Multiple interaction points**

So far we have treated beam-beam interactions at a single interaction point and the consequences of symmetry breaking between the two beams. In case of two or more interaction points, additional symmetry breaking effects exist. Like in the case with one interaction point but unequal tunes, the symmetry can be broken when the phase advance between multiple interaction points is not the same [38]. It can be demonstrated [38, 39] that the symmetry and periodicity of the head-on and long-range interaction points play an important role in the excitation of beam-beam resonances.

With many bunches colliding, the frequency spectrum will change. When a bunch collides with more than one other bunch, the bunches will couple and more modes than only the 0-mode and $\pi$-mode will be excited [40, 41, 45]. This coupling can be either through head-on or long-range interactions. In the case of the LHC, every nominal bunch couples with 90 other bunches through long-range interactions and eventually all bunches are coupled together. This would result in a large number of possible modes. Their frequencies must all lie between the 0-mode and the $\pi$-mode which is the mode with the highest frequency shift. This is demonstrated in Fig. 28 (left) where the spectrum of one bunch out of 36 bunches is shown which undergo two head-on interactions. These results were obtained from a rigid bunch model to show all possible modes [46, 47]. For such a model no incoherent spectrum is visible and Landau damping cannot occur. The equivalent simulation with a truly self-consistent multi-particle simulation
Fig. 28: Coherent beam-beam modes with many bunches coupled together by head-on and long range beam-beam interactions. Left figure with rigid bunch model, right figure with self-consistent multi-particle model.

In Fig. 29 a similar result is shown, this time for a collision scheme which exhibits more coherent modes (unsymmetric collision scheme). The comparison between the left (rigid bunches) and the right picture (multi particle simulation) shows again that all modes whose frequencies are inside the incoherent spectrum are damped, while the two discrete modes just outside this continuum are still visible.

In general, symmetry breaking effects and additional degrees of freedom make it more difficult for coherent effects to sustain. For future colliders with many interaction points and head-on as well as multiple long-range collisions, it is very likely that coherent beam-beam oscillations can easily be damped by the lack of symmetry and the help of the synchrotron motion.
6 Compensation of beam-beam effects

For the case the beam-beam effects limit the performance of a collider, several schemes have been proposed to compensate all or part of the detrimental effects. The basic principle is to design correction devices which act as non-linear 'lenses' to counteract the distortions from the non-linear beam-beam 'lens'. For both, head-on and long-range effects schemes have been proposed:

- **Head on effects:**
  - Electron lenses
  - Linear lens to shift tunes
  - Non-linear lens to decrease tune spread

- **Long-range effects:**
  - At large distance: beam-beam force changes like $1/r$
  - Same force as a wire!

6.1 Electron lenses

The basic principle of a compensation of proton-proton (or antiproton) collisions with an "electron lens” implies that the proton (antiproton) beam travels through a counter-rotating high current electron beam ("electron lens”) [48, 49]. The negative electron space charge can reduce the effect from the collision with the other proton beam.

An electron beam with a size much larger than the proton beam can be used to shift the tune of the proton beam ("linear lens”). When the current in the electron bunches can be varied fast enough, the tune shift can be different for the different proton bunches, thus correcting PACMAN tune shifts.

When the electron charge distribution is chosen to be the same as the counter-rotating proton beam, the non-linear focusing of this proton beam can be compensated ("non-linear lens”). When it is correctly applied, the tune spread in the beam can be strongly reduced.

Such lenses have been constructed at the Tevatron at Fermilab [49] and experiments are in progress.

6.2 Electrostatic wire

To compensate the tune spread from long-range interactions, one needs a non-linear lens that resembles a separated beam. At large enough separation, the long-range force changes approximately with $1/r$ and this can be simulated by a wire parallel to the beam [50]. In order to compensate PACMAN effects, the wires have to be pulsed according to the bunch filling scheme. Tests are in progress at the SPS to study the feasibility of such a compensation for the LHC.

6.3 Möbius scheme

The beam profiles of $e^+e^-$ colliders are usually flat, i.e. the vertical beam size is much smaller than the horizontal beam size. Some studies indicate that the collision of round beams, even for $e^+e^-$ colliders, show more promise for higher luminosity since larger beam-beam parameters can be achieved. Round beams can always be produced by strong coupling between horizontal and vertical planes. A more elegant way is the so-called Möbius lattice [51, 52]. In this lattice, the horizontal and vertical betatron oscillations are exchanged by an insertion. A horizontal oscillation in one turn becomes a vertical oscillation in the next turn and vice versa.

Tests with such a scheme have been done at CESR at Cornell [52].
7 Summary

A lecture on beam-beam effects can only try to give a flavour of the different types of effects caused by the beam-beam interaction. I have tried to concentrate on those that are presently under study and which are considered to be a limit for foreseen and present colliders.

A list (not exhaustive !) of additional effects which are caused by beam-beam interactions must include:

- Effects in linear colliders
- Asymmetric beams (different particles types, energy etc.)
- Flip-flop effects
- Coasting beams
- Beamstrahlung
- Synchrobetatron coupling
- Monochromatization (to reduce spread of centre of mass energy)
- Beam-beam experiments
- ... and many more

Furthermore, it is impossible to treat or to mention even a fraction of the different analytical and numerical techniques developed for the study of the beam-beam interaction which largely helped to understand the phenomena and which were extensively used in the design of colliders such as HERA, LHC or B-factories.
References


[18] W. Herr, *Concept of Luminosity*, this school.


References:

[34] A. Hofmann, Landau damping, This school.
[54] O. Brünig, *Non-linear imperfections*, This school.
8 APPENDIX I

In practice, one usually derives the potential \( U(x, y, z) \) from the Poisson equation which relates the potential to the charge density distribution \( \rho(x, y, z) \).

\[
\Delta U = -\frac{1}{\epsilon_0} \rho(x, y, z)
\]

(26)

and computes the fields from:

\[
\vec{E} = -\nabla U(x, y, \sigma_x, \sigma_y)
\]

(27)

The Poisson equation can be solved using e.g. the Green’s function method (e.g. [55]) since the Green’s function for this boundary value problem is well known. The formal solution using a Green’s function \( G(x, y, z, x_0, y_0, z_0) \) is:

\[
U(x, y, z) = \frac{1}{\epsilon_0} \int G(x, y, z, x_0, y_0, z_0) : \rho(x_0, y_0, z_0) dx_0 dy_0 dz_0
\]

(28)

For the solution of the Poisson equation we get [56]:

\[
U(x, y, z, \sigma_x, \sigma_y, \sigma_z) = \frac{1}{4\pi \epsilon_0} \int \int \int \frac{\rho(x_0, y_0, z_0) dx_0 dy_0 dz_0}{\sqrt{(x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2}}
\]

(29)

In the case of a beam with Gaussian beam density distributions, we can factorize the density distribution \( \rho(x_0, y_0, z_0) = \rho(x_0) \cdot \rho(y_0) \cdot \rho(z_0) \) with r.m.s. of \( \sigma_x, \sigma_y \) and \( \sigma_z \):

\[
\rho(x_0, y_0, z_0) = \frac{Ne}{\sigma_x \sigma_y \sigma_z \sqrt{2\pi}} e^{-\frac{x_0^2}{2\sigma_x^2} - \frac{y_0^2}{2\sigma_y^2} - \frac{z_0^2}{2\sigma_z^2}}
\]

(30)

Here \( N \) is the number of particles in the bunch. We therefore have:

\[
U(x, y, z, \sigma_x, \sigma_y, \sigma_z) = \frac{1}{4\pi \epsilon_0 \sigma_x \sigma_y \sigma_z \sqrt{2\pi}} \int \int \int \frac{e^{-\frac{x_0^2}{2\sigma_x^2} - \frac{y_0^2}{2\sigma_y^2} - \frac{z_0^2}{2\sigma_z^2}} dx_0 dy_0 dz_0}{\sqrt{(x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2}}
\]

(31)

This is difficult to solve and we rather follow the proposal by S. Kheifets [7] and solve the diffusion equation:

\[
\Delta V - A^2 \cdot \frac{\partial V}{\partial t} = -\frac{1}{\epsilon_0} \rho(x, y, z) \quad \text{(for } t \geq 0)\]

(32)

and obtain the potential \( U \) by going to the limit \( A \to 0 \), i.e.

\[
U = \lim_{A \to 0} V
\]

(33)

The reason for this manipulation is that the Green’s function to solve the diffusion equation takes a more appropriate form [56]:

\[
G(x, y, z, t, x_0, y_0, z_0) = \frac{A^3}{(2\sqrt{\pi t})^3} \cdot e^{-A^2/4t \cdot ((x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2)}
\]

(34)

and we get for \( U(x, y, z, \sigma_x, \sigma_y, \sigma_z) \):

\[
\frac{Ne}{\sigma_x \sigma_y \sigma_z \sqrt{2\pi}} \cdot \frac{1}{\epsilon_0} \int_0^t dt \int \int e^{-\frac{x_0^2}{2\sigma_x^2} - \frac{y_0^2}{2\sigma_y^2} - \frac{z_0^2}{2\sigma_z^2}} \frac{A^3}{(2\sqrt{\pi t})^3} \cdot e^{-A^2/4t \cdot ((x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2)} dx_0 dy_0 dz_0
\]

(35)

407
This allows to avoid the denominator in the integral and to collect the exponential expressions which can then be integrated. Changing the independent variable $\tau = \frac{q}{\sqrt{\pi}}$ and using the formula [56] for the three integrations:

$$\int_{-\infty}^{\infty} e^{-(a\tau^2+2bu+c)} d\tau = \sqrt{\frac{\pi}{a}} e^{(2u^2)} \quad (\text{for } u = x_0, y_0, z_0)$$

(36)

we can integrate (28) and with (33) we get the potential $U(x, y, z, \sigma_x, \sigma_y, \sigma_z)$ [6, 7]:

$$U(x, y, z, \sigma_x, \sigma_y, \sigma_z) = \frac{1}{4\pi \epsilon_0} \frac{Ne}{\sqrt{\pi}} \int_0^\infty \frac{\exp(-\frac{x^2}{2\sigma_x^2+q} - \frac{y^2}{2\sigma_y^2+q} - \frac{z^2}{2\sigma_z^2+q})}{\sqrt{(2\sigma_x^2 + q)(2\sigma_y^2 + q)(2\sigma_z^2 + q)}} dq$$

(37)

Since we are interested in the transverse fields, we can work with the two-dimensional potential:

$$U(x, y, \sigma_x, \sigma_y) = \frac{ne}{4\pi \epsilon_0} \int_0^\infty \frac{\exp(-\frac{x^2}{2\sigma_x^2+q} - \frac{y^2}{2\sigma_y^2+q})}{\sqrt{(2\sigma_x^2 + q)(2\sigma_y^2 + q)}} dq$$

(38)

where $n$ is the line density of particles in the beam, $e$ is the elementary charge and $\epsilon_0$ the electrostatic constant. In this case we do not yet make any assumptions on the longitudinal distribution.
For the one-dimensional case, we write the betatron motion of a single particle as a simple harmonic oscillator and use the "smooth approximation":

\[ r = R \cos(\Phi) \]  

(39)

and its derivative:

\[ r' = -\frac{R}{\beta} \sin(\Phi) \]  

(40)

(41)

After a small kick from the beam-beam interaction the phase \( \Phi \) will be shifted and we can calculate \( d\Phi \)

\[ \delta r' \bigg|_{r \rightarrow 0} = \frac{Nr_0 \cdot r}{\gamma \sigma^2} = -\frac{Nr_0 \cdot R \cos(\Phi)}{\gamma \sigma^2} \]  

(44)

\[ 2\pi \Delta Q_i = -\frac{Nr_0 \cdot R \cos(\Phi) \cdot \cos(\Phi) / \beta}{R} \]  

(45)

\[ Fig. 30: \] Phase space before and after the beam-beam kick. Change of phase \( d\Phi \).
\[ \Delta Q_i = - \frac{N r_0 \beta}{2 \pi \gamma \sigma^2} \cdot \cos^2(\Phi) \] (46)

After averaging \( \Phi \) from 0 to 2\( \pi \)
\[ \Delta Q = - \frac{1}{2 \pi} \int_0^{2\pi} \Delta Q_i d\Phi \] (47)
we get the linear beam-beam tune shift:
\[ \Delta Q = - \frac{N r_0 \beta}{4 \pi \gamma \sigma^2} \] (48)

For the non-linear tune shift we must not linearize the beam-beam force and get for the instantaneous tune shift:
\[ \Delta Q_{i, nl} = - \frac{N r_0 \beta}{\pi \gamma} \cdot \frac{1 - e^{-\frac{R^2}{2 \pi} \cos^2(\Phi)}}{R^2} \] (49)

To perform the integral we first substitute the \( \cos^2(\Phi) \) term in the exponential by the expression:
\[ \cos^2(\Phi) = \frac{1}{2}(1 + \cos(2\Phi)) \] (50)
and then perform the integral using:
\[ \frac{1}{\pi} \int_0^{\pi} e^x \cos(\Phi) d\Phi = I_0(x) \] (51)
where \( I_0(x) \) is the modified Bessel function and get the formula for the non-linear detuning with the amplitude \( J \):
\[ \Delta Q(J) = \frac{2}{J} \cdot (1 - I_0(J/2) \cdot e^{-J/2}) \] (52)
which is \( J = \epsilon \beta /2\sigma^2 \) in the usual units.
Abstract
Analytical and numerical field computation methods for the design of conventional and superconducting accelerator magnets are presented. This report is an extract from a continuously updated eBook that can be downloaded from http://russ.home.cern.ch/russ.

1 Guiding fields for charged particles
A charged particle moving with velocity \( \mathbf{v} \) through an electro-magnetic field is subjected to the Lorentz force according to
\[
\mathbf{F} = q(\mathbf{v} \times \mathbf{B} + \mathbf{E}).
\] (1)

While the particle moves from the location \( \mathbf{r}_1 \) to \( \mathbf{r}_2 \) with \( \mathbf{v} = \frac{d\mathbf{r}}{dt} \), it changes its energy by
\[
\Delta E = \int_{\mathbf{r}_1}^{\mathbf{r}_2} \mathbf{F} d\mathbf{r} = q \int_{\mathbf{r}_1}^{\mathbf{r}_2} (\mathbf{v} \times \mathbf{B} + \mathbf{E}) d\mathbf{r} .
\] (2)

The particle trajectory \( d\mathbf{r} \) is always parallel to the velocity vector \( \mathbf{v} \). Therefore the vector \( \mathbf{v} \times \mathbf{B} \) is perpendicular to \( d\mathbf{r} \), i.e., \( (\mathbf{v} \times \mathbf{B}) d\mathbf{r} = 0 \). The magnetic field cannot contribute to a change in the particle’s energy. However, if forces perpendicular to the particle trajectory are needed, magnetic fields can serve for guiding and focusing of particle beams. At relativistic speed, electric and magnetic fields have the same effect on the particle trajectory if \( E \equiv cB \). A magnetic field of 1 T is then equivalent to an electric field of strength \( E = 3 \cdot 10^8 \) V/m. A magnetic field of one tesla strength can easily be achieved with conventional magnets (superconducting magnets on an industrial scale can reach up to 10 T), whereas electric field strength in the giga volt / meter range are technically not to be realized. This is the reason why for high energy particle accelerators only magnetic fields are used for guiding the beam.

Assuming a constant circular motion on the bending radius \( r = R \) (also called the radius of gyration) gives
\[
p = mv_0 = qB_z R .
\] (3)

In charged particle dynamics it is customary to refer to the “momentum” \( pc \) which has the dimension of an energy and to express it in units of GeV. With \( q \) expressed in units of the electronic charge, the particle momentum in GeV/c is determined by:
\[
\{p\}_{\text{GeV/c}} \approx 0.3\{q\}_e\{R\}_m \{B_z\}_T .
\] (4)

The term \( B_z R \) is called the magnetic rigidity and is a measure of the beams “stiffness” in the bending field. In circular proton machines the maximum energy is basically limited by the strength of the bending magnets. According to Eq. (4) the trajectory radius of the particle increases with the particle momentum. As both the maximum field and the maximum dimensions of the magnets are limited, the magnetic field must be ramped synchronously with the particle energy. Note that the effective radius is between 60% and 70% of the tunnel radius because of the dipole “filling factor” (space needed for focusing quadrupoles, interconnection regions, cavities etc.) and the straight parts around the collision points.
2 Conventional and superconducting magnets

Fig. 1 left, shows the conventional dipoles for the LEP (Large Electron Positron collider) and the single aperture dipole model used for testing the dipole coil manufacture for the LHC. The field calculations were performed with the CERN field computation program ROXIE. The field representations in the iron yokes are to scale, the size of the field vectors changes with the different field levels. The maximum magnetic induction in the LEP dipoles is about 0.13 T. In order to reduce the effect of remanent iron magnetization, the yoke is laminated with a filling factor of only 0.27. It can be seen that the field is dominated by the shape of the iron yoke.

If the excitational current is increased above a density of about 10 A/mm², superconducting technology has to be employed. Neglecting the quantum-mechanical nature of the superconducting material, it is sufficient to notice that the maximum achievable current density in the superconducting coil is by the factor of 500 higher than in copper coils. At higher field levels the field quality in the aperture is increasingly affected by the coil layout. Notice the large difference between the field in the aperture (8.3 T) and the field in the iron yoke (max. 2.8 T), which has merely the effect of shielding the fringe field.

3 Field quality in accelerator magnets

The quality of the magnetic field is essential to keep the particles on stable trajectories for about $10^8$ turns. The magnetic field errors in the aperture of accelerator magnets can be expressed as the coefficients of the Fourier-series expansion of the radial field component at a given reference radius (in the 2-dimensional case). In the 3-dimensional case, the transverse field components are given at a longitudinal position $z_0$ or integrated over the entire length of the magnet. For beam tracking it is sufficient to consider the transverse field components, since the effect of the z-component of the field (present only in the magnet ends) on the particle motion can be neglected. Assuming that the radial component of the magnetic flux density $B_r$ at a given reference radius $r = r_0$ inside the aperture of a magnet has been measured or calculated as a function of the angular position $\varphi$, we get for the Fourier-series expansion of the field

$$B_r(r_0, \varphi) = \sum_{n=1}^{\infty} (B_n(r_0) \sin n\varphi + A_n(r_0) \cos n\varphi),$$

Fig. 1: Magnetic field strength in the iron yoke and field vector presentation of accelerator magnets. Left: C-Core dipole ($N\cdot I = 2 \times 5250$ A, $B_1 = 0.13$ T) with a filling factor of the yoke laminations of 0.27. Right: LHC single aperture coil test facility ($N\cdot I = 480000$ A, $B_1 = 8.33$ T). Notice that even with increased field in the aperture the field strength in the yoke is reduced in the $\cos \Theta$ magnet design.
with
\[ A_n(r_0) = \frac{1}{\pi} \int_{-\pi}^{\pi} B_r(r_0, \varphi) \cos n\varphi d\varphi, \quad (n = 1, 2, 3, \ldots) \] (6)
\[ B_n(r_0) = \frac{1}{\pi} \int_{-\pi}^{\pi} B_r(r_0, \varphi) \sin n\varphi d\varphi. \quad (n = 1, 2, 3, \ldots) \] (7)

If the field components are related to the main field component \( B_N \) we get with \( N = 1 \) for the dipole, \( N = 2 \) for the quadrupole, etc.:
\[ B_r(r_0, \varphi) = B_N(r_0) \sum_{n=1}^{\infty} \left( b_n(r_0) \sin n\varphi + a_n(r_0) \cos n\varphi \right). \] (8)

The \( B_n \) are called the normal and the \( A_n \) the skew components of the field given in tesla, \( b_n \) the normal relative, and \( a_n \) the skew relative field components. They are dimensionless and are usually given in units of \( 10^{-4} \) at a 17 mm reference radius. In practice the \( B_r \) components are calculated in discrete points \( \varphi_k = \frac{k\pi}{2P} - \pi, k = 0, 1, 2, \ldots, 2P-1 \) in the interval \([-\pi, \pi)\) and a discrete Fourier transformation is carried out, i.e., for the normal component:
\[ B_n(r_0) \approx \frac{1}{P} \sum_{k=0}^{2P-1} B_r(r_0, \varphi_k) \sin n\varphi_k. \] (9)

The expression of field quality through the field components is perfectly in line with magnetic measurements using so-called harmonic coils, where the periodic variation of flux in radial or tangential rotating coils is analyzed with a Fast Fourier Transformation (FFT).

Consider a so-called tangential coil as sketched in Fig. 2 (left) rotating in the aperture of a magnet. With \( \varphi = \omega t + \Theta \), where \( \omega \) is the angular velocity, the flux linkage through the coil at time \( t \) is given by
\[ \Phi(t) = NL \int_{\varphi-\delta/2}^{\varphi+\delta/2} B_r(r_c, \varphi) r_c d\varphi \]
\[ = \sum_{n=1}^{\infty} \frac{2NLr_c}{n} \sin\left(\frac{n\delta}{2}\right) \left[ B_n(r_c) \sin(n\omega t + n\Theta) + A_n(r_c) \cos(n\omega t + n\Theta) \right], \] (10)

Fig. 2: Left: Cross-section of the long ceramic measuring shaft for the LHC magnets with the three tangential coils centered and aligned with ceramic pins. Right: Radial coil assembly.
where \( N \) is the number of turns in the rotating coil, \( rc \) is the coil radius, \( L \) is the length of the rotating coil, \( \delta \) is the opening angle of the coil and \( \Theta \) is the positioning angle at \( t = 0 \). The voltage signal at time \( t \) is then

\[
U(t) = -\frac{d\Phi}{dt} = \sum_{n=1}^{\infty} 2NLr_c\omega \sin\left(\frac{n\delta}{2}\right) [-B_n(rc) \cos(n\omega t + n\Theta) + A_n(rc) \sin(n\omega t + n\Theta)].
\]  

(11)

With the geometric parameters of the measurement coil resulting in a constant factor (which can be calculated and calibrated) the field harmonics can be obtained by means of the FFT of the voltage signal.

4 Maxwell’s equations

In this section we will present Maxwell’s equations in global and integral form, and in the form of classical vector-analysis. For the solving of Maxwell’s equations in various circumstances we further need the constitutive equations as well as the boundary and interface conditions. After a study of the properties of soft and hard magnetic materials we will be able to approximately calculate the main field in conventional accelerator magnets. We first summarize the governing laws of electromagnetism in their global form for all geometrical objects at rest.

\[
V_m(\partial a) = I(a) + \frac{d}{dt}\Psi(a),
\]

(12)

\[
U(\partial a) = -\frac{d}{dt}\Phi(a),
\]

(13)

\[
\Phi(\partial V) = 0,
\]

(14)

\[
\Psi(\partial V) = Q(V).
\]

(15)

Eq. (12) is Ampère’s magnetomotive force law and Eq. (13) is Faraday’s law of electromagnetic induction. Eq. (14) is the magnetic flux conservation law and Eq. (15) is Gauss’ fundamental theorem of electrostatics. In SI units, \( U \) denotes the electric voltage \([U] = 1 \text{ V}\) and \( V_m \), the magnetomotive force \([V_m] = 1 \text{ A}\) along the boundary \( \partial a \) of a surface \( a \). The electric flux through the boundary surface \( a = \partial V \) of a volume \( V \), is denoted \( \Psi \) with \([\Psi] = 1 \text{ C} = 1 \text{ A} \cdot \text{s}\), the magnetic flux is denoted \( \Phi \) with \([\Phi] = 1 \text{ Wb} = 1 \text{ V} \cdot \text{s}\). \( I \) is the electric current \([I] = 1 \text{ A}\) across the surface \( a \). \( Q \) is the electric charge in a volume \( V \), \([Q] = 1 \text{ C} = 1 \text{ A} \cdot \text{s}\). In integral form, Maxwell’s equations read for the stationary case in SI units:

\[
\int_{\partial a} \mathbf{H} \cdot d\mathbf{s} = \int_a \mathbf{J} \cdot d\mathbf{a} + \frac{d}{dt} \int_a \mathbf{D} \cdot d\mathbf{a},
\]

(16)

\[
\int_{\partial a} \mathbf{E} \cdot d\mathbf{s} = -\frac{d}{dt} \int_a \mathbf{B} \cdot d\mathbf{a},
\]

(17)

\[
\int_{\partial V} \mathbf{B} \cdot d\mathbf{a} = 0,
\]

(18)

\[
\int_{\partial V} \mathbf{D} \cdot d\mathbf{a} = \int_V \rho dV.
\]

(19)

The vector fields \( \mathbf{E}(t, r), \mathbf{H}(t, r) \) are the electric and magnetic field intensities, \( \mathbf{D}(t, r), \mathbf{B}(t, r) \) are the electric and magnetic induction (or flux density), \( \mathbf{J}(t, r) \) is the electrical current density. These vector fields are assumed to be finite in the entire domain and to be continuous functions of position and time. Discontinuities in the field vectors may occur, however, on surfaces with an abrupt change of the physical properties of the medium. Such discontinuities must therefore be excluded until we have treated the interface conditions in Section 7. The field intensities \( \mathbf{E} \) and \( \mathbf{H} \) are integrated along a line, \( [\mathbf{E}] = 1 \text{ V/m} \), \( [\mathbf{H}] = 1 \text{ A/m} \), whereas the flux and current densities \( \mathbf{D}, \mathbf{B} \) and \( \mathbf{J} \) are integrated over a surface, \( [\mathbf{D}] = 1 \text{ A} \cdot \text{s/m}^2 \), \( [\mathbf{B}] = 1 \text{ V} \cdot \text{s/m}^2 \), \( [\mathbf{J}] = 1 \text{ A/m}^2 \). The electric charge density \( \rho \) is integrated on a
volume, \([\rho]=1 \text{ As/m}^3\). The field intensity vectors and the flux density vectors have different natures. The line integrals of \(E\) and \(H\) are the electric voltage and magnetomotive force, respectively. The surface integrals of \(D, B, J\) are the electric flux, the magnetic flux and the electric current across the surface.

As long as the necessary conditions (continuously differentiable vector fields, smooth surfaces with simply connected, closed, piecewise smooth and consistently oriented boundary, volumes with piecewise smooth, closed and consistently oriented surface) hold for the application of the Stokes and Gauss theorem which read

\[
\int_a \text{curl} \mathbf{g} \cdot d\mathbf{a} = \int_{\partial a} \mathbf{g} \cdot d\mathbf{s}, \tag{20}
\]

\[
\int_V \text{div} \mathbf{g} \, dV = \int_{\partial V} \mathbf{g} \cdot d\mathbf{a}, \tag{21}
\]

and if we assume that the surfaces and volumina are at rest, the field equations can be written as follows:

\[
\int_a \text{curl} \mathbf{H} \cdot d\mathbf{a} = \int_a \left( \mathbf{J} + \frac{\partial}{\partial t} \mathbf{D} \right) \cdot d\mathbf{a}, \tag{22}
\]

\[
\int_a \text{curl} \mathbf{E} \cdot d\mathbf{a} = -\int_a \frac{\partial}{\partial t} \mathbf{B} \cdot d\mathbf{a}, \tag{23}
\]

\[
\int_V \text{div} \mathbf{B} \, dV = 0, \tag{24}
\]

\[
\int_V \text{div} \mathbf{D} \, dV = \int_V \rho \, dV. \tag{25}
\]

The equations can only be true for arbitrary volumes and surfaces if the following equation hold for the integrands:

\[
\text{curl} \mathbf{H} = \mathbf{J} + \partial_t \mathbf{D}, \tag{26}
\]

\[
\text{curl} \mathbf{E} = -\partial_t \mathbf{B}, \tag{27}
\]

\[
\text{div} \mathbf{B} = 0, \tag{28}
\]

\[
\text{div} \mathbf{D} = \rho. \tag{29}
\]

This is the classical vector analytical form of Maxwell’s equations. The use of notation \(\partial_t\) instead of \(\frac{\partial}{\partial t}\) is a way to establish \(\partial_t\) as an operator on the same footing as the differential operators \(\text{grad}, \text{div}\) and \(\text{curl}\). Eq. (26) - (29) are Maxwell’s equations in classical vector notation which is mainly due to O. Heaviside who eliminated the vector-potential and the scalar potential in Maxwell’s original set of equations. Divergence free vector fields such as the magnetic induction are said to be \(\text{solenoidal}\).

From the first Poincaré lemma, \(\text{div} \, \text{curl} \mathbf{g} = 0\), it follows directly that

\[
\text{div} \left( \mathbf{J} + \partial_t \mathbf{D} \right) = \text{div} \mathbf{J} + \partial_t \rho = 0 \tag{30}
\]

which is called the conservation of charge law. The commutation of the \(\text{div}\) and \(\partial_t\) operators is admissible if the fields and charge distributions are smooth. The law can be written in integral form as

\[
\int_{\partial V} \mathbf{J} \cdot d\mathbf{a} + \frac{d}{dt} \int_V \rho \, dV = 0 \tag{31}
\]

or in global form as

\[
I(\partial V) + \frac{d}{dt} Q(V) = 0. \tag{32}
\]
If at every point within a volume $V$ the charge density is constant in time we get

$$\text{div } J = 0, \quad (33)$$
$$\int_{\partial V} J \cdot da = 0, \quad (34)$$
$$I(\partial V) = 0, \quad (35)$$

the latter being known as Kirchhoff’s node-current law applied in network analysis.

5 Constitutive equations

Maxwell’s equations constitute two vector equations ($2 \times 3$ equations) and two scalar equations (altogether 8 equations) for the unknown $E, D, H, B, J$ and $\rho$ (16 unknowns). If we also consider that Eq. (28) follows from Eq. (27), then Maxwell’s equations can only be solved with the additional 9 material relations which are called the constitutive equations

$$B = \mu H, \quad D = \varepsilon E, \quad J = z E, \quad (36)$$

where $\mu, \varepsilon, z$ are the permeability $[\mu] = \frac{W}{A\cdot m}$, the permittivity $[\varepsilon] = \frac{C}{A \cdot V}$ and the conductivity $[z] = \frac{S}{V \cdot m}$, respectively. (The international IEC standard recommends to use the symbol $\sigma$ for the conductivity which is, however, also used for the surface charge density. Therefore we use the symbol $\kappa$ as proposed in DIN 1324). These most simple forms of constitutive equations hold only for linear (field independent), homogeneous (position independent), isotropic (direction independent) and stationary media. The material parameters may, however, depend on the spatial position.

If the physical properties in a specimen are the same in all directions, the material is said to be isotropic. In this case it is customary to express the permeability and permittivity as a function of the free space (vacuum) field constants with $\mu = \mu_r \mu_0$ and $\varepsilon = \varepsilon_r \varepsilon_0$ where $\mu_0 = 4\pi \cdot 10^{-7}$ H/m and $\varepsilon_0 = 8.8542\ldots \cdot 10^{-12}$ F/m. The permeability and the permittivity of free space are related through the velocity of light in vacuum by

$$c_0 = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} = 299 792 458 \text{ m/s}. \quad (37)$$

In a more general case, e.g., with a permanent magnetic or electric polarization (which are volume densities of magnetic and electric dipole moments, respectively) it will prove convenient to introduce new vectors: the electric polarization $P_{\text{el}}$ and the magnetic polarization $P_{\text{mag}}$. Often the magnetic polarization is replaced by the magnetization $M$ in units of A/m. The material relations can then be expressed as

$$B = \mu_0 H + P_{\text{mag}}(H) = \mu_0 (H + M(H)), \quad (38)$$
$$D = \varepsilon_0 E + P_{\text{el}}(E). \quad (39)$$

Note that the definition of the magnetization is not unique in literature and sometimes $M$ contains $\mu_0$.

The polarization vectors are associated with matter and vanish in free space. For linear isotropic material the polarization vectors are parallel to the field vectors and are found to be proportional according to $P_{\text{el}} = \chi_e \varepsilon_0 E$ and $M = \chi_m H$ so that for magnetic materials

$$B = \mu_0 H + \mu_0 \chi_m H = \mu_0 (1 + \chi_m) H = \mu_0 \mu_r H = \mu H, \quad (40)$$

where $\mu_r = 1 + \chi_m$ is the relative permeability, $[\mu_r] = 1_E$ and $\chi_m$ is called magnetic susceptibility $[\chi_m] = 1_E$. 

416
Fig. 3: Maxwell’s “House” [3] with scalar and vector-potential as well as the material relations. Left facade: Faraday complex. Right facade: Maxwell’s complex. Front facade: Electric fields and charges. Rear facade: Magnetic fields and impressed currents. Note how Ohm’s law and the absence of free magnetic charges spoils the otherwise perfect symmetry.

6 Maxwell’s “House”

Employing the second Lemma of Poincaré we can express the magnetic field by means of a magnetic vector potential

$$B = \text{curl} A,$$  
(41)

as the magnetic flux density is source free. If $B$ in Eq. (27) is replaced by $\text{curl} A$ one obtains $\text{curl}(E + \partial_t A) = 0$ and therefore the electric field can be expressed as

$$E = -\text{grad}\phi - \partial_t A.$$  
(42)

In the electrostatic case with $\partial_t = 0$ the electric field is curl free (irrotational), $\text{curl} E = 0$, and hence $E = -\text{grad}\phi$. Since the curl of the magnetic flux density is, in general, non-zero, it cannot always be written as the gradient of a scalar potential function. If, however, a vector-field $T$ is found such that

$$\text{curl} T = J,$$  
(43)

then the vector field $H - T$ is curl free, i.e., $\text{curl}(H - T) = 0$ and therefore $H$ can be expressed as

$$H = -\text{grad}\phi_{\text{red}} + T.$$  
(44)

$T$ is called the electric vector potential in the context of the so-called $T-\Omega$ method for steady-state field problems and $\phi_{\text{red}}$ is the reduced magnetic scalar potential. Several options for the choice of $T$ exist, [7]. The most straightforward one is to use the Biot-Savart field $H_s$ computed from the impressed current distribution. The structure of the Maxwell equations and the electromagnetic potentials is revealed in a construction called Maxwell’s “House” in [3]. The house is displayed in Fig. 3.

For magneto(quasi)static problems with vanishing time derivative ($\partial_t = 0$) only the back facade of Maxwell’s house remains standing erect. Maxwell’s equations reduce to

$$\text{curl} H = J,$$  
(45)

$$\text{div} B = 0,$$  
(46)

with the constitutive equation

$$B = \mu H.$$  
(47)
7 Boundary and interface conditions

Subsequently, the closed domain (either 2D or 3D) in which the electromagnetic field is to be calculated will be denoted as \( \Omega \). The field quantities \( B \) and \( H \) satisfy boundary conditions on the piecewise smooth boundary \( \Gamma = \partial \Omega \) of the domain \( \Omega \), see Fig. 4.

Two types of boundary conditions, prescribed on the two disjoint smooth boundaries denoted \( \Gamma_H \) and \( \Gamma_B \) with \( \Gamma = \Gamma_H \cup \Gamma_B \), cover all practical cases:

On the part \( \Gamma_B \) of the boundary the normal component of the magnetic flux density is prescribed. On symmetry planes parallel to the field, on far boundaries, or on outer boundaries of iron yokes surrounded by air (where it can be assumed that no flux leaves the outer boundary) the normal component of the flux density (denoted \( B_n \)) is zero. In some special cases the distribution of \( B_n \) can be estimated along a physical surface, e.g., the flux distribution in the air gap of an electrical machine can be assumed to be sinusoidal. These boundary conditions can be written in the form

\[
B_n = B \cdot n = \sigma_{\text{mag}} \quad \text{on } \Gamma_B, \tag{48}
\]

where \( \sigma_{\text{mag}} \) is the surface density of a fictitious magnetic charge. A surface charge is defined as a charge with infinite density, while the charge per unit surface remains finite: Consider a thin layer of thickness \( d \) in which a charge of density \( \rho_{\text{mag}} \) is present, see Fig. 5 left. In some surface \( \Delta x \Delta y \) of this layer the total charge \( \Delta Q = \Delta x \Delta y \rho_{\text{mag}} \) is present which is \( d \rho_{\text{mag}} \) per unit surface. If we let \( \rho_{\text{mag}} \to \infty \) and \( d \to 0 \) so that \( d \rho_{\text{mag}} \) remains finite, we get the surface charge with the density \( \sigma_{\text{mag}} = d \rho_{\text{mag}} \), \( [\sigma_{\text{mag}}] = 1 \text{ V} \cdot \text{s/m}^2 \).

On the part \( \Gamma_H \) of the boundary the tangential components of the magnetic field are prescribed. In many cases (as on symmetry planes perpendicular to the field) and on infinitely permeable iron poles, where the field enters at right angle, the tangential components of the field (denoted \( H_t \)) are zero. The tangential components of \( H \) can also be determined by a real or fictitious surface current density. All these boundary conditions can be written in the form

\[
H \times n = \alpha \quad \text{on } \Gamma_H, \tag{49}
\]

where \( \alpha \) is the density of a real or fictitious electric surface current. A surface current is defined as a current with infinite density on a surface, while the current per unit length remains finite: Consider a thin layer of thickness \( d \) in which a current of density \( J \) flows, see Fig. 5 right. In some length \( \Delta l \) of this layer flows the total current \( \Delta I = J d \Delta l \) which is \( J d \) per unit length. If we let \( J \to \infty \) and \( d \to 0 \) so that \( J d \) remains finite, we get the surface current with the density \( \alpha = J d \), \( [\alpha] = 1 \text{ A/m} \). The condition

Fig. 4: Composite material domain with boundary and interface.
that the tangential components are zero on the boundary implies
\[ \mathbf{H}_t = 0 \quad \rightarrow \quad \mathbf{n} \times (\mathbf{H} \times \mathbf{n}) = 0. \] (50)

In order to establish the interface conditions on a smooth surface (outer oriented by a given crossing direction) between two regions with different magnetic properties, consider two domains \( \Omega_1 \) with permeability \( \mu_1 \) and \( \Omega_2 \) with permeability \( \mu_2 \) as shown in Fig. 6.

Consider a surface element which penetrates the interface and where the vector \( da \) lies in the interface plane, as shown in Fig. 6, left. Applying Ampère’s law \( \int_{\partial a} \mathbf{H} \cdot ds = \int_a \mathbf{J} \cdot da \) to the rectangular loop while letting the height \( \delta \to 0 \) yields
\[ \int_c (\mathbf{H}_2 \cdot \frac{n_2 \times da}{da} - \mathbf{H}_1 \cdot \frac{n_2 \times da}{da}) ds = \int_c \alpha \cdot \frac{da}{da} ds. \] (51)

Eq. (51) holds for any curve \( c \), if the integrands obey
\[ (\mathbf{H}_2 - \mathbf{H}_1) \cdot \frac{n_2 \times da}{da} = \alpha \cdot \frac{da}{da}, \] (52)
\[ (\mathbf{H}_2 - \mathbf{H}_1) \times n_2 \cdot da = \alpha \cdot da. \] (53)

Eq. (53) holds for any surface element \( da \) in the plane of the interface. It yields
\[ \alpha = (\mathbf{H}_2 - \mathbf{H}_1) \times n_2 = (\mathbf{H}_1 - \mathbf{H}_2) \times n, \] (54)
where the surface normal vector \( \mathbf{n} \) points from \( \Omega_2 \) to \( \Omega_1 \) as shown in Fig. 6. If no real or fictitious electric surface currents exist, the tangential components of the magnetic field strength are continuous at the interface

\[
\mathbf{H}_{t1} = \mathbf{H}_{t2} \quad \equiv \quad (\mathbf{H}_1 - \mathbf{H}_2) \times \mathbf{n} = 0 .
\]  

(55)

Now consider the volume of the “pill-box” as shown in Fig. 6 middle. With the flux conservation law \( \int_{\partial V} \mathbf{B} \cdot d\mathbf{a} = 0 \) which holds for any closed simply connected surface we get for \( \delta \to 0 \),

\[
\int_a \sigma_{\text{mag}} da = \int_a \mathbf{B}_1 \cdot da_1 + \mathbf{B}_2 \cdot da_2 = \int_a (\mathbf{B}_1 - \mathbf{B}_2) \cdot \mathbf{n}_1 da .
\]  

(56)

Eq. (56) holds for any surface \( a \) if the integrands obey

\[
\sigma_{\text{mag}} = (\mathbf{B}_1 - \mathbf{B}_2) \cdot \mathbf{n} .
\]  

(57)

If no fictitious magnetic surface charge density exists, the normal component of the magnetic flux density is continuous at the interface

\[
B_{n1} = B_{n2} \quad \equiv \quad (\mathbf{B}_1 - \mathbf{B}_2) \cdot \mathbf{n} = 0 .
\]  

(58)

For a boundary of isotropic materials free of surface currents (Fig. 6 right) it follows that

\[
\frac{\tan \alpha_1}{\tan \alpha_2} = \frac{\mu_1}{\mu_2} = \frac{B_{t1}}{B_{t2}} = \frac{\mu_1 H_{t1}}{\mu_2 H_{t2}} = \frac{\mu_1}{\mu_2} ,
\]  

(59)

at all points \( \mathbf{x} \in \Gamma_{12} \). For \( \mu_2 \gg \mu_1 \) it follows that \( \tan \alpha_2 \gg \tan \alpha_1 \). Therefore for all angles \( \pi/2 > \alpha_2 > 0 \) we get \( \tan \alpha_1 \approx 0 \). The field exits vertically from a highly permeable medium into a medium with low permeability. We will come back to this point when we discuss ideal pole shapes of conventional magnets, see Section 12.

Remark: Both \( \mathbf{B} \) and \( \mathbf{H} \) are discontinuous at \( \mathbf{x} \in \Gamma_{12} \) and therefore \( \text{curl} \mathbf{H} \) and \( \text{div} \mathbf{B} \) cease to make sense there. For any boundary value problem defined on \( \Omega \) to be well posed, the interface conditions have to be implied. Ways of doing so include weak formulations in the Finite Element technique.

8 Magnetic anisotropy in laminated iron yokes

In case of anisotropic magnetic material the permeability has the form of a diagonal rank 2 tensor, so that \( \mathbf{B} = [\mu] \mathbf{H} \) with

\[
[\mu] = \begin{pmatrix}
\mu_x & 0 & 0 \\
0 & \mu_y & 0 \\
0 & 0 & \mu_z \\
\end{pmatrix} .
\]  

(60)

In many materials, such as in rolled metal sheets, the fabrication process produces some regularity in the crystal structure and consequently a dependence of the magnetic properties on the direction. The most well known (and strongest) anisotropy in magnetic materials can be achieved by laminating the iron yokes. Between each of the ferromagnetic laminations of thickness \( \ell_{Fe} \) (magnetically isotropic to first order) there is a non-magnetic (\( \mu = \mu_0 \)) layer of thickness \( \ell_0 \), as shown schematically in Fig. 7.

Consider a lamination in \( z \)-direction and the field components \( \mathbf{B}_t \) in the \( xy \)-plane. Because of the continuity condition \( \mathbf{H}_{t0} = \mathbf{H}_{t0}^{\text{iso}} = \overline{\mathbf{H}}_t \) we get for the effective macroscopic tangential flux density

\[
\overline{\mathbf{B}}_t = \frac{1}{\ell_{Fe} + \ell_0} \left( \ell_{Fe} \mu_0 \overline{\mathbf{H}}_t + \ell_0 \mu_0 \overline{\mathbf{H}}_t \right) .
\]  

(61)
As the normal component of the magnetic flux density is continuous, i.e., $B_z^0 = B_{zE}^0 = B_z$, the average magnetic field intensity can be calculated from

$$H_z = \frac{1}{l_{Fe} + l_0} \left( l_{Fe} \frac{B_z}{\mu} + l_0 \frac{B_z}{\mu_0} \right). \quad (62)$$

With the packing factor

$$\lambda = \frac{l_{Fe}}{l_{Fe} + l_0} \quad (63)$$

which is 0.985 for the LHC yokes, we get for the average permeability in the plane of the lamination

$$\mu_t = \lambda \mu + (1 - \lambda) \mu_0 \quad (64)$$

and normal to the plane of the lamination

$$\mu_z = \left( \frac{\lambda}{\mu} + \frac{1 - \lambda}{\mu_0} \right)^{-1}. \quad (65)$$

We have obtained a simple equation for the packing factor scaling of the material characteristic. For laminations in the $x$ and $y$ direction, i.e., with the plane of the laminations normal to the 2D cross-section, the laminations have a strong directional effect and the packing factor scaling is no longer appropriate. A macroscopic model for these circumstances is developed in [5].

9 Magnetic material

Although all materials are either ferro-, dia- or paramagnetic it is customary to talk of magnetic material only in case of ferromagnetic behavior with either a wide hysteresis curve (hard ferromagnetic material and permanent magnets) or soft ferromagnetic material with narrow hysteresis as used for yoke laminations in magnet technology.

In **diamagnetic** substances (e.g. Cu, Zn, Ag, Au, Bi) the orbit and spin magnetic moments cancel in the absence of external magnetic fields. An applied field causes the spin moments to slightly exceed the orbital moments, resulting in a small net magnetic moment which opposes the applied field. The permeability is less than $\mu_0$. In the case of water the magnetic susceptibility $\chi_m$ is $-8.8 \cdot 10^{-6}$. Superconductors in the Meissner phase represent the limiting case of $\mu = 0$, the ideal diamagnet with a
complete shielding of the external field. Diamagnetic samples brought to either pole of a magnet will be repelled.

The diamagnetic effect in materials is so low that it is easily overwhelmed in materials where the spin and orbit magnetic moments are unequal. In ideal paramagnets the individual magnetic moments do not interact with each other and take random orientation in space due to thermal agitation. When an external field is applied, the magnetic moments line up in the field direction resulting in positive susceptibilities $\chi_m$ in the order of $10^{-5}$ to $10^{-3}$ basically independent of field strength and without hysteresis behavior. Paramagnetic substances include the rare earth elements, platinum, sodium and oxygen.

9.1 Ferromagnetic material

Ferromagnetic substances (which include iron, nickel and cobalt as well as alloys of these elements) cannot be characterized by simple, single valued constitutive laws, as different $B(H)$ relations (called magnetization curves) can be measured depending on the history of the excitation. The magnetization curves, see Fig. 8, can be measured by means of so-called permeameters consisting of an annulus of ferromagnetic material with $N$ toroidal windings that excite the field of modulus $H = \frac{NI}{2\pi r}$. (66)

The induced voltage in the pick-up coil (which is wound directly onto the specimen) is proportional to the rate of change of the flux

$$U = \frac{d}{dt} \Phi = \frac{d}{dt} B_a,$$

where $a$ is the cross-section of the ring. Time integration ($\int U \, dt = B_a$) yields the corresponding values of $I$ and $U$ and consequently the hysteresis curve for $H$ and $B$. For an easy exchange of specimen these permeameters are made with split coils for $B(H)$ measurements at ambient temperatures, see Fig. 9 left.
However, the total resistance of the excitation coil is inherently high due to the large number of contacts (two per turn) resulting in high power dissipations at high currents. Thus for low temperature measurements involving superconducting excitation coils, the coil (consisting of more than 3000 turns) is wound directly onto a toroidal glass-epoxy case with an automatic winding machine. A specimen prepared for measurements is shown in Fig. 9 right.

If the applied field to a specimen is increased to saturation and then decreased, the decrease in flux density is not as rapid as the increase along the initial (virgin) magnetization curve. When $H$ reaches zero there remains a residual flux density or remanence $B_r$. In order to reduce $B$ to zero, a negative field $-H_B$ has to be applied which is called the coercive field. The phenomenon that causes $B$ to lag behind $H$, so that the magnetization curves for increasing and decreasing fields are not the same, is called hysteresis. Hysteresis curves for soft and hard ferromagnets are shown in Fig. 10. A hysteresis loop can be represented in terms $B(H)$ or $M(H)$. In a soft ferromagnet, the fields involved in the hysteresis loop are much smaller than the corresponding magnetization (Fig. 10 left) and plotting $B(H)$ instead of $M(H)$ makes only a tiny difference. However, in permanent magnet material $H$ and $M$ have the same order of magnitude and the $B(H)$ loops differs considerably from the $M(H)$ curve, see Fig. 12 left. The coercive field $H_c^B$ is in the order of 50-100 A/m in non-oriented Si-Fe alloys and low-carbon steels used in electrical motors. The low-carbon steel used for the LHC yoke laminations is specified to have a coercivity of less than 60 A/m. Low-carbon steels are good choices for yoke lamination because they are easy to handle (draw, bend, and punch) and are fairly inexpensive. Coercive fields are decreased to about 10 A/m in grain-oriented Si-Fe alloys used in transformer cores. Extremely soft materials can be obtained from nickel alloys (usually called permalloys) with about 80% of nickel and 20% of iron.

Fig. 11 shows the measured virgin $B(H)$ curve of the low carbon steel laminations used for the yoke of the LHC main dipoles, the corresponding $M(B)$ curve and the relative permeability as a function of the magnetic induction. The measurements were performed at 4.2 K with a ring specimen, a toroidal superconducting excitation coil and a copper search coil, in magnetic flux densities of up to 7.4 tesla. Table 1 gives the measured temperature and stress dependence on the coercive field, remanence and maximum permeability of yoke laminations used in a LHC model magnet.

The properties of the magnetization curves $M(H)$ are governed by two mechanisms known as exchange coupling and anisotropy. Exchange coupling between electron orbitals in the crystal lattice favors long-range spin ordering over macroscopic distances and is isotropic in space. At temperatures above a critical value (for iron about 770° C) which is called the Curie temperature, the exchange cou-

---

1Thus the need to distinguish between the two coercive fields $H_c^B$ and $H_c^M$. 

---

Fig. 9: Left: Split coil permeameter used for the warm measurements of LHC yoke iron samples (180 turns for the excitation coil, 90 turns for the pick-up coil. Right: Superconducting excitation coil (about 3000 turns) and pickup coil wound directly onto a glass-epoxy box containing the ferromagnetic specimen for cryogenic measurements.
pling disappears. Anisotropy favors spin orientation along certain symmetry axes of the lattice. The study of the quantum origin of these mechanisms is not needed in our phenomenological treatment of the material properties in field computation.

Many of the phenomena of the magnetization curve, i.e., the three sections between the “toe”, the “instep” and the “knee” can be described by means of the domain theory by Weiss, see for example [4]. In an unmagnetized (and unstrained) piece of iron the directions in which the domains are magnetized are either distributed at random (in parallel to one of the six crystal axes) or in such a way that the resultant magnetization of the specimen is zero. Application of a magnetic field only changes the direction of the magnetization in a given volume and not the magnitude. This is attained by a reversible and later irreversible boundary displacement of the domains. Saturation in high field is attained by a reversible process of rotation within the domains.

9.2 Magnetostriction

A ferromagnetic specimen changes its dimensions by some parts per million when it is magnetized. This effect is referred to as magnetostriction (positive for materials showing expansion and negative for contraction). The effect is due to magneto-crystalline anisotropy which gives rise to energy variations when the relative positions of magnetic ions in the lattice are modified. It is usually distinguished between Joule magnetostriction (the change of dimension transversely to the field) and volume magnetostriction. In case of inverse magnetostriction or stress anisotropy, the deformation caused by an externally applied stress favors certain magnetization directions.

Table 1 gives the measured temperature and stress dependence on the coercive field, remanence and maximum $\mu_r$ of yoke laminations for a LHC model magnet. An aluminum ring around the ring specimen provided for mechanical stress in the order of 20 MPa.

The main dipole magnets for LEP were built with a small packing factor of 0.27, realized by regularly spaced magnetic steel laminations and spaces filled with cement mortar. This solution provided for mechanical rigidity at low price. Mortar shrinkage at hydration had an effect on the longitudinal magnet geometry which was well controlled by means of four tie rods. In the transverse plane, however, the steel laminations opposed the shrinkage of the mortar layers so that tensile stresses built up in the mortar (about 10 MPa, near the upper limit of mortar yield strength) and compressive stresses built up in the iron laminations (at about 30 MPa due to a different elastic modulus and thickness of the layers). This resulted in an unacceptable $\Delta B/B$ in the bending field at low excitation caused by the reduction of the

![Graphs showing M(B) curves for different materials.](image-url)

Fig. 10: Left: $M(B)$ hysteresis curve for soft (3%Si-Fe) grain oriented laminations used in transformer cores. Right: $M(B)$ hysteresis curve for a sintered Fe$_{77}$Nd$_{15}$B$_8$ permanent magnet. Loop width differ by a factor of $10^{-5}$. The low-carbon steel used for the LHC yoke laminations is specified to have a coercive field $H_c$ of less than $80 \pm 10$ A/m at room temperature.
maximum $\mu_r$ due to magnetostriction. By means of a hydraulic system, transverse forces were exerted all along the poles in order to relieve the mortar induced compressive stresses in the yoke laminations.

### 9.3 Permanent magnets

In dealing with permanent magnets, the section of the hysteresis loop in the second quadrant of the $B(H)$ and $M(H)$ diagrams are of interest. If the loop is the major hysteresis loop, it is called the demagnetization curve. It is desirable that the material has a high remanence (as it determines the maximum possible flux density in a circuit) and a high coercive field $H_c^M$ so that the magnet will not easily be de-magnetized. Therefore the maximum product $(BH)_{\text{max}}$ is a good figure of merit. A discussion of the operation point of permanent magnets in a magnetic iron circuit is given in Section 10.3.

Rare earth materials like SmCo$_5$ and NdFeB are sintered from a powder with grain sizes of about 5$\mu$m. These grains are magnetically highly anisotropic along one crystalline direction. The powder is then exposed to a strong magnetic field (so that the grains rotate until their magnetically preferred axis is aligned to the magnetic field), subjected to a high pressure and sintered. Finally the sintered and machined material is exposed to a very strong magnetization field in parallel to the previously established direction. A typical $B(H)$ relationship for $B$ and $H$ parallel to the magnetically preferred axis of the grains is shown in Fig. 12 left. The demagnetization curve is for rare earth materials basically a straight line with a differential permeability of $\frac{dB}{dH} \approx 1.04\mu_0 - 1.08\mu_0$, so that the coercive field $\mu_0H_c^B$ is about 4-8% less than $B_c$. In the perpendicular direction, the typical values for the relative differential permeability are in the range of 1.02 to 1.08. Because the permeabilities are so close to $\mu_0$ we will treat the material as vacuum with an impressed (field dependent) magnetization.

![Fig. 11: Measured virgin $B(H)$ curve of the iron yoke laminations for the LHC main dipole (stressed at 20 MPa and measured at cryogenic temperature of 4.2 K), the corresponding $\mu_0M(B)$ curve and the relative permeability as a function of the magnetic induction [2].](image)

<table>
<thead>
<tr>
<th>Temperature $T$ K</th>
<th>Stress $\sigma$ MPa</th>
<th>Coercive field $H_c^B$ A/m</th>
<th>Remanence $B_r$ T</th>
<th>max $\mu_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>0</td>
<td>68.4</td>
<td>1.07</td>
<td>5900</td>
</tr>
<tr>
<td>77</td>
<td>0</td>
<td>79.6</td>
<td>1.12</td>
<td>5600</td>
</tr>
<tr>
<td>4.2</td>
<td>0</td>
<td>85.1</td>
<td>1.06</td>
<td>4800</td>
</tr>
<tr>
<td>4.2</td>
<td>20</td>
<td>110.</td>
<td>0.67</td>
<td>2460</td>
</tr>
</tbody>
</table>

Table 1: Measured temperature and stress dependence on the coercive field $H_c^B$, remanence $B_r$ and maximum $\mu_r$ of the LHC yoke laminations.
9.4 Magnetization currents and fictitious magnetic charges

In the presence of ferromagnets the magnetic field can be calculated as in vacuum, if all currents (including the magnetization currents) are explicitly considered

\[
\text{curl} \mathbf{B} = \mu_0 (\mathbf{J}_{\text{free}} + \mathbf{J}_{\text{mag}}) = \mu_0 \mathbf{J}_{\text{free}} + \mu_0 \text{curl} \mathbf{M}(\mathbf{H}).
\]

Hence

\[
\text{curl} \left( \frac{\mathbf{B} - \mu_0 \mathbf{M}(\mathbf{H})}{\mu_0} \right) = \mathbf{J}_{\text{free}}.
\]

The magnetic induction \(\mathbf{B}\) is always source free but the magnetic field \(\mathbf{H}\) is not:

\[
\text{div} \mathbf{H} = \text{div} \left( \frac{\mathbf{B} - \mu_0 \mathbf{M}(\mathbf{H})}{\mu_0} \right) = -\text{div} \mathbf{M}(\mathbf{H})
\]

which gives rise to the definition of a fictitious magnetic charge of density

\[
\rho_{\text{mag}} = -\text{div} \mu_0 \mathbf{M}(\mathbf{H}).
\]

A fictitious magnetic surface charge density, c.f. Fig. 13, is then defined as

\[
\sigma_{\text{mag}} = \mu_0 \mathbf{M}(\mathbf{H}) \cdot \frac{\mathbf{d}a}{\mathbf{d}a} = \mu_0 \mathbf{M}(\mathbf{H}) \cdot \mathbf{n}.
\]
This allows formally a treatment of magnetostatic problems in the same way as electrostatic problems using a magnetic scalar potential.

10 One-dimensional field calculation for conventional magnets

10.1 C-core dipole

Consider the magnetic circuit shown in Fig. 14 (left), a c-shaped iron yoke with two coils of all-together \( N \) turns around it. If the gap thickness is small compared to all other dimensions, the fringe field around the gap will be small and we assume that the flux of \( B \) through any cross section of the yoke (and across the air-gap) will be constant. With Ampère’s law \( \int_{\partial a} \mathbf{H} \cdot d\mathbf{s} = \int_a \mathbf{J} \cdot d\mathbf{a} \), we can write

\[
H_i s_i + H_0 s_0 = V_m, \tag{73}
\]

\[
\frac{1}{\mu_0 \mu_r} B_i s_i + \frac{1}{\mu_0} B_0 s_0 = N I. \tag{74}
\]

With \( \mu_r \gg 1 \) we get the easy relation

\[
B_0 = \frac{\mu_0 N I}{s_0}. \tag{75}
\]

10.2 Quadrupole

For the quadrupole we can split up the integration path as shown in Fig. 14 right, from the origin to the pole (\( s_1 \)) along an arbitrary path through the iron yoke (\( s_2 \)) and back inside the aperture along the x-axis (\( s_3 \)). Neglecting the magnetic resistance of the yoke we get

\[
\oint \mathbf{H} \cdot d\mathbf{s} = \int_{s_1} \mathbf{H}_1 \cdot d\mathbf{s} + \int_{s_3} \mathbf{H}_3 \cdot d\mathbf{s} = N I. \tag{76}
\]

As we will see later, in a quadrupole the field is defined by it’s gradient \( g \) with \( B_x = gy \) and \( B_y = gx \). Therefore the modulus of the field along the integration path \( s_1 \) is

\[
H = \frac{g}{\mu_0} \sqrt{x^2 + y^2} = \frac{g}{\mu_0} r. \tag{77}
\]

Along the x-axis (\( s_3 \)) the field integral is zero because \( \mathbf{H} \perp s \). Therefore

\[
\int_0^{r_0} H dr = \frac{g}{\mu_0} \int_0^{r_0} r dr = \frac{g r_0^2}{2} = N I, \tag{78}
\]

Fig. 14: Magnetic circuit of a conventional dipole magnet (left) a quadrupole magnet (right). Neglecting the magnetic resistance of the iron yoke, an easy relation between the air gap field and the required excitational current can be derived.
Fig. 15: Left: C-core dipole with excitation coil and varying yoke surface. Notice the orientation of the surface for the integration of the magnetic field strength and the surface of the integration of the magnetic flux density. Right: C-core dipole with permanent magnet excitation. Disregarding the magnetic resistance of the iron yoke and all fringe fields, an easy relation between the air-gap field and the size of the permanent magnet can be derived.

or

\[ g = \frac{2\mu_0 NI}{r_0^2}. \]  

(79)

Notice that for a given \( NI \) the field decreases linearly with the gap size of the dipole, whereas the gradient in a quadrupole magnet is inverse proportional to the square of the aperture radius \( r_0 \).

10.3 Permanent magnet excitation

For a magnet with permanent magnet excitation as shown in Fig. 15 (right), we write the equation for the magnetic circuit as

\[ H_1 s_1 + H_0 s_0 + H_m s_m = 0. \]  

(80)

In the absence of fringe fields we get with the pole surface \( a_0 \) and the magnet surface \( a_m \):

\[ B_m a_m = B_0 a_0 = \mu_0 H_0 a_0. \]  

(81)

For \( \mu_r \gg 1 \) we can again neglect the magnetic resistance of the yoke and from Eq. (80) it follows that

\[ H_0 s_0 = -H_m s_m; \]  

(82)

\[ \frac{1}{\mu_0} B_m a_m \frac{s_0}{a_0} = -H_m s_m; \]  

(83)

\[ \frac{B_m}{\mu_0 H_m} = \frac{-s_m a_0}{s_0} a_m = P; \]  

(84)

where \( P \) is called the permeance coefficient which becomes zero for \( s_0 \gg s_m \) (open circuit) and becomes \(-\infty \) for \( s_m \gg s_0 \) (short circuit). The case of \( a_m > a_0 \) is usually referred to as the “flux concentration” mode. The permeance coefficient defines the point on the demagnetization curve which is the branch of the permanent magnet hysteresis curve in the second quadrant. From Eq. (81) and (82) we derive

\[ B_m a_m s_m = \mu_0 H_0 a_0 \frac{-H_0 s_0}{H_m}. \]  

(85)

Therefore

\[ H_0 = \sqrt{\frac{(a_m s_m)(-B_m H_m)}{\mu_0 (a_0 s_0)}} = \sqrt{\frac{V_m(-B_m H_m)}{\mu_0 V_0}}. \]  

(86)
For a given magnet volume, the maximum air gap field can be obtained by dimensioning the magnetic circuit in such a way that $B_m H_m$ is maximum.

Neglecting the leakage flux may be a rough treatment of the field problem, in particular for magnetic flux densities exceeding 1 T in the yoke, or for large air gaps. Flux leakage is proportional to the magnetic potential difference, i.e., the m.m.f. $V_m = \int H \cdot ds$ between the poles. The design shown in Fig. 16 (bottom) is therefore a poor one, as there are large areas at high potential differences. The structure in Fig. 16 (top) with the permanent magnets brought to the air gap shows considerably less leakage flux. For the use of permanent magnet material in accelerator magnets, demagnetization due to irradiation and thermal fluctuations has to be considered. As no permanent damage to the crystalline structure of the magnet occurs, it is always possible to re-magnetize the magnet after irradiation to the nominal level.

11 Potential formulations for magnetostatic field problems

We will now show that in the aperture of a magnet (two-dimensional, current free region) both the magnetic scalar-potential as well as the vector-potential can be used to solve Maxwell’s equations:

$$\mathbf{H} = -\nabla \Phi = -\frac{\partial \Phi}{\partial x} \mathbf{e}_x - \frac{\partial \Phi}{\partial y} \mathbf{e}_y, \quad (87)$$

$$\mathbf{B} = \text{curl}(\mathbf{e}_z A_z) = \frac{\partial A_z}{\partial y} \mathbf{e}_x - \frac{\partial A_z}{\partial x} \mathbf{e}_y, \quad (88)$$

and that both formulations yield the Laplace equation. These fields are called harmonic and the field quality can be expressed by the fundamental solutions of the Laplace equation. Lines of constant vector-potential give the direction of the magnetic field, whereas lines of constant scalar potential define the ideal pole shapes of conventional magnets. Every vector field can be split into a source free and a curl free part. In case of the magnetic field with

$$\mathbf{H} = \mathbf{H}_s + \mathbf{H}_m, \quad (89)$$

the curl free part $\mathbf{H}_m$ arises from the induced magnetism in ferromagnetic materials and the source free part $\mathbf{H}_s$ is the field generated by the prescribed sources (can be calculated directly by means of Biot Savart’s law). With $\text{curl}\mathbf{H}_m = 0$ it follows that

$$\mathbf{H} = -\nabla \Phi_m + \mathbf{H}_s \quad (90)$$
and we get:

\[
\text{div } \mathbf{B} = 0, \\
\text{div } \mu (\nabla \Phi_m + \mathbf{H}_s) = 0, \\
\text{div } \mu \nabla \Phi_m = \text{div } \mu \mathbf{H}_s.
\]

While a solution of Eq. 93 is possible, the two parts of the magnetic field \( \mathbf{H}_m \) and \( \mathbf{H}_s \) tend to be of similar magnitude (but opposite direction) in non-saturated magnetic materials, so that cancellation errors occur in the computation. For regions where the current density is zero, however, \( \text{curl } \mathbf{H} = 0 \) and the field can be represented by a total scalar potential

\[
\mathbf{H} = -\nabla \Phi
\]

which is the Laplace equation for the scalar potential. The vector-operator Nabla is defined in Cartesian coordinates as

\[
\nabla = \left( \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right)
\]

and the Laplace operator

\[
\Delta = \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.
\]

The Laplace operator itself is essentially scalar. When it acts on a scalar function the result is a scalar, when it acts on a vector function, the result is a vector. Because of \( \text{div } \mathbf{B} = 0 \) a vector potential \( \mathbf{A} \) can be introduced: \( \mathbf{B} = \text{curl } \mathbf{A} \). We then get

\[
\text{curl } \mathbf{A} = \mu_0 (\mathbf{H} + \mathbf{M}),
\]

\[
\mathbf{H} = \frac{1}{\mu_0} \text{curl } \mathbf{A} - \mathbf{M},
\]

\[
\frac{1}{\mu_0} \text{curl } \text{curl } \mathbf{A} = \mathbf{J} + \text{curl } \mathbf{M},
\]

\[
\frac{1}{\mu_0} (-\nabla^2 \mathbf{A} + \text{grad } \text{div } \mathbf{A}) = \mathbf{J} + \text{curl } \mathbf{M},
\]

Since the curl (rotation) of a gradient field is zero, the vector-potential is not unique. The gradient of any (differentiable) scalar field \( \psi \) can be added without changing the curl of \( \mathbf{A} \):

\[
\mathbf{A}_0 = \mathbf{A} + \text{grad } \psi.
\]

Eq. (103) is called a gauge-transformation between \( \mathbf{A}_0 \) and \( \mathbf{A} \). \( \mathbf{B} \) is gauge-invariant as the transformation from \( \mathbf{A} \) to \( \mathbf{A}_0 \) does not change \( \mathbf{B} \). The freedom given by the gauge-transformation can be used to set the divergence of \( \mathbf{A} \) to zero

\[
\text{div } \mathbf{A} = 0,
\]
which (together with additional boundary conditions) makes the vector-potential unique. Eq. (104) is called the Coulomb gauge, as it leads to a Poisson type equation for the magnetic vector-potential. Therefore, from Eq. (102) we get after incorporating the Coulomb gauge:

\[ \nabla^2 \mathbf{A} = -\mu_0 (\mathbf{J} + \text{curl} \mathbf{M}). \] (105)

In the two-dimensional case with no dependence on \( z \), \( \frac{\partial}{\partial z} = 0 \) and \( J = J_z \), \( \mathbf{A} \) has only a \( z \)-component and the Coulomb gauge is automatically fulfilled. Then we get the scalar Poisson differential equation

\[ \nabla^2 A_z = -\mu_0 J_z. \] (106)

For current-free regions (e.g. in the aperture of a magnet) Eq. (106) reduces to the Laplace equation, which reads in cylindrical coordinates

\[ r^2 \frac{\partial^2 A_z}{\partial r^2} + r \frac{\partial A_z}{\partial r} + \frac{\partial^2 A_z}{\partial \varphi^2} = 0. \] (107)

### 11.1 Harmonic fields

A solution of the homogeneous differential equation (107) reads

\[ A_z(r, \varphi) = \sum_{n=1}^{\infty} (C_1 r^n + C_2 r^{-n})(D_1 \sin n\varphi + D_2 \cos n\varphi). \] (108)

Considering that the field is finite at \( r = 0 \), the \( C_2, n \) have to be zero for the vector-potential inside the aperture of the magnet while for the solution in the area outside the coil all \( C_1, n \) vanish. Rearranging Eq. (108) yields the vector-potential in the aperture:

\[ A_z(r, \varphi) = \sum_{n=1}^{\infty} r^n (C_n \sin n\varphi - D_n \cos n\varphi), \] (109)

and the field components can be expressed as

\[ B_r(r, \varphi) = \frac{1}{r} \frac{\partial A_z}{\partial \varphi} = \sum_{n=1}^{\infty} nr^{n-1}(C_n \cos n\varphi + D_n \sin n\varphi), \] (110)

\[ B_\varphi(r, \varphi) = -\frac{\partial A_z}{\partial r} = -\sum_{n=1}^{\infty} nr^{n-1}(C_n \sin n\varphi - D_n \cos n\varphi). \] (111)

Each value of the integer \( n \) in the solution of the Laplace equation corresponds to a different flux distribution generated by different magnet geometries. The three lowest values, \( n=1,2,3 \) correspond to a dipole, quadrupole and sextupole flux density distribution. The solution in Cartesian coordinates can be obtained from the simple transformations

\[ B_x = B_r \cos \varphi - B_\varphi \sin \varphi, \] (112)

\[ B_y = B_r \sin \varphi + B_\varphi \cos \varphi. \] (113)

For the dipole field \((n=1)\) we get

\[ B_r = C_1 \cos \varphi + D_1 \sin \varphi, \] (114)

\[ B_\varphi = -C_1 \sin \varphi + D_1 \cos \varphi, \] (115)

\[ B_x = C_1, \] (116)

\[ B_y = D_1. \] (117)
This is a simple, constant field distribution according to the values of $C_1$ and $D_1$. Notice that we have not yet addressed the conditions necessary to obtain such a field distribution. For the pure quadrupole (n=2) we get from Eq. (110) and (111):

$$B_r = 2r C_2 \cos 2\varphi + 2r D_2 \sin 2\varphi,$$

$$B_\varphi = -2r C_2 \sin 2\varphi + 2r D_2 \cos 2\varphi,$$

$$B_x = 2(C_2 x + D_2 y),$$

$$B_y = 2(-C_2 y + D_2 x).$$

The amplitudes of the horizontal and vertical components vary linearly with the displacements from the origin, i.e., the gradient is constant. With a zero induction in the origin, the distribution provides linear focusing of the particles. It is interesting to notice that the components of the magnetic fields are coupled, i.e., the distribution in both planes cannot be made independent of each other. Consequently a quadrupole focusing in one plane will defocus in the other.

Repeating the exercise for the case of the pure sextupole (n=3) yields:

$$B_r = 3r C_3 \cos 3\varphi + 3r D_3 \sin 3\varphi,$$

$$B_\varphi = -3r C_3 \sin 3\varphi + 3r D_3 \cos 3\varphi,$$

$$B_x = 3C_3(x^2 - y^2) + 6D_3 xy,$$

$$B_y = -6C_3 xy + 3D_3(x^2 - y^2).$$

Along the x-axis (y=0) we then get the expression for the y-component of the field:

$$B_y = D_1 + 2D_2 x + 3D_3 x^2 + 4D_4 x^3 + ...$$

If only the two lowest order elements are used for steering the beam, forces on the particles are either constant or vary linear with the distance from the origin. This is called a linear beam optic. It has to be noted that the treatment of each harmonic separately is a mathematical abstraction. In practical situations many harmonics will be present and many of the coefficients $C_n$ and $D_n$ will be non-vanishing. A successful magnet design will, however, minimize the unwanted terms to small values. It has to be stressed that the coefficients are not known at this stage. They are defined through the (given) boundary conditions on some reference radius or can be calculated from the Fourier series expansion of the (numerically) calculated field, ref. Eq. (5), in the aperture using the relations

$$A_n = nr_0^{n-1}C_n$$

and

$$B_n = nr_0^{n-1}D_n.$$  

(127)

### 12 Ideal pole shapes of conventional magnets

Equipotentials are surfaces where $\Phi$ is constant. For a path $ds$ along the equipotential it therefore results

$$d\Phi = \text{grad}\Phi \cdot ds = 0$$

(128)

i.e., the gradient is perpendicular to the equipotential. With the field lines (lines of constant vector potential) leaving highly permeable materials perpendicular to the surface (ref. Section 7), the lines of total magnetic scalar potential define the pole shapes of conventional magnets. As in 2D (with absence of magnetization and free currents) the z-component of the vector potential and the magnetic scalar potential both satisfy the Laplace equation, we already have the solution for a dipole field:

$$\Phi = C_1 x + D_1 y.$$  

(129)

So $C_1 = 0, D_1 \neq 0$ gives a vertical (normal) dipole field, $C_1 \neq 0, D_1 = 0$ yields a horizontal (skew) dipole field. The equipotential surfaces are parallel to the x-axis or y-axis depending on the values of $C$
and $D$ and results in a simple flux density distribution used for bending magnets in accelerators. For the quadrupole:

$$\Phi = C_2(x^2 - y^2) + 2D_2xy$$

(130)

with $C_2 = 0$ giving a normal quadrupole field and with $D_2 = 0$ giving a skew quadrupole field (which is the above rotated by $\pi/4$). The quadrupole field is generated by lines of equipotential having hyperbolic form. For the $C_2 = 0$ case, the asymptotes are the two major axes.

In practice, however, the magnets have a finite pole width (due to the need of a magnetic flux return yoke and space for the coil). To ensure a good field quality with these finite approximations of the ideal shape, small shims are added at the outer ends of each pole. The shim geometry has to be optimized using numerical field computation tools. Fig. 17 shows the pole shape of a conventional dipole and quadrupole magnet, with magnetic shims.

Fig. 18 shows the cross-section of the LEP dipole and quadrupole magnets with iso-surfaces of constant vector-potential (for the dipole) and magnetic field modulus in the iron yoke for the quadrupole. The field quality in the dipole was improved by adding shims on the pole surface. In case of the quadrupole, however, the pole shape is defined as a combination of a hyperbola, a straight section and an arc. The points at which the segments are connected was found in an optimization process not only considering the multipole components in the cross-section, but also to provide for a part compensation of the end-effects.

13 Coil field of superconducting magnets

For coil dominated superconducting magnets with fields well above one tesla the current distribution in the coils dominate the field quality and not the shape of the iron yoke. The problem therefore remains how to calculate the field harmonics from a given current distribution. It is reasonable to focus on the fields generated by line-currents, since the field of any current distribution over an arbitrary cross-section can be approximated by summing the fields of a number of line-currents distributed within the cross-section. For a set of $n_e$ of these line-currents at the position $(r_i, \Theta_i)$ carrying a current $I_i$, the multipole coefficients are given by [16]

$$B_n(r_0) = -\sum_{i=1}^{n_e} \frac{\mu_0 I_i}{2\pi} \frac{r_i^{n-1}}{r_i^n} \left(1 + \frac{\mu_t - 1}{\mu_t + 1} \left(\frac{r_i}{R_{yoke}}\right)^{2n}\right) \cos n\Theta_i, \quad (131)$$

433
where $R_{\text{yoke}}$ is the inner radius of the iron yoke with the relative permeability $\mu_r$. The field of any current distribution over an arbitrary cross-section can be approximated by summing the fields of a number of line-currents distributed within the cross-section. As superconducting cables are composed of single strands with a diameter of about 1 mm, a good computational accuracy can be obtained by representing each cable by two layers of equally spaced line-currents at the strand position. Thus the grading of the current density in the cable due to the different compaction on its narrow and wide side is automatically considered.

With equations (131) and (132), a semi-analytical method for calculating the fields in superconducting magnets is given. The iron yoke is represented by image currents (the second term in the parentheses). At low field level, when the saturation of the iron yoke is low, this is a sufficient method for optimizing the coil cross-section. Under that assumption some important conclusions can be drawn:

- For a coil without iron yoke the field errors scale with $1/r^n$ where $n$ is the order of the multipole
and $r$ is the mid radius of the coil. It is clear, however, that an increase of coil aperture causes a linear drop in dipole field. Other limitations of the coil size are the beam distance, the electromagnetic forces, yoke size, and the stored energy which results in an increase of the hot-spot temperature during a quench.

- For certain symmetry conditions in the magnet, some of the multipole components vanish, i.e. for an up-down symmetry in a dipole magnet (positive current $I_0$ at $(r_0, \Theta_0)$ and at $(r_0, -\Theta_0)$) no $A_n$ terms occur. If there is an additional left-right symmetry, only the odd $B_1, B_3, B_5, B_7, ..$ components remain.

- The relative contribution of the iron yoke to the total field (coil field plus iron magnetization) is for a non-saturated yoke ($\mu_t \gg 1$) approximately $\left(1 + \left(\frac{R_{\text{yoke}}}{r}\right)^{2n}\right)^{-1}$. For the main dipoles with a mean coil radius of $r = 43.5 \text{ mm}$ and a yoke radius of $R_{\text{yoke}} = 89 \text{ mm}$ we get for the $B_1$ component a 19% contribution from the yoke, whereas for the $B_5$ component the influence of the yoke is only 0.07%.

It is therefore appropriate to optimize for higher harmonics first using analytical field calculation, and include the effect of iron saturation on the lower-order multipoles only at a later stage.

14 The generation of pure multipole fields

Consider a current shell $r_i < r < r_e$ with a current density varying with the azimuthal angle $\Theta$, $J(\Theta) = J_0 \cos m\Theta$, then we get for the $B_n$ components

$$B_n(r_0) = \int_{r_1}^{r_0} \int_0^{2\pi} r_0^{n-1} - \frac{\mu_0 J_0 r_0^{n-1}}{2\pi r^{n}} \left(1 + \frac{\mu_t - 1}{\mu_t + 1} \left(\frac{r}{R_{\text{yoke}}}\right)^{2n}\right) \cos m\Theta \cos n\Theta r d\Theta dr.$$

With $\int_0^{2\pi} \cos m\Theta \cos n\Theta d\Theta = \pi \delta_{m,n}$ ($m, n \neq 0$) it follows that the current shell produces a pure $2m$-polar field and in the case of the dipole ($m = n = 1$) one gets

$$B_1(r_0) = -\frac{\mu_0 J_0}{2} \left( r_e - r_1 \right) \left( \frac{\mu_t - 1}{\mu_t + 1} \frac{1}{R_{\text{yoke}}^2} \frac{1}{3} (r_e^3 - r_1^3) \right).$$

Obviously, since $\int_0^{2\pi} \cos m\Theta \sin n\Theta d\Theta = 0$, all $A_n$ components vanish. A shell with $\cos \Theta$ and $\cos 2\Theta$ dependent current density is displayed in Fig. 21. Because of $|B| = \sum_n \sqrt{B_n^2 + A_n^2}$ the modulus of the field inside the aperture of the shell dipole without iron yoke is given as

$$|B| = \frac{\mu_0 J_0}{2} (r_e - r_1).$$

14.1 Coil-block arrangements

Usually the coils do not consist of perfect cylindrical shells because the conductors itself are either rectangular or keystoned with an insufficient angle to allow for perfect sector geometries. Therefore the shells are subdivided into coil-blocks, separated by copper wedges. The dipole coil consists of two layers with cables of the same height but of different width. Both layers are connected in series so that the current density in the outer layer, being exposed to a lower magnetic field, is about 40% higher. The conductor for the inner layer consists of 28 strands of 1.065 mm diameter, the outer layer conductor has 36 strands of 0.825 mm diameter. The strands are made of thousands of filaments of NbTi material embedded in a copper matrix which serves for stabilizing the conductor and to take over the current in case of a quench. The field generated by this coil layout is calculated with the line-current approximation of the superconducting cable. Real coil-geometries with one and two layers of coil-blocks are shown in Fig. 21.
Magnets for particle accelerators have always been a key application of numerical methods in electromagnetism. Hornsby [12], in 1963, developed a code based on the finite difference method for the solving of elliptic partial differential equations and applied it to the design of magnets. Winslow [22] created the computer code TRIM (Triangular Mesh) with a discretization scheme based on an irregular grid of plane triangles by using a generalized finite difference scheme. He also introduced a variational principle and showed that the two approaches lead to the same result. In this respect, the work can be viewed as one of the earliest examples of the finite element method applied to the design of magnets. The POISSON code which was developed by Halbach and Holsinger [11] was the successor of this code and was still applied for the optimization of the superconducting magnets for the LHC during the early design stages. Halbach had also, in 1967, [10] introduced a method for optimizing coil arrangements and pole shapes of magnets based on the TRIM code, an approach he named MIRT. In the early 1970's a general purpose program (GFUN) for static fields had been developed by Newman, Turner and Trowbridge that was based on the magnetization integral equation and was applied to magnet design.

When the LHC magnets are ramped to their nominal field of 8.33 T, two nonlinear effects on the multipole field components appear: At low field due to the superconducting filament magnetization and...
Fig. 22: Variation of the transfer function $B/I$ and the relative quadrupole field component as a function of the excitational current. Note the effect of the iron saturation at higher field levels and the superconductor magnetization at low excitational levels.

at high field due to the saturation of the iron yoke. Fig. 22 shows the transfer function and the quadrupole field component in the main bending magnets as a function of the excitational current in these magnets.

For magnets with saturating iron yoke, numerical methods have to be used to replace the imaging method. It is advantageous to use numerical methods that do not require the modeling of the coils in the finite-element mesh and allow a distinction between the coil-field and the iron magnetization effects, to confine both modeling problems on the coils and FEM-related numerical errors on the magnetization effects in the iron yoke. The integral equation method of GFUN would qualify but leads to a very large (fully populated) matrix of the linear equation system.

The disadvantage of the finite-element method is that only a finite domain can be discretized, and therefore the field calculation in the magnet coil-ends with their large fringe-fields requires a large number of elements in the air region. The relatively new boundary-element method is defined on an infinite domain and can therefore solve open boundary problems without approximation with far-field boundaries. The disadvantage is that non-homogeneous materials are difficult to consider. The BEM-FEM method couples the finite-element method inside magnetic bodies $\Omega_i = \Omega_{\text{FEM}}$ with the boundary-element method in the domain outside the magnetic material $\Omega_a = \Omega_{\text{BEM}}$, by means of the normal derivative of the vector-potential on the interface $\Gamma_{ai}$ between iron and air. The principle of the method is shown in Fig. 23. The application of the BEM-FEM method to magnet design has the following intrinsic advantages:

- The coil field can be taken into account in terms of its source vector potential $A_s$, which can be obtained easily from the filamentary currents $I_s$ by means of Biot-Savart type integrals without the meshing of the coil.

- The BEM-FEM coupling method allows for the direct computation of the reduced vector potential $A_r$ instead of the total vector potential $A$. Consequently, errors do not influence the dominating contribution $A_s$ due to the superconducting coil.

- Because the field in the aperture is calculated through the integration over all the BEM elements, local field errors in the iron yoke cancel out and the calculated multipole content is sufficiently accurate even for very sparse meshes.

- The surrounding air region need not be meshed at all. This simplifies the preprocessing and avoids artificial boundary conditions at some far-field-boundaries. Moreover, the geometry of the permeable parts can be modified without regard to the mesh in the surrounding air region, which strongly
supports the feature based, parametric geometry modeling that is required for mathematical optimization.

- The method can be applied to both 2D and 3D field problems.

15.1 LHC main dipole cross-section

The dipole magnet, its connections, and the bus-bars are enclosed in the stainless steel shrinking cylinder closed at its ends and form the dipole cold-mass, a containment filled with static pressurized superfluid helium at 1.9 K. The cold-mass, weighing about 24 tons, is assembled inside its cryostat, which comprises a support system, cryogenic piping, radiation insulation, and thermal shield, all contained within a vacuum vessel.

Fig. 24 and 25 show the quadrilateral (higher order) finite element mesh of collar and yoke for the LHC main dipole, the magnetic field strength, the magnetic vector potential and the relative permeability in the iron yoke at nominal field level.

REFERENCES

Fig. 24: Left: Quadrilateral (higher order) finite element mesh of collar and yoke for the LHC main dipole. Right: Magnetic field strength in collar and yoke

Fig. 25: Left: Magnetic vector potential; Right: Relative permeability in the iron yoke at nominal field level

[20] The LHC study group, Large Hadron Collider, The accelerator project, CERN/AC/93-03
Insertion Devices

Johannes Bahrdt
BESSY, Berlin, Germany

Abstract
The interaction of an insertion device with the electron beam in a storage ring is discussed. The radiation property including brightness, flux and polarization of an ideal and real planar and helical / elliptical device is described. The magnet design of planar, helical, quasiperiodic devices and of devices with a reduced on axis power density are resumed.

1 Introduction
Undulators and wigglers, also called insertion devices (IDs), are periodic magnetic arrays which are installed at linear or circular accelerators. The electrons are forced to move on sinusoidal or elliptical trajectories inside the devices. The strength of the IDs and hence the amplitude of the oscillations can be modiﬁed whereas the period length is usually ﬁxed. Outside of the device the trajectory is unchanged.

There are basically two areas of applications for insertion devices:
1: IDs can modify electron beam parameters such as momentum compaction factor, damping times, emittance, energy spread, spin polarization time and degree of spin polarization.
2: IDs produce synchrotron radiation with a high brightness (photons / s / mm² / mrad²) over a wide energy range and a tunable polarization. This paper concentrates mainly on this aspect.

The ﬁrst undulator has been operated at a linear accelerator of 5 MeV (100 MeV) for the generation of radiation in the mm (visible) region [1].

In 1976 with a superconducting helical undulator installed at a linear accelerator for free electron laser operation a ﬁnite gain at 10.6 µm has been demonstrated. One year later the system, now equipped with a resonator, operated above threshold at 3.4 µm [2].

A few years later the ﬁrst insertion devices for light generation had been installed in storage rings (SSRL [3], LURE [4], VEPP3 [5] [6]).

At the same time the ﬁrst so called wavelength shifters have been installed in storage rings (VEPP3 3.5 T [6], SRS T [7], VEPP2M 7.5 T [8]). These superconducting devices have signiﬁcantly higher magnetic ﬁelds than the bending magnets and hence, the radiation spectrum is shifted to the higher energies.

In ﬁrst and second generation synchrotron radiation facilities the bending magnets played a domi-
nant role as radiation sources. Third generation facilities are optimized for the operation of IDs which are superior to bending magnets with respect to brightness (ﬂux) by up to four (two) orders of magnitude (assuming a 1 mrad fan for a bending magnet).

Today about ﬁfteen third generation machines are in operation and the demand for synchrotron radiation is still increasing.

Free electron lasers have been developed at linacs and storage rings [9]. The achievable photon energies are restricted to the near VUV because the reﬂectivity of mirrors decreases dramatically with increasing energy. To overcome this problem SASE-FELs which start from noise and do not need a resonator, have been proposed and successfully built (BNL VISA [10], APS LEUTL [11], DESY TTF [12]). Saturation has been achieved and measurements agree well with theory. The shortest wavelength at which saturation has been demonstrated is 82 nm [12].

The next generation of light sources will be SASE-FEL devices [13]. Two projects aiming for 1 Å radiation have been started (DESY TESLA [14], SLAC LCSL [15]). The undulator structures for these facilities will have a length of 100–200 m.
In chapter 2 the dependence of basic beam parameters on the synchrotron radiation integrals will briefly be summarized. Chapter 3 concentrates on the motion of charged particles in permanent magnet fields. In chapter 4 the radiation characteristics including polarization will be discussed. Various technologies and undulator designs will be presented in chapter 5.

2 Beam parameters and synchrotron radiation integrals

The synchrotron radiation integrals \([16]\) describe the basic parameters of a circular accelerator.

\[
\begin{align*}
I_1 &= \int \frac{\eta(s)}{\rho} ds \\
I_2 &= \int \frac{1}{\rho^2} ds \\
I_3 &= \int \frac{1}{|\rho|^3} ds \\
I_{3a} &= \int \frac{1}{\rho^3} ds \\
I_4 &= \int \frac{(1 - 2n(s))\eta(s)}{\rho^3} ds \\
I_5 &= \int \frac{H(s)}{|\rho(s)|^3} ds \\
H(s) &= \frac{1}{\beta} [\eta^2 + <\beta \eta' - 0.5 \beta' \eta>] \\
n(s) &= \rho^2 \frac{\partial}{\partial x}(1/\rho)
\end{align*}
\]

\(\eta\) and \(\beta\) are the dispersion and the betatron function. The integrals are taken along the circumference of the ring. The following parameters depend on the synchrotron radiation integrals:

energy loss per revolution

\[
\Delta E = \frac{2}{3} \gamma_e \frac{E^4}{3(m_e c^2)^3} I_2
\]

momentum compaction factor

\[
\alpha = I_1/2\pi R
\]

damping partition numbers

\[
\begin{align*}
J_x &= 1 - I_4/I_2 \\
J_z &= 1 \\
J_\epsilon &= 2 + I_4/I_2
\end{align*}
\]

damping times

\[
\tau_i = \frac{3T_0}{r_0 \gamma^3 J_i I_2} \\
i = x, z, \epsilon
\]
energy spread

\[
\frac{(\sigma_E E)^2}{E^2} = C_q \gamma^2 \frac{I_3}{2I_2 + I_4}
\]

\[C_q = \frac{55}{32\sqrt{3}} \frac{\hbar}{m_e c}\]

emittance

\[\epsilon = C_q \gamma^2 \frac{I_5}{I_2 - I_4}\]

horizontal beam size

\[
\frac{\sigma_x^2(s)}{\beta(s)} = C_q \beta \frac{I_5}{I_2 - I_4} + \eta^2 \frac{(\sigma_E E)^2}{E^2}
\]

spin polarization time

\[1/\tau_p = \frac{5\sqrt{3} \hbar r_{ce}}{8 m_e c^2} \left( \frac{E_0}{m_e c^2} \right)^5 \frac{I_3^3}{2\pi R}\]

achievable degree of polarization

\[P_{\text{max}} = \frac{8}{5\sqrt{3}} \frac{I_{3a}}{I_3}\]

The magnetic field of an insertion device modifies the synchrotron radiation integrals. The dispersion at the location of the ID is the sum of the dispersion of the ring without ID and the dispersion of the ID [17]:

\[\eta(s) = \eta_0 + \eta'(s) + \cos(ks)/(\rho_0 k^2)\] (2)

with the undulator parameters \( \rho_0 = eB_0/\gamma m_e c \) and \( k = 2\pi/\lambda_0 \), where \( \lambda_0 \) is the undulator period. Using this dispersion and the ID local bending radius, the synchrotron radiation integrals can be integrated and the modification due to the ID can be estimated [17].

Gradient wigglers have been proposed and built in order to enhance the radial damping in combined function machines [18], [19]. These devices are installed in sections of finite dispersion and reduce the integral \( I_4 \). \( J_x \) is enhanced and the emittance is reduced. On the other hand \( J_x \) is reduced and the energy spread is increased.

Planar wigglers have been installed in damping rings to reduce the emittance (damping wigglers). It has also been proposed to install damping wigglers in future light sources in order to reduce the emittance below the limits of existing third generation storage rings [20].

Another effect of high field wigglers is the reduction of the spin polarization time in storage rings [21]. Conventional wigglers diminish the achievable degree of polarization. Asymmetric wigglers having positive and negative poles of different field strengths (keeping the net kick of the device zero) increase again the achievable polarization \( P_{\text{max}} \). At the storage ring LEP 12 asymmetric polarization wigglers had been installed [22] to provide a high polarization rate and degree.

We refer to [17] for a detailed overview on the impact of IDs on the beam parameters.
3 Charged particles in periodic magnetic structures

3.1 The equations of motion

The movement of charged particles in arbitrary magnetic fields are determined by the Lorentz force. The exact equations of motion for arbitrary magnetic fields are given by the following differential equations [23]:

\[ \begin{align*}
    x'' &= C \cdot \left[ y' B_x - (1 + x'^2) B_y + x' y' B_x \right] \\
    y'' &= -C \cdot \left[ x' B_y - (1 + y'^2) B_x + x' y' B_y \right] \\
    C &= \frac{1}{\left( \sqrt{1 + x'^2 + y'^2} \right)} \left( B \rho \right)
\end{align*} \tag{3} \]

The derivatives are taken with respect to the longitudinal coordinate \( s \). In small emittance accelerators the angles \( x' \) and \( y' \) are small and the equations can be simplified:

\[ \begin{align*}
    x'' &= \frac{1}{B \rho}[x'y'B_x - (1 + \frac{1}{2}y'^2 + \frac{3}{2}x'^2)B_y + y'B_s] \\
    y'' &= -\frac{1}{B \rho}[x'y'B_y - (1 + \frac{1}{2}x'^2 + \frac{3}{2}y'^2)B_x + x'B_s] \tag{4}
\end{align*} \]

In linear optics only (vertical) dipole and quadrupole terms are taken into account and the equations of motion reduce to [24]:

\[ \begin{align*}
    x''(s) + \left( 1/\rho^2(s) - \kappa(s) \right) \cdot x(s) &= 0 \\
    y''(s) + \kappa(s) \cdot y(s) &= 0 \tag{5}
\end{align*} \]

\( \rho \) is the dipole bending radius and \( \kappa = (e/\gamma mc) \cdot (\partial B_y)/(\partial x) \) is the quadrupole term. The coordinates \( x \) and \( y \) now relate to the closed orbit whereas in Eq. (3) a fixed cartesian coordinate system has been used.

The field of a planar undulator with \( k = 2\pi/\lambda_0 \) can be described as follows:

\[ \begin{align*}
    B_x &= \frac{k_x}{k_y} B_0 \cdot \sinh(k_x x) \cdot \sinh(k_y y) \cdot \cos(ks) \\
    B_y &= B_0 \cdot \cosh(k_x x) \cdot \cosh(k_y y) \cdot \cos(ks) \\
    B_s &= -\frac{k}{k_y} B_0 \cdot \cosh(k_x x) \cdot \sinh(k_y y) \cdot \sin(ks) \tag{6}
\end{align*} \]

Higher Fourier components of the magnetic field can be included with appropriate wave numbers \( k_n = k \cdot n \) where \( n \) is the expansion order. For elliptical or helical devices further field terms have to be added which are shifted in phase by \( \lambda_0/4 \).

\[ \begin{align*}
    \tilde{B}_x &= \frac{k_x}{k_y} \tilde{B}_0 \cdot \sinh(\tilde{k}_x x) \cdot \sinh(\tilde{k}_y y) \cdot \sin(ks) \\
    \tilde{B}_y &= \tilde{B}_0 \cdot \cosh(\tilde{k}_x x) \cdot \cosh(\tilde{k}_y y) \cdot \sin(ks) \\
    \tilde{B}_s &= \frac{k}{k_y} \tilde{B}_0 \cosh(\tilde{k}_x x) \cdot \cosh(\tilde{k}_y y) \cdot \cos(ks) \tag{7}
\end{align*} \]

This ansatz satisfies Maxwell’s equations if the wave numbers \( k \) fulfill the following relation:

\[ \begin{align*}
    k_x^2 + k_y^2 &= k^2 \\
    \tilde{k}_x^2 + k_y^2 &= k^2 \tag{8}
\end{align*} \]

The parameters \( k_x \) and \( k_y \) describe the field variation in transverse direction. The values are real (imaginary) if the fields increase (decrease) off axis.

The quantities \( x' \), \( y' \), \( 1/B \rho \) are small and therefore, the solutions of the equation of motion can be expanded analytically up to a certain order in these quantities. This will be discussed later.
3.2 Focussing of undulators

The fringe field of a dipole gives rise to focussing effects in the horizontal and vertical plane. In the horizontal plane two effects are observed [24]. The so called weak focussing in sektor magnets is due to the different path lengths of particles which are transversely displaced. The edge focussing can be observed if the electron path has a finite angle with respect to the normal of the pole face. This angle also leads to different path lengths of particles with different transverse positions. For rectangular dipoles both effects cancel each other in the horizontal plane. In the vertical plane the effect due to the edge focussing remains. An undulator can be interpreted as a series of dipoles with alternating polarities and the focussing is a sum of fringe field effects.

The vertical focussing can be readily understood from the undulator field distribution of Eq. (6). The longitudinal component has a linear dependence on the vertical coordinate. In combination with the oscillating trajectory the Lorentz force points always towards the beam axis resulting in a net focussing.

A quantitative analysis which includes also horizontal field variations starts form Eq. (5). The parameters $\rho$ and $\kappa$ vary within an undulator period and average terms $\langle K_x \rangle$ and $\langle K_y \rangle$ have to be derived where

$$\langle K_x(s) \rangle = \langle 1/\rho^2(s) - \kappa(s) \rangle$$
$$\langle K_y(s) \rangle = \langle \kappa(s) \rangle$$

For a curved trajectory we have:

$$\kappa = \frac{e}{\gamma mc} \frac{\partial B_y}{\partial \xi}$$
$$= \frac{e}{\gamma mc} \frac{\partial B_y}{\partial y}$$
$$= \frac{e}{\gamma mc} \left[ \frac{\partial B_y}{\partial x} \frac{\partial x}{\partial \xi} + \frac{\partial B_y}{\partial s} \frac{\partial s}{\partial \xi} \right]$$
$$\approx \frac{e}{\gamma mc} \left[ \frac{\partial B_y}{\partial x} - \frac{\partial B_y}{\partial s} \cdot x' \right]$$

$\xi$ is the local horizontal coordinate which is perpendicular to the trajectory. Using this expression for $k$ and the undulator fields of Eq. (6) the averaged focussing can be derived [17]:

$$\langle K_x \rangle = \frac{1}{2\rho_0^2} \left( \frac{k_x}{k} \right)^2$$
$$\langle K_y \rangle = \frac{1}{2\rho_0^2} \left( \frac{k_y}{k} \right)^2$$

The constraints on $k_x$ and $k_y$ given by Eq. (8) transform into a relation between the focussing parameters $\langle K_x \rangle$ and $\langle K_y \rangle$. In a planar undulator with infinitely wide poles the horizontal focussing is zero. Depending on the shape of the poles, horizontal focussing or defocussing can be achieved. As a consequence, the vertical focussing changes since the sum over $\langle K_x \rangle$ and $\langle K_y \rangle$ must be constant.

For arbitrary magnetic fields the focussing parameters are [25]:

$$\int K_x \, ds = \frac{1}{2} \frac{e}{\gamma mc} \int \frac{\partial^2}{\partial x^2} \left[ \langle (\int \! B_x \, ds)^2 + (\int \! B_y \, ds)^2 \rangle \right] \, ds$$
$$\int K_y \, ds = \frac{1}{2} \frac{e}{\gamma mc} \int \frac{\partial^2}{\partial y^2} \left[ \langle (\int \! B_x \, ds)^2 + (\int \! B_y \, ds)^2 \rangle \right] \, ds$$
3.3 Tracking of particles

To obtain reliable results typically several thousand turns have to be evaluated for dynamic aperture calculations. It is essential that the phase space volume does not artificially blow up due to an improper integration scheme. Phase space is conserved by canonical particle tracking. In ref. [26] and [27] an analytical approach for canonical particle tracking in arbitrary magnetic fields is presented.

In a first step, a Taylor-series expanded transformation map from the initial coordinates to the final coordinates \((x_i, x_i', y_i, y_i') \Rightarrow (x_f, x_f', y_f, y_f')\) is derived. Expansion parameters are the transverse angles \(x'\) and \(y'\) and a parameter \(x_3\) which is proportional to \(1/B\rho\). The transformation map has the form:

\[
x(s) = x_i + s \cdot x_i' + \sum_{k,l,m} a_{klm}(x_i, y_i, s) \cdot x_i'^k \cdot y_i'^l \cdot x_3^m
\]

\[
y(s) = y_i + s \cdot y_i' + \sum_{k,l,m} b_{klm}(x_i, y_i, s) \cdot x_i'^k \cdot y_i'^l \cdot x_3^m
\]

where the subscript \(i\) denotes the initial position \(s = 0\). The three magnetic field components have to be expanded in transverse direction at the location \(s\):

\[
B(x(s), y(s), s) = B(x_i, y_i, s) + \sum_{k,l,m} \frac{1}{k!l!m!} \frac{\partial^2 B(x_i, y_i, s)}{\partial x^k \partial y^l} \cdot \Delta x^k \cdot \Delta y^l
\]

with \(\Delta x = (x - x_i)\) and \(\Delta y = (y - y_i)\). The expanded coordinates and fields are inserted into Eq. (4). A comparison of coefficients of the same order provides the recursive formula for the derivatives of the expansion coefficients \(a_{klm}\) and \(b_{klm}\) in Eq. (12). The transfer map is calculated by integrating the derivatives twice from 0 to \(s\). The focussing terms of the last section are part of this general solution.

It has to be emphasized that the possibility of solving the coefficients recursively is a special feature of this particular expansion. A Taylor series expansion with respect to the transverse coordinates and angles would lead to a system of implicit differential equations which can not be solved analytically.

In a second step the generating function is derived. The variables have to be changed to canonical variables:

\[
qx = x
\]
\[
px = \tilde{A}_x/(B\rho) + x'/\sqrt{1 + x'^2 + y'^2}
\]
\[
qy = y
\]
\[
py = \tilde{A}_y/(B\rho) + y'/\sqrt{1 + x'^2 + y'^2}
\]

where \(\tilde{A}\) is the vector potential. The resulting transformation \((qx_i, px_i', qy_i, py_i') \Rightarrow (qx_f, px_f', qy_f, py_f')\) has to be converted to \((qx_i, px_i', qy_i, py_i') \Rightarrow (qx_f, px_f', qy_f, py_f')\). Based on this transformation a generating function is calculated:

\[
F = F_{00} + F_{10} \cdot px_f + F_{01} \cdot py_f + F_{20} \cdot px_f^2 + F_{11} \cdot px_f \cdot py_f + F_{02} \cdot py_f^2
\]

A canonical transformation is given by:

\[
qx_f = \frac{\partial F}{\partial px_f}
\]
\[
px_f = \frac{\partial F}{\partial qx_i}
\]

and similar for \(qy_f\) and \(py_f\). We get implicit equations for the initial and final coordinates and momenta which can easily be solved using the Newton–Raphson iterative method.

The procedure described is well suited for undulators and wigglers. For superconducting wavelength shifters the bending radius is small and hence the expansion parameter \(1/B\rho\) is rather large. As
a consequence the convergence is weak. A numerical method has been developed [28] to cope with this problem. Various trajectories with different initial coordinates are calculated and a Taylor-expansion of the generating function is fitted to these data. For the BESSY-WLS a fourth order expansion was sufficient.

3.4 Undulator–beam interaction

In third generation synchrotron radiation facilities many undulators are operated individually by the users. If the energy range of a spectrum to be measured is wider than the width of an undulator harmonic both the monochromator and the undulator have to be moved simultaneously.

It is necessary that the undulators are transparent for electron beam. Even a perfect undulator has an impact on the machine which has to be compensated:

The undulator has to be terminated such that the trajectory is without angular kick and orbit displacement. Various pole excitation patterns can be realized [29] for this purpose. At least in low energy rings active correction coils operated in correlation with the gap setting have to be used to compensate residual trajectory errors. At higher electron energies coils may not be needed if the first field integrals are below 0.1 Gm at all gaps (e.g., most IDs at the ESRF).

The edge focussing of an undulator results in a tune shift and a beta-beat. Both effects have to be compensated using an appropriate setting of the quadrupoles according to the gap variation [30].

Residual sextupole and decapole terms can be avoided if the polarity of the undulator ends are opposite.

Strong multipole wigglers produce a lengthening of the closed orbit. This effect has to be cancelled by detuning the cavity frequency.

Furthermore, magnetic field errors exist which result from imperfect magnetic material, or geometrical errors during ID assembly. These errors are usually quantified in terms of two dimensional multipoles. They have to be minimized with shimming techniques which will be described later.

4 Radiation properties

Moving charges produce two types of electric fields: the velocity field is the field of an uniformly moving charge whereas the acceleration field is related to particles which experience momentum variations due to external forces. The acceleration field as derived from the Lienard–Wiechert potential is given by:

\[
\vec{E}(t) = \frac{e}{4\pi\epsilon_0 c} \cdot \left[ \frac{1}{R} \cdot \vec{n} \times \left( \vec{n} - \vec{\beta} \right) \times \vec{\beta} \right] \cdot \text{ret} \tag{17}
\]

The bracket has to be evaluated at the retarded time \( t' = t - R(t')/c \). With \( \vec{S} = \vec{E} \times \vec{H} \) being the poynting vector, the emitted energy per time and solid angle is:

\[
\frac{\partial^2 I}{\partial t \partial \Omega} = |\vec{S}| R^2 \tag{18}
\]

A Fourier transformation yields the energy emitted per frequency interval and solid angle [31]:

\[
\frac{\partial^2 I}{\partial \omega \partial \Omega} = \frac{e^2}{16\pi^3 \epsilon_0 c} \left[ \frac{1}{R} \cdot \vec{n} \times \left( \frac{\vec{n} - \vec{\beta}}{1 - \vec{\beta} \cdot \vec{n}} \right) \right] \cdot \text{ret} \tag{19}
\]

Changing the integration variable and applying the far field approximation (\( \vec{n} = \vec{n}_0 \) and \( R(t') \approx R_0(t') - \vec{n}_0 \cdot \vec{r}(t') \), see Fig. 1) Eq. (19) becomes

\[
\frac{\partial^2 I}{\partial \omega \partial \Omega} = \frac{e^2}{16\pi^3 \epsilon_0 c} \left[ \frac{1}{R} \cdot \vec{n} \times \left( \frac{\vec{n} - \vec{\beta}}{1 - \vec{\beta} \cdot \vec{n}} \right) \right] e^{i\omega(t-\vec{n}\vec{r})} dt \tag{20}
\]
4.1 Bending magnets

In a bending magnet a charged particle moves on a circular path and Eq. (19) can be integrated [31].

\[
\begin{align*}
\frac{\partial^2 I}{\partial \omega \partial \Omega} &= \frac{3e^2}{16\pi^3c\varepsilon_0}y^2\gamma^2(1 + X^2)^2 \left| K_{2/3}(\xi), -i\sqrt{\frac{X^2}{1 + X^2}}K_{1/3}(\xi) \right|^2 \\
\xi &= \frac{y}{2}(1 + (\gamma\theta_y)^2)^{3/2} \\
y &= \frac{\omega}{\omega_c} \\
X &= \gamma\theta_y \\
\omega_c &= (3\gamma^3c)/(2\rho)
\end{align*}
\]

with \( K \) being the modified Bessel functions of the second kind and \( \omega_c \) the critical frequency which divides the power spectrum into equal parts. At the critical energy the vertical opening angle is roughly \( 1/\gamma \). For smaller (larger) energies it is given by \( \frac{1}{\gamma}(\omega_c/\omega_c)^{1/3} \) and \( \frac{1}{\gamma}(\omega_c/\omega_c)^{1/2} \) [31].

![Fig. 1: Coordinate system](image1.png)

![Fig. 2: Functions G_1 and H_2](image2.png)

The critical energy can be shifted to higher energies either with an increased electron energy or with a higher magnetic field (shorter bending radius \( \rho \)). Wavelength shifters with fields significantly higher compared to bending magnets open the hard X-ray regime for users at low energy rings. The ALS has successfully replaced three conventional bending magnets with superconducting devices to serve more beamlines for protein crystallography [32].

The first expression in the bracket of Eq. (21) describes the radiation polarized in plane and the second one the radiation polarized perpendicular to the plane of motion. The radiation is 100% linearly polarized in the plane and the fraction of vertical polarization increases for increasing observation angles.

On axis the photon flux density (photons per second, 0.1% bandwidth, mrad\(^2\)) in practical units is given by

\[
\frac{\partial^2 \tilde{F}}{\partial (\Delta \omega /\omega) \partial \Omega} = 1.327 \cdot 10^{13} \cdot E^2(\text{GeV}) \cdot I(A) \cdot H_2(y)
\]
and an integration over the vertical angle $\theta_y$ yields

\[
\frac{\partial^2 \tilde{F}}{\partial (\Delta \omega / \omega) \partial \theta_x} = 2.457 \cdot 10^{13} E(\text{GeV}) \cdot I(A) \cdot G_1(y)
\]  

(24)

The functions $G_1$ and $H_2$ are plotted in Fig. 2

4.2 Wigglers

Both undulators and wigglers are devices with a periodic magnetic field which forces the charged particles to move on sinusoidal trajectories. Undulators and wigglers differ in the $K$ parameter which can be interpreted as the ratio between the radiation divergence ($\sim 1/\gamma$) of a dipole and the maximum deflection angle of the charged particle ($K/\gamma$). For small $K$-values the light cones emitted at individual poles overlap and due to interference produce a sharply peaked line spectrum with only a few harmonics. The harmonics are well separated and the on axis flux density is proportional to a factor between $N$ and $N^2$ (depending on e-beam and beamline parameters), where $N$ denotes the number of magnet periods.

Wigglers have large $K$-values. They produce hundreds to thousands of harmonics. The harmonic number $n(\lambda)$ is given by:

\[
n = \frac{3/4K(1 + K^2/2)}{\lambda/\lambda_e}
\]  

(25)

The relative separation of the harmonics is small ($1/n$) and at high photon energies (well above the first harmonic) they are smeared due to emittance, energy spread and beamline acceptance. In this regime the spectrum can be approximated with an incoherent superposition of bending magnet spectra. At low photon energies the individual harmonics are still visible and the spectra are described with the equations described in the next section.

Although the brightness is low in the wiggler mode (the on axis flux density scales only with $N$), wigglers are widely used because they provide photons in the high energy regime which is not accessible by an undulator. The horizontal phase space distribution of the wiggler radiation can be evaluated from Eq. (26) [33]:

\[
B(x, x') = \int_{-L/2}^{L/2} \frac{\partial^2 \tilde{F}}{\partial (\Delta \omega / \omega) \partial \theta_x} \frac{1}{\rho(s)} \frac{1}{2\pi \sigma_x \sigma_{x'}} \frac{e^{-\left(\frac{(x-x_c(s)+x'(s)+x(s))}{2\sigma_x^2}\right)^2}}{e^{-\left(\frac{(x'-x'(s))}{2\sigma_{x'}^2}\right)^2}}
\]  

(26)
$x_c$ and $x'_c$ are the displacement and angle of the electrons travelling on the central orbit and $s$ is the longitudinal coordinate. The horizontal phase space of the superconducting HMI-wiggler installed at BESSY [34] has been derived using this equation (Fig. 3). A field of 7 T provides a critical energy of 16.8 keV at an electron energy of 1.9 GeV. The device will be used up to 60 keV. The two peaks separated by about 1.3 mm originate form the positive and negative poles respectively. Only the light from the poles of one polarity are used in the beamline.

### 4.3 Planar undulators

Compared to a dipole source the radiation pattern of an undulator is far more complicated. The reason are interference effects of the radiation originating at different poles of the device. Due to the sinusoidal electron trajectory and the slight difference between the speed of the electrons and the speed of light the slippage per period between the photon and the electron is

$$\lambda = \frac{\lambda_0}{2\gamma^2} \left( 1 + K^2/2 + \gamma^2 \theta^2 \right)$$

$$K = 93.4 \cdot \lambda_0 \cdot B_0$$

We observe constructive interference around the fundamental wavelength $\lambda$ and also for higher harmonics ($\lambda/n, n=$ order of harmonic) and destructive interference in between.

The electron trajectory is given by

$$x(t) = \frac{Kc}{\gamma \omega_u} \sin(\omega_u t)$$

$$s(t) = \beta ct - \frac{K^2 c}{8 \gamma^2 \omega_u} \sin(2\omega_u t)$$

where $\omega_u = 2\pi \bar{\beta} c/\lambda_0$ and $\bar{\beta}$ is the averaged velocity $\beta = v/c$ in s-direction. In a moving coordinate system the electron performs a figure eight motion since the frequency in longitudinal direction is twice the frequency in transverse direction. This longitudinal motion is the reason for a non sinusoidal electric field at the observer point. The frequency spectrum is the Fourier transform of the electric field at the retarded time. The amplitude of the longitudinal motion increases with the magnitude of the parameter $K$ and therefore the higher harmonic intensity grows (see Fig. 4).

On axis the electric field amplitudes are equally spaced and the Fourier spectrum contains only odd harmonics. Off axis the periodicity of the electric field maxima is disturbed and therefore even harmonics appear.

For periodic magnetic structures Eq. (20) can be splitted into two parts [35]:

$$\frac{\partial^2 I}{\partial \omega^2} = \frac{e^2 \omega^2 N^2}{16 \pi^4 \epsilon_0 c} \left| \int_0^{T_0} \left[ \hat{n} \times \left[ \hat{n} \times \hat{\beta} \right] \right] e^{i\omega(t - \hat{n} \cdot \hat{r} / c)} dt \right|^2,$$

$$\omega_1 = \frac{2\pi c}{\lambda_0} \frac{2\gamma^2}{1 + K^2/2 + \gamma^2 \theta^2}$$

$N$ is the number of periods and the integration is taken over one period. The expression consists of two parts: The integral is slowly varying with $\omega$ over one period whereas the last term (line shape function) is oscillating. For large $N$ sharp peaks at the location of the harmonics show up (Fig. 5). The first zero values of this function are located at $\Delta \omega = \pm \omega_1/N$ and the spectral width (FWHM) of the $n^{th}$ harmonic therefore is approximately given by

$$\frac{\Delta \omega}{\omega_n} = \frac{1}{nN}.$$
The opening angle of the $n^{th}$ harmonic can be estimated from the angle $\theta_0$ at which destructive interference is observed.

\[(N\lambda_0)(1 - \cos(\theta_0)) = \frac{\lambda}{2\lambda/L}\]

with $L = N\lambda_0$. For $\sigma_{r'} \approx \theta_0/2$ we get $\sigma_{r'} = \sqrt{\lambda/2L}$. The rms-value of the photon beam size at the center of the undulator is approximately

\[\sigma_r = \frac{1}{\pi\sqrt{2}} \sqrt{\lambda L}\]

as determined numerically. The expressions for $\sigma_r$ and $\sigma_{r'}$ are rough estimates since the phase space distribution of the photon beam for a long undulator is quite different from being Gaussian (see Fig. 15) [36], [37].
Fig. 5: The on axis lineshape function is plotted for a device with 20 periods (left). The line shape function is red shifted for off axis radiation (right, contour plot for a device with 10 periods).

For a sinusoidal magnetic field the integral of Eq. (29) can be evaluated analytically at the harmonic frequencies ($\omega = n \cdot \omega_1$) [38] [39]. The intensity distribution close to the harmonic number $n$ is then given by

$$\frac{\partial^2 I}{\partial \omega \partial \Omega} = \frac{e^2 \gamma^2 N^2}{4 \pi \epsilon_0 c} \cdot F_n(K_x, K_y, \gamma \theta, \phi) \cdot \frac{\sin^2 \left( N \pi \frac{\Delta \omega}{\omega_1(\theta)} \right)}{N^2 \sin^2 \left( \frac{\Delta \omega}{\omega_1(\theta)} \right)}$$

(33)

where $\Delta \Omega = \omega - n \omega_1$.

The terms $F_n$ are infinite sums over Bessel functions of the first kind. Walker [40] derived an analytical expression for $F_n$:

$$F_n(K_x, K_y, \gamma \theta, \gamma \phi) = \frac{n^2}{(K_x^2/2 + K_y^2/2 + (\gamma \theta^2))^2} |A_x, A_y|^2$$

(34)

In the paper the results for planar, elliptical and helical undulators are discussed. The results are in agreement with previous results which are restricted to specific designs (41 elliptical ID; [38] [42] helical ID and [38] [39] planar ID).

The interpretation of the quantities $A_x$ and $A_y$ provide a better understanding of the complicated radiation properties including polarization effects (see next chapter). It is difficult to get a simple picture for the radiation properties from a direct integration of Eq. (29). The spatial distribution of the two terms contributing to $F_n(K_x, K_y, \gamma \theta, \gamma \phi)$ in case of a planar undulator are plotted in Fig. 6. Cuts in horizontal and vertical direction are presented in Fig. 7 for planar and helical undulators.

On axis ($\theta = 0$) the photon flux density (photons per second, 0.1% bandwidth, mrad$^2$) at the $n^{th}$ harmonic is:

$$\frac{\partial^2 \hat{F}}{\partial (\Delta \omega / \omega) \partial \Omega} = 1.744 \cdot 10^{14} \cdot N^2 \cdot E^2(\text{GeV}) \cdot I(A) \cdot F_n(K)$$

(35)

$$F_n(K) = \frac{K^2 n^2}{(1 + K^2/2)^2} \cdot \left[ J_{n+1}(\zeta) - J_{n+1}(\zeta) \right]^2$$

$$\zeta = \frac{nK^2}{4(1 + K^2/2)}$$
Fig. 6: The contributions to the function $F_n(K_x, K_y, \gamma \theta_x, \gamma \theta_y)$ of Eq. (34) which depend on $|A_x|^2$ (solid lines) and $|A_y|^2$ (dotted lines), respectively, with $K_x = 0$ and $K_y = 1$.

The flux over the central cone for harmonic $n$ can be derived by integration of the line shape function over $\theta$ and $\phi$. On resonance the flux is:

$$\frac{\partial \tilde{F}}{\partial (\Delta \omega / \omega)} = 1.431 \cdot 10^{14} \cdot N \cdot Q_n \cdot I(A)$$

$$Q_n = (1 + K^2/2) \cdot F_n/n$$

The functions $F_N(K)$, $Q_n(K)$ for a planar undulator are plotted in Fig. 8 and Fig. 9. The photon flux density has a maximum exactly at the $n^{th}$ harmonic. The maximum of the photon flux is red shifted by an amount of $\Delta \omega / \omega_1 = 1/N$ and it has twice the value compared to the flux exactly at the $n^{th}$ harmonic.

In a dipole series model [43] the undulator is represented by a number of dipoles with alternating polarities. The function $F_n$ in Eq. (34) can be written as:

$$F_n(K, \gamma \theta, \gamma \phi) = \frac{3}{\pi} y^2 (1 + X^2)^2 \cdot \left[ \sin^2(\frac{\Delta}{2}) K_{2/3}^2(\xi), \cos^2(\frac{\Delta}{2}) \frac{X^2}{1 + X^2} K_{2/3}^2(\xi) \right]$$

$$\Delta = \frac{\omega}{\omega_1} (\pi + 2 \sin^{-1}(\alpha) + 3 \frac{K^2}{A} \alpha \sqrt{1 - \alpha^2})$$

$$A = 1 + K^2/2 + \gamma^2(\theta_x^2 + \theta_y^2)$$

$$\alpha = \gamma \theta_x / K.$$ 

The first (second) term in brackets refer to the horizontally (vertically) polarized light. It has been shown cited dipoles series that this expression is accurate at the level of 15% for low K and small harmonic number $n$, and is even better for higher K and $n$.

The equations above have been derived in the far field approximation. Far off axis this approximation is not valid anymore [44]. An additional $\theta$-dependent term in the expansion of $R$ is introduced:

$$R = R_0 - \tilde{n} \cdot \tilde{r} + \frac{r^2 \theta^2}{2R_0}$$

This term gives rise to an additional phase factor in Eq. (19). An analytical model is derived which describes the impact of this phase factor on the radiation properties. It is shown that the spectral and spatial width off axis can significantly be broader than expected from the far field approximation [44].
4.4 Helical and elliptical undulators

Helical (elliptical) undulators are widely used for the production of helically (elliptically) polarized light. A horizontal field $B_x$ is superimposed to the vertical field $B_y$ with a relative phase of 90 deg between the fields. The condition for constructive interference is now:

$$
\lambda = \frac{\lambda_0}{2\gamma^2} \left( 1 + K_x^2/2 + K_y^2/2 + \gamma^2 \theta^2 \right)
$$

(39)

$K_x$ and $K_y$ are defined in analogy to $K$ substituting the peak magnetic field $B_0$ with $B_x$ and $B_y$, respectively.

In a helical device the longitudinal velocity $\dot{s}(t)$ is constant, the electric field is purely sinusoidal and therefore, no higher harmonics are generated.

The photon flux density on axis and the flux integrated over the central cone can be described byFig. 7: Function $F_n(\theta, \phi)$ for a planar device with $K_y = 1.0$ horizontally (top left), vertically (top right) and for a circular device with $K_y = K_x = 1.0$ (bottom left). The harmonics $n$ 1–5 are plotted.
Eq. (35) and Eq. (36) replacing $F_n$ and $Q_n$ by

\begin{align*}
F_n(K_x, K_y) &= \frac{n^2}{(1 + K_x^2/2 + K_y^2/2)^2} \left( A_x^2 + A_y^2 \right) \\
A_x &= K_y \left( J_{n+1}(Y) - J_{n-1}(Y) \right) \\
A_y &= K_x \left( J_{n+1}(Y) + J_{n-1}(Y) \right) \\
Y &= \frac{n(K_y^2 - K_x^2)}{4(1 + K_x^2/2 + K_y^2/2)} \\
Q_n(K_x, K_y) &= \frac{(1 + K_x^2/2 + K_y^2/2)}{n} \cdot F_n(K_x, K_y)
\end{align*}

4.5 Polarization

The state of polarization for a plane monochromatic wave is characterized by the Stokes vector [45]:

\begin{align*}
S_0 &= E_x^2 + E_y^2 \\
S_1 &= E_x^2 - E_y^2 \\
S_2 &= 2 \cdot E_x E_y \cos(\delta) \\
S_3 &= 2 \cdot E_x E_y \sin(\delta)
\end{align*}

$\delta$ is the relative phase between the electric field components $E_x$ and $E_y$. The four quantities are not independent:

\begin{equation}
S_1^2 + S_2^2 + S_3^2 \leq S_0^2
\end{equation}

where the equal sign is valid for fully polarized light. $S_1$, $S_2$ and $S_3$ represent the difference between horizontal and vertical, $45^\circ$ and $-45^\circ$ and right and left handed circularly polarized light, respectively.

In a bending magnet, the phase relation between the horizontal and vertical fields can be understood from symmetry considerations. The fields are related by the expression $\vec{n} \times \left[ (\vec{n} - \vec{\beta}) \times \vec{\beta} \right]$. As illustrated in Fig. 10 and Fig. 11, for $\beta_y = \dot{\beta}_y = 0$ the electric fields have the following symmetries:
Fig. 10: Acceleration of an electron in a dipole field

Fig. 11: Off axis electric fields in 10 m distance from a 1.3 T dipole. The off axis angle is $2\gamma$ at a beam energy of 1.7 GeV

Fig. 12: Normalized Stokes parameters for dipole radiation with energies of 0.3, 1.0 and 3.0 times the critical energy.
Assuming \( n_x = 0 \), \( \Delta E_x \) is symmetric and \( \Delta E_y \) is antisymmetric with respect to the angular point of the electron trajectory. Hence, horizontal and vertical fields have a relative phase of \( \pi/2 \). The degree of circularly polarized light increases with increasing observation angle relative to the orbit plane (Fig. 12) and has opposite sign above and below the midplane.

On axis the polarization of a planar undulator is fully linearly polarized in horizontal direction. Vertically off axis, left and right circularly polarized light of successive poles add to horizontal linearly polarized light. For \( 0 \leq |\phi| \leq 90^\circ \) the angle of polarization is tilted and \( S_2 \) becomes non zero \( (S_3 \) is still zero) (see Fig. 13). For a wiggler at high harmonics there is no interference between light from neighbouring poles. Intensities and Stokes parameters increase in a cumulative way. The \( S_3 \) component of successive poles cancel which gives rise to a certain amount of unpolarized light.

On axis the Stokes parameters of elliptical (helical) devices are

\[
S_1 / S_0 = (A_x^2 - A_y^2) / (A_x^2 + A_y^2) \quad (45)
\]

\[
S_2 / S_0 = 0
\]

\[
S_3 / S_0 = -2A_xA_y / (A_x^2 + A_y^2)
\]
Fig. 14: Polarized flux density $F_p$ and degree of polarization for a helical device ($K_x = K_y = 1.0$) for the first harmonic.
where the quantities $A_x$ and $A_y$ have been defined in Eq. (40). Vertically off axis, the ellipticity of the polarization ellipse changes ($S_2$ is still zero). For $0 \leq |\phi| \leq 90^\circ$ the ellipse is tilted and all three Stokes parameters have now a finite value [46] (see Fig. 14).

4.6 Beam emittance and energy spread

The equation for undulator radiation described so far are valid for a single electron moving on an ideal trajectory. In reality, electron beam parameters and magnetic field errors degrade the spectral performance. Here, we will discuss the impact of the electron beam emittance and the energy spread. Field errors will be discussed later.

The horizontal and vertical phase space distribution of the electron beam can be approximated with two ellipses with dimensions $\sigma_{x/y}$ and $\sigma_{x'/y'}$ for the width and the divergence of the beam. In a storage ring, the area of each ellipse remains constant and is called the beam emittance. In general, the photon beam distribution has to be convoluted with these quantities. In specific cases, however (horizontal plane and low photon energies in third generation machines), the radiation is already diffraction limited and the vertical emittance can be neglected.

Kim [36] introduces the brightness $B_0$ as a quantity describing the angular and spatial photon flux density:

$$B_0(\vec{x}, \vec{\Phi}) = c \cdot \int d^2\xi \cdot A(\vec{x}, \vec{\xi}) \cdot \exp(i \cdot \frac{2\pi}{\lambda} \cdot \vec{\Phi} \cdot \vec{\xi})$$

$$A(\vec{x}, \vec{\xi}) = E_y^*(\vec{x} + \vec{\xi}/2) \cdot E_y(\vec{x} - \vec{\xi}/2) + E_z^*(\vec{x} + \vec{\xi}/2) \cdot E_z(\vec{x} - \vec{\xi}/2).$$

(46)

$B_0$ is real but not positive definite. It depends on the spatial ($\vec{x}$) and angular ($\vec{\Phi}$) coordinates and can be convoluted with a Gaussian profile according to the beam emittance. An integration over $\vec{x}$ ($\vec{\Phi}$) yields the angular (spatial) flux density [37]. These quantities are plotted in Fig. 15 for the BESSY U41 undulator.

With the assumption of a Gaussian photon beam the convolution can be done more simply. The photon angular flux density given in Eq. (35) is converted to a angular and spatial flux density which is then folded with the electron beam emittance.

$$\frac{\partial^3 \tilde{F}_\epsilon}{\partial(\Delta\omega/\omega) \partial \vec{x} \partial \Omega} = \frac{\partial^2 \tilde{F}}{\partial(\Delta\omega/\omega) \partial \Omega} \cdot \frac{\sigma^2}{\sqrt{\sigma_x^2 + \sigma_y^2} \sqrt{\sigma_x'^2 + \sigma_y'^2}}$$

(47)

The electron beam energy which is typically $8 \cdot 10^{-4}$. It is important if the increase in energy width of a harmonic due to off energy particles is in the order of the natural line width of the harmonic. For an undulator with 80 periods the 5th harmonic is already broadened by about 20%.

Many computer codes have been written for the calculation of the undulator radiation and polarization properties. Beam emittance and energy spread as well as the finite acceptance of the beamline can be included. The analytical approach (Eq. (33)) is fast but does not include the undulator end fields and field errors. A direct integration of Eq. (19) provides more information on a non-perfect undulator. Today, the cpu-time for the numerical integration of Eq. (19) is acceptable.

The evaluation of wiggler spectra in the regime of high harmonics is difficult due to numerical problems arising from the interference of radiation from neighbouring poles. Often, the wiggler spectrum is calculated for each single pole and the individual intensities are added. This procedure is valid if emittance, energy spread and beamline acceptance are dominant and if the harmonics are broadened.

4.7 Field errors and spectral performance

The equations for the on-axis flux density presented so far are valid only for ideal undulators. In reality field errors cause a reduction of the spectral performance. In electromagnetic and superconducting devices the errors are due to geometric errors. In permanent magnet undulators the errors of the magnetic
Fig. 15: Normalized electric field (top left), brightness (top right), spatial flux density (bottom left) and angular flux density (bottom right) of the BESSY U41 undulator: \( E = 1.7 \) GeV, \( \lambda = 41.2 \) mm, no of periods=79.

Material are also important. Typical errors of die pressed magnets are below 0.3\% rms for the dipole moment and below 0.3\% \( \text{irc} \) rms for the dipole angular orientation. Additionally, higher order multipoles (called inhomogeneities) are present and dominate the field quality at small gaps. These values are still too large for the construction of a high quality undulator without additional shimming.

Kincaid analytically evaluated the impact of rms-field errors on the spectral performance for two cases characterized by \( g < 1 \) and \( g > 1 \) with \( g = \sigma^2 N^3 \) [47] (\( \sigma \) is the rms peak field error and \( N \) is the number of poles). In the first case (short undulators and small errors) the phase errors produced by the random walk can be treated in a Gaussian approximation. This is not possible for large \( g \). Kincaid presents general curves for both cases describing the spectral degradation. In this context uncompensated and compensated trajectories have to be distinguished.

The correlation between the rms peak field error and the phase error is rather weak and hence the spectral performance of an undulator is not well characterized by peak field errors only [48]. Bobbs et al. [49] demonstrate that the phase error correlates much better with the spectral output. The phase error describes the deviation from the ideal slippage of one optical wavelength per period between the electron
Fig. 16: On axis flux density for the BESSY U41 undulator ($\lambda = 41.2$ mm, 79 periods). The data are given for a perfect and the actual device without emittance and energy spread (solid line), with emittance only (dashed line), with energy spread only (dotted dashed line), and with emittance and energy spread (thick solid line).

and the light and can be evaluated from the difference between the ideal and the real trajectory length:

$$\Delta \Phi = \frac{2\pi}{\beta \chi \rho_0^2} \cdot \int_{z_0}^{z} \left[ \int_{z_0}^{z'} B_y^{\text{id}} dz'' \cdot \int_{z_0}^{z'} B_y^{\text{res}} dz'' \right] dz'$$

$$+ \frac{1}{\beta \chi \rho_0^2} \cdot 0.5 \cdot \int_{z_0}^{z} \left[ \int_{z_0}^{z'} B_y^{\text{res}} dz'' \cdot \int_{z_0}^{z'} B_y^{\text{res}} dz'' \right] dz'$$

where $B_y^{\text{id}}$ and $B_y^{\text{res}}$ are the ideal magnetic field and the field errors derived from a measured field distribution. Usually, the rms-field error is not affected by the undulator shimming whereas the rms-phase error is reduced significantly.

Walker includes a Gaussian distribution of phase errors in his dipole series model and derives quantitative expressions for the on axis flux density. In Eq. (37) more terms describing the phase error are added. The reduction $R$ of the odd harmonics due to phase errors is well reproduced by the expression:

$$R = \frac{1 - e^{-\sigma_y^2 \cdot N^2}}{N^2}$$

For increasing phase errors the expression converges to $1/N$. If the trajectory errors are large, the phase errors are not Gaussian distributed anymore. Eq. (48) is still valid for the odd harmonics, the model however does not reproduce the increase of even harmonics on axis.

Today, undulators can be built with phase errors below $2^\circ$. In this regime the brightness degradation due to undulator errors is small compared to the brightness reduction due to beam emittance and energy spread (Fig. 16).
4.8 Frequency integrated power

Schwinger [50] derived a general expression for the emitted power of an electron moving along an arbitrary trajectory. According to Elleaume [51] the expression can be rewritten as:

$$\frac{\partial P}{\partial \Omega} (\text{W/mrad}^2) = 0.01344 \cdot E(\text{GeV})^2 \cdot I(\text{A}) \cdot N \cdot \int_{-\lambda_0/2}^{\lambda_0/2} \left[ \frac{v_x^2 + v_y^2}{D^3} - \frac{(v_x^2)' + (v_y^2)'}{D^5} \right] ds$$

$$D = 1 + v_x^2 + v_y^2$$

$$v_{x/y} = \gamma(\beta_{x/y} - \theta_{x/y})$$

where $\theta_x$ and $\theta_y$ are the angles of observation. The derivatives inside the integrand are taken with respect to $s$. At large distances $D$ the unit W/mrad$^2$ can be converted to W/mm$^2$ by multiplying the expression with the factor $D^2$. Close to the insertion device (typically a few meters) the depth of field effects are important and the power per area has to be derived directly from Eq. (18).

For a planar undulator the equation reduces to the expression given by Kim [52]:

$$\frac{\partial P}{\partial \Omega} (\text{W/mrad}^2) = 10.84 \cdot B_0(\text{T}) \cdot E^4(\text{GeV}) \cdot I(\text{A}) \cdot N \cdot G(K) \cdot f_k(\theta_x, \theta_y)$$

For $K \geq 1$ the function $G(K)$ is nearly one. The function $f_k(\theta_x, \theta_y)$ has a width of roughly $\pm K/\gamma$ horizontally and $1/\gamma$ vertically. In Fig. 17 the angular flux densities of planar and elliptical devices are compared.

5 Insertion device technology

5.1 Permanent magnet undulators

Electromagnetic undulators are attractive because they are mechanically simple. The gap is fixed and the variation of the photon energy is done by tuning a power supply. They are used at longer period lengths.
(above \( \approx 80 \) mm) and at moderate peak fields. At shorter period lengths the achievable magnetic fields drop significantly. The current density has a technical limit at about 2 kA/cm\(^2\). Undulators including permanent magnets are the preferred devices if shorter period lengths and higher fields are required.

In 1981 Halbach [53] proposed the design of a pure permanent magnet undulator. Each period consists of \( M \) magnets above and \( M \) magnets below the midplane. The magnetization direction between neighbouring magnets is rotated by \( 360^\circ / M \) clockwise above and counter clockwise below midplane. For infinitely wide magnets the field of such a device can be expressed analytically:

\[
B = 2B_r \sum_{\nu=0}^{\infty} \cos(nk_s) \cdot \exp(-nk_g/2) \cdot [1 - \exp(-nk_L)] \cdot \frac{\sin(n\pi/M)}{n\pi/M} \tag{51}
\]

with \( B_r = \) remanence, \( M = \) number of magnets per \( 360^\circ \) \( \epsilon r c \), \( L = \) height of magnets, \( g = \) gap and \( \epsilon = \) filling factor. In most pure permanent magnet undulators, four magnets per period are used. Six magnets per period are employed for a small period undulator [54] which enhances the field by 6\% compared to a \( M = 4 \) device.

A hybrid design was proposed by Halbach [55] in 1983. Highly permeable poles are magnetized by longitudinally oriented permanent magnets which are located between the poles. At a ratio of \( g/\lambda_0 = 0.4 \) this design provides a 10\% higher field compared to the pure permanent magnet design. The difference becomes even larger for smaller \( g/\lambda_0 \) ratios. The iron poles are saturated close to the gap and therefore less effective in that region. A wedged pole design [56] with longitudinal thicker poles reduces this local saturation and provides about 15\% higher fields compared to a straight pole design. The peak field of permanent magnet undulators can be approximated by the equation [57]:

\[
B = a \cdot \exp(b \frac{g}{\lambda} + c(\frac{g}{\lambda})^2) \tag{52}
\]

In [57] the parameters \((a, b, c)\) are given for various designs including planar pure permanent magnet devices \((a, b, c = 2.076, -3.24, 0)\) and hybrid devices \((a, b, c = 3.694, -5.068, 1.520)\).

The magnetic material which is commercially available for a reasonable price still does not have the quality with respect to dipole errors and inhomogeneities to build a high performance undulator without magnet sorting and field tuning after assembly. Careful magnet characterization permits the precise prediction of residual field errors (below 1.5 Gm for a 4m device) which is essential for a successful sorting procedure [58] [59].

After sorting, field optimization with respect to spectral performance and ID-storage ring interaction is still necessary. The field quality is improved by adding magnetic material such as iron pieces or small permanent magnets [60]. State of the art undulators are transparent to the storage ring and phase errors are typically \( 2^\epsilon r c \) which provides good performance even at the 15\(^{th}\) harmonic (about 30\% loss for an undulator with 80 periods).

### 5.2 Quasiperiodic devices

Generally, the higher harmonics of an undulator are passed through the monochromator to the sample and contaminate the spectrum. Sasaki proposed a quasiperiodic undulator which produces higher harmonics at slightly shifted positions relative to integer multiples of the fundamental [61]. The quasiperiodicity of the magnetic field can be derived from the projection of a two dimensional squared lattice onto an inclined line with inclination angle \( \alpha \). Later, Sasaki introduced a second design parameter \( r \) [62], which gives the ratio of the lattice parameters of a two dimensional rectangular grid which is again projected onto an inclined line. The quasiperiodicity is given by

\[
z_m = \frac{d}{r \cdot \tan(\alpha)} \left( m + (r \cdot \tan(\alpha) - 1) \left[ \frac{\tan(\alpha)}{r + \tan(\alpha)} n + 1 \right] \right) \tag{53}
\]
The rectangular bracket means that the largest integer number smaller than the argument of the bracket is used. This expression defines the quasiperiodic distribution of the harmonics and the first factor defines their relative intensities.

In the original design with $M = 2$ (Fig. 18) two different pole distances ($z_{m+1} - z_m$) have been implemented according to Eq. (53) with $r = 1$ [63]. A more compact magnetic structure based on a conventional undulator design with $M = 4$ (Fig. 19) has been proposed in [64] [65]. Here, certain magnets (defined by Eq. (53)) are removed [64] or retracted [65]. The magnet retraction offers the possibility to optimize the flux of certain harmonics.

Measurements illustrate the effectivity of higher order suppression [66] [67]. The quasiperiodic scheme has been adopted also to the hybrid undulator (Fig. 20) [85] or electromagnetic undulator [68]. Fig. 21 shows the spectra of two BESSY undulators with a periodic and quasiperiodic field distribution.

5.3 Asymmetric wigglers and elliptical wigglers

In the high energy regime which can not be reached with undulators elliptical and asymmetric wigglers are the only choice though they have a rather low brightness compared to undulators.

The spectral characteristics of asymmetric or elliptical wigglers can be understood from the bending magnet properties since the contributions from the individual poles add incoherently. In a conventional undulator the magnetic field distribution of positive and negative poles have the same shape. Off axis the right and left handed circular polarized light of successive poles cancel. An asymmetric wiggler is composed of strong and weak poles with opposite polarity. The weak poles extend over a larger region in order to guarantee a zero field integral within one period. At a distance of $\lambda_0/2$ to the strong pole the field is close to zero.

Well above the critical energy for the weak poles the flux from the strong poles dominate the spectrum and off axis circular polarized light from the strong poles can be observed. Below the critical
energy the flux for the weak poles increases, and the degree of circularly polarized light decreases.

Asymmetric wigglers for the production of circularly polarized light have been proposed [69] [70]
and built [71], [72], [73] at various laboratories.

Different technologies have been applied to produce asymmetric fields. Pure permanent magnets
[71] and hybrid designs [72] [73] [74] have been realized. Recently a 3.1 T (11 mm gap) 2 pole hybrid
device has been built at the ESRF [75].

A superconducting device installed at DELTA can be operated either in a symmetric mode with
2.8 T peak field or in an asymmetric mode yielding 5.3 T [76].

Asymmetric wigglers have several disadvantages compared to elliptical wigglers. Due to the off
axis observation the vertical source size increases (dependent on the length of the device) which translates
into a reduced brightness. The polarized flux is a factor of 2–4 smaller. Finally, the slope of the flux with
observation angle is finite which makes the photon flux at the experiment sensitive to angle errors of the
electron orbit or of beamline components.

Elliptical wigglers have been proposed in [77]. They consist of two magnet arrays with strong and
weak magnetic fields perpendicular to each other. Due to the horizontal magnetic field the observation
angle between the trajectory and the midplane changes sign at each pole. As a consequence, the circularly
polarized light has always the same helicity and adds in a cumulative fashion over the length of the device.

The first two elliptical wigglers have been installed at the accumulation ring of the TRISTAN ring
and at the Photon Factory [78], [79]. The devices have a pure permanent magnet design. The two arrays
which are perpendicular to each other can be moved in gap and phase individually. The maximum fields
on axes are 1.0 T vertically and 0.2 T horizontally. Helicity switching can be done by moving the arrays
by λ/0 with respect to each other.

At Spring-8 a planar elliptical wiggler has been built [80]. Three permanent magnet rows above
and below the midplane provide both field components. The outer arrays can be moved longitudinally
for polarization control.

For many experiments it is useful to switch the polarization with 10–100 Hz which can not be
done mechanically. This demand initiated the development of partly electromagnetic devices. Walker
et al. proposed to use electromagnets for the horizontal structure to increase the switching rate [81]. A
five pole device built at the APS in collaboration with the Budker institute and the NSLS and installed at
the NSLS has electromagnets for the horizontal field and a wedged pole hybrid structure vertically [82].
Later, a similar device with 36 poles (straight poles) has been installed at the APS [83]. The devices can
be operated with switching rates of 10 Hz [83] and 100 Hz [82], respectively.

5.4 Helical (elliptical) undulators

Helical devices generate only one harmonic on axis and the degree of circular polarization is close to one.
Higher harmonics can only be produced for with non helical fields. The degree of polarization is reduced
in this case. The figure of merit usually used is the product \( S_0 \cdot (S_3/S_0)^2 \). This product is plotted in
Fig. 22 versus the ratio of horizontal to vertical fields. With increasing harmonic number the maximum
of the figure of merit shifts towards lower field ratios: \( K_x/K_y = 1.0, 0.4, 0.32, 0.27 \) for the harmonics
1, 3, 5, 7. The corresponding degrees of polarization are 1.0, 0.78 – 0.89 (1 ≤ K ≤ 2.5), 0.73 – 0.84
(1.3 ≤ K ≤ 2.5), 0.70 – 0.80 (1.5 ≤ K ≤ 2.5) for the harmonics 1, 3, 5, 7, respectively [102].

5.4.1 Electromagnetic devices

Electromagnetic elliptical undulators have the potential for fast helicity switching. The current variation
follows a trapezoidal shape and the rise time has to be minimized in order to optimize the duty cycle.
Switching frequencies are usually limited to 10 Hz in contrast to elliptical wigglers where a sinusoidal
current variation of up to 100 Hz is possible.
A fast switching elliptical undulator ($\lambda_0 = 80$ mm) has been installed at the ESRF [84]. The horizontal (vertical) fields are produced from permanent magnets (electromagnets). The helicity can be switched within 6 ms and switching rates are 10 Hz. For long period devices electromagnets can be used for both planes which simplifies the mechanics and reduces the costs. Various devices of such kind have been built. The ELETTRA device (212 mm) [87] can switch with 1 Hz and 100 Hz in trapezoidal and sinusoidal mode, respectively. In a similar device for the SLS (212 mm) [68] fast switching is not intended and therefore, a solid iron joke is used. The APS undulator (128 mm) can switch with 10 Hz with a rise time of 20 ms [86].

High switching frequencies introduce eddy currents in the vacuum chamber. The time dependent field integrals have to be minimized with a sophisticated feed forward correction scheme.

### 5.4.2 Permanent magnet helical (elliptical) undulators

Small period lengths require either the permanent magnet or the superconducting technology. Halbach [53] proposed a helical devices made of $N \cdot m$ permanent magnet dipoles which are rotated with respect to each other by $2\pi/m$. The wavelength and polarization of such a device is fixed and there is no access from the side.

Onuki [88] proposed and built an elliptical / helical undulator with two perpendicular magnet arrays. The arrays could be shifted and the gap could be changed for an independent polarization and wavelength control. The minimum gap was restricted to 5.7 cm and there was no side access. At SuperACO [89] an electromagnetic version of the Onuki design has been installed. A gap variation is not needed whereas the polarization is modified via a longitudinal movement of the coils.

A crossed undulator has been proposed independently by Moiseev [90] and Kim [91]. A short crossed undulator has been built at BESSY [92]. The device consists of two planar devices tilted by $90^\circ$ with respect to each other. Between the undulators a modulator matches the relative phase between the light and the electron beam. The two wave trains are elongated in a monochromator and overlap to produce circularly polarized light at the sample. The disadvantage of the crossed undulator is the strong variation of the state of polarization over the width of the harmonic and over the central cone. Emittance
and energy spread reduce the degree of polarization especially for higher harmonics.

About a decade ago several planar helical (elliptical) undulators have been invented and realized. The degree of polarization is insensitive to emittance and energy spread. The devices have no magnets at the side and therefore, have no limitation for the minimum gap. The magnetic measurements can be done conveniently with a bench at the side of the device and the installation of the device into the storage ring is simple.

The horizontal and vertical fields of the ESRF design [93] are produced by the upper and lower magnet arrays, respectively. Both arrays can be moved independently in vertical direction and in addition the upper array can be moved longitudinally to provide full tuning capability with respect to wavelength and polarization. The vertical focusing of the lower structure has to be compensated with a long coil.

The ELETTRA design [94] provides higher fields and hence a larger tuning range. The drawback is the fixed state of polarization (circular) and the lack of higher harmonics.

In 1992 Sasaki [95] invented the so called APPLE (Advanced Polarizing Photon Light Emitter) undulator. This device consists of four magnet arrays, two above and two below the midplane. Row shift and gap movement permit a wavelength tuning and a complete polarization control. In an improved design [96] the blocks originally magnetized under 45° were replaced by vertically magnetized blocks. This device produces the highest fields among all planar helical (elliptical) undulators and is, therefore, the preferred design in many laboratories.

The fields of an APPLE undulator are composed of the fields from four individual rows. Assuming that two diagonal rows are fixed and the other two rows are moved by $\phi_1$ and $\phi_2$ the resulting fields $A_{y/z}$ and the corresponding phases are given by:

$$B_{y/z} = B^0_{y/z} \cdot \{2 \cdot \cos(k \cdot z) \pm \cos(k \cdot z + \phi_1) \pm \cos(k \cdot z + \phi_2)\} \tag{54}$$

$$A_{y/z} = \frac{B^0_{y/z}}{\cos(\delta_{y/z})} \cdot \{2 \pm \cos(\phi_1) \pm \cos(\phi_2)\}$$

Fig. 23: Relative phases between the horizontal and vertical field of an APPLE undulator
The relative phase $\delta_{y/z}$ is plotted in Fig. 23 versus the phases $\phi_1$ and $\phi_2$. Moving the two rows parallel the relative phase between the field components is $\pi/2$ whereas it is 0 if the rows are moved in opposite direction. In this way circular (elliptical) or linear polarization can be realized. The inclination angle of the linear polarized light can be controlled via the row phase (see Fig. 24) between 0 and 90°. This range can be enlarged to 180° if the other two rows are movable as well.

The ratio of $B_y/B_z$ of the APPLE II changes with gap and, hence, the magnet rows have to be moved during an energy scan in order to keep the polarization fixed. At SPRING-8 [97] a helical undulator with three magnet rows has been developed. The center rows producing the vertical fields are moved. The magnets are designed such that the ratio $B_y/B_z$ is nearly independent from gap. This magnet design is suitable for in vacuum undulators where a longitudinal drive system can hardly be realized. Similar devices have been installed also at HISOR [98] and UVSOR [99]. Grooves in the magnets of the center row improve the transverse field homogeneity [97] [98].

A novel magnet design [100] provides a degree of polarization close to one over a wide energy range. The undulator has different period lengths at the upper ($\lambda_0$) and the lower ($\lambda_0/3$) magnet array and can be used in three different modes. Helical fields are produced using either the upper or the lower

\[
\delta_{y/z} = \arctan \frac{\pm \sin(\phi_1) \pm \sin(\phi_2)}{2 \pm \cos(\phi_1) \pm \cos(\phi_2)}
\]
magnet array. In the combined mode the projected trajectory has the shape of a rombus and higher harmonics with a high degree of polarization are observed.

5.5 Helicity switching

The helicity switching with permanent magnet undulators has to be done mechanically and the switching frequency is limited to about 1 Hz. Higher frequencies can be realized using a double undulator which produces both helicities simultaneously. Appropriate choppers and apertures periodically absorb one or the other beam.

The first double undulator has been installed at the ESRF applying the HELIOS design [101], [51]. The photon beams of the two undulators are spatially separated with chicane magnets and a chopper in a beamline selects either right or left handed polarized light. At BESSY a similar scheme is realized. Two double undulators of the APPLE II type have been installed [102]. Rotatable permanent magnets permit the adjustment of the angle of separation between the two light beams.

At the SLS [103] an APPLE II type double undulator and an electromagnetic double undulator is in operation. In contrast to the ESRF and the BESSY design the two light beams are parallel with respect to each other and displaced in transverse direction. In the focal plane of a focusing mirror the beams are separated and can be selected with a chopper.

Recently, another switching scheme has been realized at SPRING 8. The electron beam is deflected periodically in such a way that the light of either the upstream or the downstream undulator is passed through the beamline [104] [97].

Up to now there are no technical solutions available for the longitudinal movement of magnet rows under vacuum conditions. Polarization switching is possible only with double undulators or with special devices. In a parabolic undulator [105] the period length for the horizontal field is half the period length of the vertical field and the relative phase is 0°. The projected trajectory has the shape of a parabola. On axis the polarization is horizontal linear and vertical linear for successive harmonics. Horizontally off axis the degree of circularly polarized light increases. Off resonance two peaks with opposite helicity are generated at positive and negative observation angles. These two beams can be focussed onto a sample and helicity switching can be done with a chopper.

The figure-8 undulator [106] has a similar magnet design except for a phase of π/2 between horizontal and vertical fields. This device has been optimized for a reduced on axis power density. The light is linearly polarized in the horizontal (vertical) direction at the harmonics 0.5, 1.5, 2.5 (1.0, 2.0, 3.0).

An asymmetric figure-8 configuration [108] provides also elliptically polarized light. In this design the vertical field is produced by electromagnets which can reverse helicity and vary the asymmetry and thus the degree of polarization.

5.6 Reduced power on axis

At high K-values planar devices deliver high on axis intensities at the higher harmonics. This can be harmful to the beamline optics and various schemes have been realized or proposed to minimize these problems: i) On axis helical devices produce only flux at the first harmonic. The power distribution has a minimum on axis and increases for larger observation angles. An appropriate pinhole removes most of the unwanted power. ii) A figure-8 undulator [106] delivers horizontal and vertical linearly polarized light at various harmonics. The power density on axis is reduced compared to a planar device. iii) Sasaki [107] proposed a double undulator where the first device produces right and the second one left handed circularly polarized light. Depending on the relative phase between the two segments the coherent superposition of the two light beams yields linearly polarized light with variable angle. No higher harmonics are present and the on axis flux density is small. Similar to the crossed undulator the double undulator is rather sensitive to parameters like beam emittance energy spread and beamline
acceptance. T. Tanaka and H. Kitamura [109] proposed a multisegment helical undulator with opposite helicities of successive segments. An analytical investigation shows that this design delivers a high degree of linear polarization with efficient suppression of higher harmonics. The sensitivity on electron beam and beamline parameters is smaller compared to the double undulator.

The problem of unwanted high on axis power is closely related to the usual design of undulators where the period length is fixed and the energy tuning is done via the tuning of the $K$-parameter. Tatchyn [110] discusses three classes of undulators with fixed period (1), fixed field (2) and fixed $K$-parameter (3). Type (2) and (3) require the variation of the period length for energy tuning. Tatchyn demonstrates that the ratio of useful flux to total flux is more favourable in the variable period devices which are operated at moderate $K$-values. The performance of variable period devices has been discussed also in [111]. The design consists of an array of staggered permeable poles surrounded by a solenoid. The energy tuning is done by moving the poles longitudinally and changing the field of the solenoid. Electromagnetic undulators have the potential to be operated at different discrete period lengths: The PTB-U180 [112] has been designed such that the period length can be doubled by rewiring the coils which can be done within a few hours.

5.7 Short period undulators

For a fixed electron energy the photon energy is tuned by variation of the magnetic field strength. At the high energy side there is a cut-off at $K = 0$ and the flux decreases already significantly for $K \leq 0.5$ (first harmonic). Higher energies can be achieved either with higher harmonics or shorter period lengths. The overlap between the first and third harmonic requires a maximum $K$-value of at least 2.5. Going to smaller period lengths this requirement can be fulfilled only by decreasing in addition the minimum magnetic gap and hence, the vacuum aperture. In a storage ring the minimum fixed vertical aperture should not be defined by the insertion devices. As a consequence, flexible vacuum chambers have been implemented at SSRL [113], DESY [114], NSLS [115], MAXLAB [116] which can restore the full vertical aperture if necessary. Minimum magnetic gaps of 5.5 mm [115] and 7.7 mm [116] have been realized this way.

The fabrication of flexible chambers is technically complicated. There is still a difference of about 2 mm between the magnetic gap and the vacuum gap. This triggered the development of a technology to install the magnets directly into the vacuum. The magnet girders of the first BESSY I undulator were installed inside a vacuum tank [117]. The magnets themselves were encapsulated in a thin metal foil to prevent a contamination of the ultra high vacuum. The first in-vacuum device having the magnets inside the ultra high vacuum has been built at KEK [118]. The sintered permanent magnets which are porous have been Ni-plated to reduce outgassing. A stainless steel foil covering the magnets and flexible joints at the undulator ends reduce image current heating.

The basic technologies for in-vacuum undulators have been further developed and optimized at SPRING8 where most of the insertion devices are in vacuum devices [119]. The magnets are coated with TiN (NdFeB) or Ni (Sm2Co17). Cu plated Ni foils on top of the magnets and water cooled RF-fingers at the undulator ends reduce the effects of image current heating and permit the use of small magnetic gaps. The magnetic material has to be chosen carefully because the material can be demagnetized by radiation [122]. Detailed studies on demagnetization effects have been done using different magnet grades from different suppliers in various geometries [123]. Appropriate thermal aging can reduce the radiation sensitivity of the magnets [124].

Conventional shimming (placing iron or permanent magnet pieces on top of the magnets) is not applicable for in-vacuum undulators. A sophisticated initial magnet sorting is essential and a rearrangement of magnets inside the assembled structure (in-situ sorting) further improves the field quality [58].

Today, in-vacuum undulators are based on a mature technology and they are used in several laboratories (SPRING8 [125], NSLS [115] [120], ESRF [126], SLS [127]).
Superconducting small period undulators can provide even higher fields compared to permanent magnet devices. Madey’s first FEL undulator consisted of a superconducting bifilar helix [2]. At Brookhaven a planar superconducting undulator has been developed for a harmonic generation FEL experiment [128]. A 3.8 mm period length superconducting undulator has been built at the Forschungszentrum Karlsruhe and tested at the Mainz microtron MAMI [129]. Based on these experiences a 14 mm period device has been developed by the Forschungszentrum Karlsruhe and ACCEL. Devices with 10 and 50 periods have been built and a 100 period device is under construction [130].

Currently many laboratories are building prototypes of superconducting small period undulators in order to push the technology [131]. Using NbTi-wires at 4 K a gain in magnetic field of almost a factor of two compared to in-vacuum undulators is expected. Nb$_3$Sn-wires which are technologically more complicated provide another factor of about two.

### 5.8 High field devices

The critical energy is proportional to the magnetic field strength. High field devices extend the energy range beyond the cut off frequency of the bending magnets. Devices with one main pole (wavelengths shifter) and with many poles (multipole wiggler) are in operation.

A permanent magnet asymmetric wiggler ($\lambda_0 = 378$ mm) with a peak field of 3.13 T (3.57 T) at a magnetic gap of 11 mm (6 mm) has been built at the ESRF [75]. The period length can be reduced to 218 mm providing the same peak field if a symmetric configuration is chosen.

At the Budker Institute a 5 T non superconducting wavelength shifter has been designed [132]. The magnet consist of coils and permanent magnets.

For higher fields and/or shorter period lengths superconducting (SC) coils are required. The strongest wavelength shifter with a peak field of 10 T has been built by the Budker Institute for SPRING8 [133].

The flux from a multipole wiggler is proportional to the number of poles. The brightness of the source is limited by depth of field effects which are less severe for shorter period lengths. At MAXLAB a SC cold bore 46-pole multipole wiggler has been developed and built [134]. The peak field is 3.5 T at a period length of 61 mm and a gap of 10.2 mm. A similar device has been installed at ELETTRA [135] (peak field = 3.5 T for $\lambda_0 = 64$ mm and gap = 16.5 mm).

A 7 T SC multipole wiggler with 17 poles has recently been installed at BESSY [34]. The period length is only 148 mm and the magnetic gap is 19 mm.

A SC multipole wiggler at DELTA [76] can be operated as a symmetric or an asymmetric wiggler which provides some flexibility to the users. The parameters for the two modes are: B = 2.79 T (5.30 T), $\lambda_0 = 144$ mm (288 mm), number of poles = 20 (10) in the symmetric and antisymmetric mode, respectively. The NSLS SC multipole wiggler [136] can be operated in three different modes: 1) 11 poles at 3 T, 2) 5 poles at 4.7 T and 3) 1 pole at 5.5 T.

More details on the technology of SC high field devices can be found in [131].

**References**


Proceedings of the 24th ICFA Beam Dynamics Workshop on Future Light Sources, SPring-8, Hyogo, Japan, 2002.


[131] Proceedings of the "Workshop on superconducting undulators and wigglers", June 30th–July 1st, organized by P. Elleaume at the ESRF, Grenoble, France
Free-electron lasers

Peter Schmüser
Institut für Experimentalphysik, Universität Hamburg, Germany

Abstract
The synchrotron radiation of relativistic electrons in undulator magnets and the low-gain Free-Electron Laser (FEL) are discussed in some detail. The high-gain FEL based on the principle of Self Amplified Spontaneous Emission is treated on a qualitative level.

1 Introduction
The Free-Electron Laser (FEL) principle has been known since the early 1970’s but for many years FEL’s have played a marginal role in comparison with conventional lasers. Only in recent years it has become clear that these devices have the potential of becoming exceedingly powerful light sources in the vacuum-ultraviolet (VUV) and X ray regime. In my talk I will first deal with undulator radiation since it is intimately related to FEL radiation, then explain the low-gain FEL and finally treat the high-gain FEL based on the principle of Self Amplified Spontaneous Emission (SASE). SASE-FEL’s are frequently considered as the fourth generation of accelerator-based light sources. In contrast to existing synchrotron radiation light sources which are mostly storage rings the FEL requirements on the electron beam quality in terms of low transverse emittance and small energy spread are so demanding that only linear accelerators can be used to provide the drive beam.

In my one-hour talk it was not possible to go much into mathematical details. The high-gain FEL is therefore treated only qualitatively. For a thorough presentation of SASE FEL’s I refer to the book “The Physics of Free Electron Lasers” by Saldin, Schneidmiller and Yurkov and to the lectures by J. Rossbach at the CERN Accelerator School on Synchrotron Radiation.

1.1 Electron accelerators as light sources
In the bending magnets of a high-energy circular accelerator the relativistic electrons emit synchrotron radiation. For a Lorentz factor $\gamma = W/(m_e c^2) \gg 1$ the radiation is emitted almost tangentially to the circular orbit. The frequency spectrum is continuous and extends from zero to frequencies beyond the “critical frequency”

$$\omega_c = \frac{3 c \gamma^3}{2 R}$$  \hspace{1cm} (1)

where $R$ is the radius of curvature in the bending magnet. The radiation power is

$$P_{\text{rad}} = \frac{e^2 c}{6 \pi \varepsilon_0} \frac{\gamma^4}{R^2}.$$  \hspace{1cm} (2)

In modern synchrotron light sources the radiation used for research is produced in wiggler or undulator magnets which are periodic arrangements of many short dipole magnets of alternating polarity. The electrons move on a wavelike orbit through such a magnet (Fig. 1) but the overall deflection of the beam is zero. Undulator radiation is far more useful than bending-magnet radiation because it is nearly monochromatic and concentrated in a narrow angular cone with an opening angle of about $\pm 1/\gamma$. The wavelength can be estimated from the following consideration. Call $\lambda_u$ the period of the magnet arrangement. In a coordinate system moving with the average speed of the beam, the relativistic length contraction reduces the period to $\lambda_u^* = \lambda_u / \gamma$, and the electrons oscillate at a correspondingly higher...
Fig. 1: Schematic representation of undulator radiation. For simplicity the alternating magnetic field and the cosine-like electron orbit have been drawn in the same plane.

frequency $\omega^* = \frac{2\pi c}{\lambda^*_u} = \gamma \frac{2\pi c}{\lambda_u}$ and emit dipole radiation. If one Lorentz-transforms this radiation into the laboratory system one gets for the light wavelength $\lambda_\ell \approx \frac{\lambda^*_u}{\gamma} = \frac{\lambda_u}{\gamma^2}$; for example, for a Lorentz factor $\gamma = 1000$ (an electron energy of 511 MeV) the radiation wavelength is a million times shorter than the undulator period. Moreover, the wavelength can be easily varied by changing the particle energy.

It is interesting to note that the total energy radiated by a relativistic electron in an undulator is the same as that in a bending magnet of equal magnetic length, however, the intensity is concentrated in a narrow spectral range. Different electrons radiate independently in bending magnets as well as in undulators, hence the total power produced by a bunch of $N$ electrons is simply $N$ times the radiation power of one electron.

1.2 Free-electron and conventional lasers

The next big improvement is given by the Free-Electron Laser. The main component is again an undulator magnet but by means of a clever mechanism (explained below) one forces a large number of electrons to emit their radiation coherently. Like undulator radiation, the FEL radiation is almost monochromatic and well collimated but the power will be $N$ times higher if one manages to achieve full coherence in the bunch.

A conventional laser (Fig. 2) consists of three basic components: the laser medium with at least 3 energy levels, an energy pump which creates a population inversion, and an optical resonator. The electrons are bound to atomic, molecular or solid-state levels, so one may call this a “bound-electron” laser in contrast to the free-electron laser where the electrons move in vacuum. In an FEL (Fig. 3 ) the role of the active laser medium and the energy pump are both taken over by the relativistic electron beam. An optical cavity is no longer possible for wavelengths below 100 nm, because the reflectivity of metals and other mirror coatings drops quickly to zero at normal incidence. Here one has to rely on the principle of Self Amplified Spontaneous Emission (SASE) where the laser gain is achieved in a single passage of a very long undulator magnet. The schematic setup of a SASE FEL is shown in Fig. 4. One big advantage of an FEL in comparison with a conventional laser is the free tunability of the wavelength by simply changing the electron energy.
2 Undulator radiation

2.1 Magnetic field of undulator

The motion of an electron in an undulator magnet is shown schematically in Fig. 5. The undulator axis is along the direction of the beam (z direction), the magnetic field points in the y direction (vertical). The period of the magnet arrangement $\lambda_u$ is in the order of 25 mm. For simplicity we assume that the horizontal width of the pole shoes is larger than $\lambda_u$, then the x dependence of the field can be neglected. The field on the axis is approximately harmonic

$$B_y(0, 0, z) = B_0 \cos(k_u z) \quad \text{with} \quad k_u = 2\pi/\lambda_u$$

(3)

In vacuum we have $\vec{\nabla} \times \vec{B} = 0$, hence the magnetic field can be written as the gradient of a scalar magnetic potential

$$\vec{B} = \nabla \phi .$$

The potential $\phi$ fulfills the Laplace equation

$$\nabla^2 \phi = 0 .$$
Making the ansatz
\[ \phi(y, z) = f(y) \cos(k_u z) \implies \frac{d^2 f}{dy^2} - k_u^2 f = 0 \]
we get for the general solution
\[ f(y) = c_1 \sinh(k_u y) + c_2 \cosh(k_u y) , \]
\[ B_y(y, z) = \frac{\partial \phi}{\partial y} = k_u (c_1 \cosh(k_u y) + c_2 \sinh(k_u y)) \cos(k_u z) . \]
The vertical field component \( B_y \) is symmetric with respect to the plane \( y = 0 \) hence \( c_2 = 0 \), and moreover \( k_u c_1 = B_0 \). So the potential is
\[ \phi(x, y, z) = \frac{B_0}{k_u} \sinh(k_u y) \cos(k_u z) . \] (4)
For \( y \neq 0 \) the magnetic field has also a longitudinal component \( B_z \).

\[
\begin{align*}
B_x &= 0 \\
B_y &= B_0 \cosh(k_u y) \cos(k_u z) \\
B_z &= -B_0 \sinh(k_u y) \sin(k_u z) .
\end{align*}
\]

In the following we restrict ourselves to the symmetry plane \( y = 0 \).

### 2.2 Electron Motion in an Undulator

#### 2.2.1 Trajectory in first order

We call \( W = E_{\text{kin}} + m_e c^2 = \gamma m_e c^2 \) the total relativistic energy of the electron. The transverse acceleration by the Lorentz force is

\[
\gamma m_e \dot{v} = -e \vec{v} \times \vec{B} .
\]

This results in two coupled equations

\[
\ddot{x} = \frac{e}{\gamma m_e} B_y \dot{z} \quad \quad \ddot{z} = -\frac{e}{\gamma m_e} B_0 \dot{x}
\]

which are solved iteratively. To obtain the first-order solution we observe that \( v_z = \dot{z} \approx v = \beta c = \text{const} \) and \( v_x \ll v_z \). Then \( \ddot{z} \approx 0 \) and

\[
x(t) \approx -\frac{eB_0}{\gamma m_e \beta c k_u^2} \cos(k_u \beta ct) , \quad z(t) \approx \beta ct .
\]

The electron travels on a cosine-like trajectory

\[
x(z) = -A \cos(k_u z) \quad \text{with} \quad A = \frac{eB_0}{\gamma m_e \beta c k_u^2} .
\]

The maximum divergence angle is

\[
\theta_{\text{max}} \approx \left[ \frac{dx}{dz} \right]_{\text{max}} = \frac{eB_0}{\gamma m_e \beta c k_u} = \frac{K}{\beta \gamma}
\]

Here we have introduced the undulator parameter

\[
K = \frac{eB_0}{m_e c k_u} = \frac{eB_0 \lambda_u}{2\pi m_e c} .
\]

Synchrotron radiation of relativistic electrons is emitted inside a cone with opening angle \( 1/\gamma \). If the particle trajectory stays within this cone one speaks of an undulator magnet (Fig. 6):

**Undulator:** \( \theta_{\text{max}} \leq 1/\gamma \quad \Rightarrow \quad K \leq 1 \).

If the trajectory extends beyond the cone the magnet is called a wiggler:

**Wiggler:** \( K > 1 \).

The special feature of an undulator is that the radiated field of an electron interferes with itself along the magnet axis. The consequence is, as we shall see, that the radiation is nearly monochromatic.

#### 2.2.2 Motion in second order

Due to the cosine-trajectory the \( z \) component of the velocity is not constant. It is given by

\[
v_z = \sqrt{v^2 - v_x^2} \approx c \left( 1 - \frac{1}{2\gamma^2} \left( 1 + \gamma^2 v_x^2/c^2 \right) \right) .
\]
We insert for $v_z = \dot{z}(t)$ the first-order solution, then the average z velocity is

$$\bar{v}_z = c \left(1 - \frac{1}{2\gamma^2} (1 + K^2/2)\right) \equiv \bar{\beta}c$$

(10)

It should be noted that the z velocity oscillates about the average\(^2\)

$$\ddot{z}(t) = \bar{\beta}c + \frac{cK^2}{4\gamma^2} \cos(2\omega_ut) \quad \text{with} \quad \omega_u = \bar{\beta}cK_u$$

The trajectory in second order reads

$$x(t) = -\frac{cK}{\gamma\omega_u} \cos(\omega_ut) \quad , \quad z(t) = \bar{\beta}ct + \frac{cK^2}{8\gamma^2\omega_u} \sin(2\omega_ut).$$

(11)

### 2.2.3 Lorentz transformation into a moving coordinate system

Consider a coordinate system $(x^*, y^*, z^*)$ moving with the average z velocity of the electrons:

$$\bar{v}_z = \bar{\beta}c, \quad \bar{\gamma} \approx \gamma = W/(m_ec^2).$$

The Lorentz transformation reads

$$t^* = \bar{\gamma}(t - \bar{\beta}z/c) = \bar{\gamma}t(1 - \bar{\beta}^2) \approx t/\gamma$$

$$x^* = x = -\frac{cK}{\gamma\omega_u} \cos(\omega_ut)$$

$$z^* = \bar{\gamma}(z - \bar{\beta}ct) \approx \frac{cK^2}{8\gamma\omega_u} \sin(2\omega_ut)$$

The electron orbit in the moving system is:

$$x^*(t^*) = -\frac{cK}{\gamma\omega_u} \cos(\omega^*t^*) \quad , \quad z^*(t^*) = \frac{cK^2}{8\gamma\omega_u} \sin(2\omega^*t^*)$$

(12)

with $\omega^* = \gamma\omega_u$. Note that $\omega_ut = \omega^*t^*$. This is mainly a transverse harmonic oscillation with the frequency $\omega^* = \gamma\omega_u$. Superimposed is a small longitudinal oscillation which will be ignored here, it leads to higher harmonics in the radiation. The motion is plotted in Fig. 7. In the moving system the electron emits dipole radiation with the frequency $\omega^* = \gamma\omega_u$ and the wavelength $\lambda_u^* = \lambda_u/\gamma$.\(^2\)

---

\(^2\) This oscillation leads to odd higher harmonics of the undulator radiation, see for example the book by K. Wille. In a helical undulator the z velocity is constant.
2.2.4 Transformation of radiation into laboratory system

The radiation characteristics of an oscillating dipole moving at relativistic speed is depicted in Fig. 8. Within increasing Lorentz factor $\gamma$ the radiation becomes more and more concentrated in the forward direction. We are interested in the light wavelength in the laboratory system as a function of the angle $\theta$ with respect to the beam axis. The Lorentz transformation of the photon energy reads

$$\bar{\omega} = \gamma \omega \left(1 - \beta \cos \theta \right)$$

$$\Rightarrow \quad \lambda_{\ell} = \frac{2\pi c}{\omega_{\ell}} = \frac{2\pi c\bar{\gamma}}{\omega^*} \left(1 - \beta \cos \theta \right) = \lambda_u \left(1 - \beta \cos \theta \right)$$

Using $\bar{\gamma} \approx \gamma$, $\beta = \left(1 - \frac{1}{2\gamma^2} \left(1 + K^2/2 \right) \right)$ and $\cos \theta \approx 1 - \theta^2/2$ we obtain for the wavelength of undulator radiation

$$\lambda_{\ell} = \frac{\lambda_u}{2\gamma^2} \left(1 + K^2/2 + \gamma^2 \theta^2 \right). \quad (13)$$

2.3 Line shape of undulator radiation

An electron passing an undulator with $N_u$ periods produces a wavetrain with $N_u$ oscillations (Fig. 9). The electric field of the light wave is written as

$$E_{\ell}(t) = \begin{cases} 
E_0 e^{i\omega_{\ell}t} & \text{if } -T/2 < t < T/2 \\
0 & \text{otherwise}
\end{cases}$$

The time duration of the wave train is $T = N_u \lambda_{\ell}/c$. Due to its finite length, this wave train is not
monochromatic but contains a frequency spectrum which is obtained by Fourier transformation

\[ A(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} E_\ell(t)e^{-i\omega t} dt = \frac{E_0}{\sqrt{2\pi}} \int_{-T/2}^{+T/2} e^{i(\omega_\ell - \omega)t} dt \]

\[ = \frac{2E_0}{\sqrt{2\pi}} \frac{\sin(\Delta \omega T/2)}{\Delta \omega} \] with \( \Delta \omega = \omega - \omega_\ell \).

The spectral intensity is

\[ I(\omega) \propto \left( \frac{\sin \xi}{\xi} \right)^2 \text{ with } \xi = \Delta \omega T/2 = \frac{\pi N_u(\omega - \omega_\ell)}{\omega_\ell}. \]

It has a maximum at \( \omega = \omega_\ell \) and a width proportional to \( 1/N_u \). The line shape for a wave train with 100 oscillations is shown in Fig. 10. The spectral resolution is

\[ \Delta \lambda/\lambda = 1/N_u \]

and amounts to 1% in the present example.

3 Low-gain FEL

3.1 Energy transfer from electron beam to light wave

We consider the case of “seeding”, where the initial light wave with wavelength \( \lambda_\ell \) is provided by an external source such as an optical laser. The schematic setup of a low-gain FEL is shown in Fig. 11.

The light wave is co-propagating with the relativistic electron beam and is described by a plane electromagnetic wave

\[ E_\ell(z, t) = E_0 \cos(k_\ell z - \omega_\ell t + \psi_0) \] with \( k_\ell = \omega_\ell/c = 2\pi/\lambda_\ell \).
Obviously the light wave, travelling with speed $c$ along the $z$ axis, slips with respect to the electrons whose average speed in $z$ direction is

$$
\bar{v}_z = c \left( 1 - \frac{1}{2\gamma^2} (1 + K^2/2) \right) < c .
$$

The question is then: how can there be a continuous energy transfer from the electron beam to the light wave? The electron energy $W = \gamma m_e c^2$ changes in the time interval $dt$ by

$$
dW = \vec{v} \cdot \vec{F} dt = -ev_x(t)E_x(t)dt .
$$

The $x$ component of the electron velocity $v_x$ and the electric vector $E_x$ of the light wave must point in the same direction to get an energy transfer from the electron to the light wave. To determine the condition for energy transfer along the entire trajectory we compute the electron and light travel times for a half period of the undulator:

$$
t_{el} = \lambda_u/(2\bar{v}_z), \quad t_{light} = \lambda_u/(2c) .
$$

Figure 12 illustrates that after a half period $v_x$ and $E_x$ are still parallel if the phase of the light wave has slipped by $\pi$, i.e.

$$
\omega_{\ell}(t_{el} - t_{light}) = \pi .
$$

(Remark: also $3\pi, 5\pi \ldots$ are possible, leading to higher harmonics of the radiation). This condition allows to compute the light wavelength:

$$
\lambda_{\ell} = \frac{\lambda_u}{2\gamma^2} \left( 1 + \frac{K^2}{2} \right)
$$

which is identical with the wavelength of undulator radiation (in forward direction).

### 3.1.1 Quantitative treatment

The energy transfer from an electron to the light wave is

$$
\frac{dW}{dt} = -ev_x(t)E_x(t)
$$

$$
= -e \frac{cK}{\gamma} \sin(k_u z) E_0 \cos(k_{\ell} z - \omega_{\ell} t + \psi_0)
$$
We consider the first term. The argument of the sine function is called the \textit{ponderomotive phase}:

\[
\psi \equiv (k_\ell + k_u)z - \omega_\ell t + \psi_0 = (k_\ell + k_u)\bar{\beta}c t - \omega_\ell t + \psi_0.
\] (14)

There will be a continuous energy transfer from the electron to the light wave if \(\psi\) is constant (independent of time) and in the range \(0 < \psi < \pi\), the optimum value being \(\psi = \psi_0 = \pi/2\). The condition \(\psi = \text{const}\) can be fulfilled only for a certain wavelength.

\[
\psi = \text{const} \iff \frac{d\psi}{dt} = (k_\ell + k_u)\bar{v}_z - k_\ell c = 0.
\] (15)

Insertion of \(\bar{v}_z\) permits to compute the light wavelength:

\[
\lambda_\ell = \frac{\lambda_u}{2\gamma^2} \left( 1 + \frac{K^2}{2} \right).
\] (16)

The condition for resonant energy transfer all along the undulator therefore yields exactly the same light wavelength as is observed in undulator radiation at \(\theta = 0\). This is the reason why the spontaneous undulator radiation can serve as a “seed radiation” in the SASE FEL.

Now we look at the second term. Here the phase of the sine function cannot be kept constant since from

\[(k_\ell - k_u)\bar{\beta}c t - \omega_\ell t + \psi_0 = \text{const}\]

we would get

\[k_\ell(1 - \bar{\beta}) = -k_u\bar{\beta} \Rightarrow k_\ell = 2\pi/\lambda_\ell < 0\]

which of course is unphysical. Hence the second sine function oscillates rapidly and averages to zero.

### 3.2 The pendulum equation

We assume “seeding” by an external light source with wavelength \(\lambda_\ell\). The \textit{resonant energy} \(W_r = \gamma_\ell m_e c^2\) is defined by the equation

\[
\lambda_\ell = \frac{\lambda_u}{2\gamma^2} \left( 1 + \frac{K^2}{2} \right).
\] (17)

Let the electron gamma factor be slightly larger, \(\gamma > \gamma_r\), and call \(\eta = (\gamma - \gamma_r)/\gamma_r\) the relative energy deviation. We assume

\[0 < \eta = \frac{\gamma - \gamma_r}{\gamma_r} \ll 1\].
The energy deviation $\eta \gamma$, $m_e c^2$ and the ponderomotive phase $\psi$ will both change due to the interaction with the radiation field. The low-gain FEL is defined by the condition that the electric field amplitude grows slowly such that $E_0 \approx \text{const}$ during one passage of the undulator.

The time derivative of the ponderomotive phase is no longer zero for $\gamma > \gamma_r$:

$$\dot{\psi} = k_u c - k_{c} \frac{1 + K^2/2}{2 \gamma^2}.$$  

We subtract $0 = k_u c - k_{c} (1 + K^2/2)/(2 \gamma^2)$, see Eq. (16), and get

$$\frac{d\psi}{dt} = \frac{k_{c} c}{2} \left( 1 + \frac{K^2}{2} \right) \left( \frac{1}{\gamma^2_r} - \frac{1}{\gamma^2} \right).$$

From this follows

$$\frac{d\psi}{dt} \approx 2k_u c \eta \equiv \omega' \quad (\omega' \ll k_u c). \quad (18)$$

The time derivative of $\eta$ is

$$\frac{d\eta}{dt} = -\frac{eE_0 K}{2m_e c \gamma^2_r} \sin \psi. \quad (19)$$

Combining Eqs. (18) and (19) we arrive at the so-called “Pendulum Equation” of the low-gain FEL

$$\ddot{\psi} + \Omega^2 \sin \psi = 0 \quad \text{with} \quad \Omega^2 = \frac{eE_0 K k_u}{m_e \gamma^2_r}.$$  

### 3.3 Phase space representation

There is a complete analogy with the motion of a mathematical pendulum (Fig. 13). At small amplitude we get a harmonic oscillation. With increasing angular momentum the motion becomes unharmonic. At very large angular momentum one gets a rotation (unbounded motion). The phase space trajectory for an electron in a FEL can be easily constructed by writing the coupled differential equations (18) and (19) as difference equations and solving these in small time steps. The trajectories for 20 electrons of different initial phases $\psi_0$ are shown in Fig. 14 for $\gamma = \gamma_r$ and $\gamma > \gamma_r$. In the first case the net energy transfer is zero since there are as many electrons which supply energy to the light wave as there which remove energy from the wave. For $\gamma > \gamma_r$, however, the phase space picture clearly shows that there is a net energy transfer from the electron beam to the light wave. This will be computed in the next section.
3.4 Computation of the FEL gain, Madey theorem

The energy (per unit volume) of the light wave is

$$ W_{\text{light}} = \frac{\varepsilon_0}{2} E_0^2 . $$

The energy increase and relative gain caused by one electron are

$$ \Delta W_{\text{light}} = -m_e c^2 \gamma_r \Delta \eta \quad G_1 = \frac{\Delta W_{\text{light}}}{W_{\text{light}}} = -\frac{2m_e c^2 \gamma_r}{\varepsilon_0 E_0^2} \Delta \eta . $$

Here $\Delta \eta$ is the change of the relative energy deviation of the electron upon passing the undulator. We use Eq. (18) to compute this change:

$$ \Delta \eta = \frac{\Delta \dot{\psi}}{2k_u c} . $$

Summing over all electrons in the bunch ($n_e$ per unit volume) the total gain becomes

$$ G = -\frac{m_e c^2 n_e}{\varepsilon_0 E_0^2 k_u} \langle \Delta \dot{\psi} \rangle . $$

Here $\langle \Delta \dot{\psi} \rangle$ denotes the change of the time derivative of the ponderomotive phase, averaged over all electrons. Hence it is this quantity we have to compute.

3.4.1 Phase change in undulator

We multiply the pendulum equation $\ddot{\psi} + \Omega^2 \sin \psi = 0$ with $2\dot{\psi}$ and integrate over time

$$ \dot{\psi}^2 - 2\Omega^2 \cos \psi = \text{const} \quad \Rightarrow \quad \dot{\psi}(t)^2 = \dot{\psi}_0^2 + 2\Omega^2 [\cos \psi(t) - \cos \psi_0] . $$

From Eq. (18)

$$ \dot{\psi}_0 = \dot{\psi}(0) = 2c k_u \eta \equiv \omega' $$

we obtain then

$$ \dot{\psi}(t) = \omega' \sqrt{1 + 2(\Omega/\omega')^2 [\cos \psi(t) - \cos \psi_0]} . $$

Fig. 14: Phase space trajectories for 20 electrons of different initial phases $\psi_0$. Left picture: $\gamma = \gamma_r$. The electrons with $\psi_0 < 0$ withdraw energy from the light wave while those with $\psi_0 > 0$ supply energy to the light wave. Obviously the net energy transfer is zero for $\gamma = \gamma_r$. Right picture: $\gamma > \gamma_r$. One can easily see that the net energy transfer is positive.
For a weak laser field one finds $(\Omega/\omega')^2 \ll 1$, so we expand the square root up to second order
\[
\sqrt{1+x} = 1 + x/2 - x^2/8 \ldots
\]
and get
\[
\dot{\psi}(t) = \omega' + \frac{\Omega^2}{\omega'} [\cos \psi(t) - \cos \psi_0] - \frac{\Omega^4}{2\omega'^3} [\cos \psi(t) - \cos \psi_0]^2. \tag{23}
\]
This equation is solved iteratively.

**Zeroth order:** $\dot{\psi}_0 = \omega'$, $\Delta \dot{\psi}_0 = 0$.

**First order:** the phase $\psi(t)$ in first order is obtained by integrating $\dot{\psi}_0$:
\[
\psi_1(t) = \psi_0 + \dot{\psi}_0 \cdot t = \psi_0 + \omega' t.
\]
We insert this in Eq. (23) to get $\psi$ in first order
\[
\dot{\psi}_1(t) = \omega' + (\Omega^2/\omega')[\cos(\psi_0 + \omega' t) - \cos \psi_0]. \tag{24}
\]
The flight time through the undulator is $T$, so the change of $\dot{\psi}_1$ when the electron passes the undulator is
\[
\Delta \psi_1 = (\Omega^2/\omega')[\cos(\psi_0 + \omega' T) - \cos \psi_0].
\]
According to Eq. (21) the gain is obtained by averaging $\Delta \dot{\psi}$ over all particles in the bunch which means that one has to average over all initial phases $\psi_0$. The result is
\[
< \Delta \dot{\psi}_1 > = 0. \tag{25}
\]
*The FEL gain is zero in first order.* The physical reason is the nearly symmetric initial phase space distribution.

**Second order:** Equation (24) is integrated to get $\psi$ in second order:
\[
\psi_2(t) = \psi_0 + \omega' \cdot t + (\Omega/\omega')^2 [\sin(\psi_0 + \omega' t) - \sin \psi_0 - \omega' t \cos \psi_0]
\]
\[
\psi_1(t)
\]
\[
\delta \psi_2(t)
\]
This is inserted in Eq. (23) to compute $\dot{\psi}$ at $t = T$ in second order
\[
\dot{\psi}_2(T) = \omega' + (\Omega^2/\omega')[\cos(\psi_0 + \omega' T + \delta \psi_2) - \cos \psi_0]
\]
\[
- \frac{\Omega^4}{(2\omega'^3)} [\cos(\psi_0 + \omega' T + \delta \psi_2) - \cos \psi_0]^2 \tag{27}
\]
$\delta \psi_2 \ll 1 \Rightarrow \cos(\psi_0 + \omega' T + \delta \psi_2) \approx \cos(\psi_0 + \omega' T) - \delta \psi_2 \sin(\psi_0 + \omega' T)$
\[
\cos(\psi_0 + \omega' T + \delta \psi_2) \approx \cos(\psi_0 + \omega' T)
\]
\[
- (\Omega/\omega')^2 \sin(\psi_0 + \omega' T) [\sin(\psi_0 + \omega' T) - \sin \psi_0 - \omega' T \cos \psi_0]
\]
Averaging over all start phases $\psi_0$ yields
\[
< \cos(\psi_0 + \omega' T + \delta \psi_2) > = (1/2)(1 - \cos(\omega' T) - \omega' T \sin(\omega' T)).
\]
From this we get
\[
< \Delta \dot{\psi}_2 > = -(\Omega^4/\omega'^3)[1 - \cos(\omega' T) - (\omega' T/2) \sin(\omega' T)].
\]
Remembering that $T = N_u \lambda_u / c$ is the flight time through the undulator and $\xi = \omega T / 2$ one obtains
\[
< \Delta \dot{\psi}_2 > = -\frac{\Omega^4}{\omega^3} [1 - \cos(2\xi) - \xi \sin(2\xi)]
\]
\[
= \frac{N_u^3 \lambda_u^3 \Omega^4}{8 e^3} \cdot \frac{d}{d\xi} \left( \frac{\sin \xi}{\xi} \right)^2
\]
The FEL gain function (21) is hence
\[
G(\xi) = -\frac{\pi e^2 K^2 N_u^3 \lambda_u^2 n_e}{4 \varepsilon_0 m_e c^2 \gamma^3} \cdot \frac{d}{d\xi} \left( \frac{\sin^2 \xi}{\xi^2} \right)
\]
(28)
We have proven the Madey Theorem which states that the FEL gain curve is given by the negative derivative of the line-shape curve of undulator radiation. This is shown in Fig. 15.

4 High-gain FEL
4.1 General principle
The essential feature of the high-gain FEL is that a large number of electrons radiate coherently. In that case, the intensity of the radiation field grows quadratically with the number of particles: $I_N = N^2 I_1$. If it were possible to concentrate all electrons of a bunch into a region which is far smaller than wavelength of the radiation then these $N$ particles would radiate like a “point macroparticle” with charge $Q = -N e$, see Fig. 16. The big problem is, however, that this concentration of some $10^9$ electrons into a tiny volume is totally unfeasible, rather even the shortest particle bunches are much longer than the FEL wavelength. The way out of this dilemma is given by the process of micro-bunching which is based on the following principle: those electrons which lose energy to the light wave travel on a cosine trajectory of larger amplitude than the electrons which gain energy from the light wave. The result is a modulation of the longitudinal velocity which eventually leads to a concentration of the electrons in slices which are shorter than $\lambda_\ell$. The result of a numerical simulation of this process is shown in Fig. 17. The particles within a micro-bunch radiate coherently. The resulting strong radiation field enhances the micro-bunching even further. The result is a “collective instability”, leading to an exponential growth of the radiation power. The ultimate power is $P \propto N_e^2$ where $N_e$ is the number of particles in a coherence region. A typical value is
\[
N_e \approx 10^6 \quad \Rightarrow \quad P_{\text{FEL}} = 10^6 P_{\text{undulator}}.
\]
4.2 Approximate analytical treatment and experimental results

An approximate analytic description of the high-gain FEL requires the self-consistent solution of the coupled pendulum equations and the inhomogeneous wave equation for the electromagnetic field of the light wave. In the one-dimensional FEL theory the dependencies on the transverse coordinates $x, y$ are disregarded. The wave equation for the radiation field $E_x$ reads

$$
\frac{\partial^2 E_x}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2 E_x}{\partial t^2} = \mu \frac{\partial j_x}{\partial t}
$$

where the current density $\vec{j}$ is generated by the electron bunch moving on its cosine-like trajectory. In addition, one has to consider the longitudinal space charge field $E_z$ which is generated by the gradually evolving periodic charge density modulation. After a lot of tedious mathematical steps and several simplifying assumptions one arrives at a third-order differential equation for the amplitude of the electric field of the light wave:

$$
\frac{d^3 \tilde{E}_x}{dz^3} - 4ik_n \eta \frac{d^2 \tilde{E}_x}{dz^2} + (k_p^2 - 4k_n^2 \eta^2) \frac{d \tilde{E}_x}{dz} - i\eta^3 \tilde{E}_x(z) = 0 .
$$

(29)
Here we have introduced the gain parameter $\Gamma$ and a parameter $k_p$

$$\Gamma = \left( \frac{\mu_0 K^2 e^2 n_e}{4\gamma^3 m_e} \right)^{\frac{1}{3}}, \quad k_p = \sqrt{\frac{4\gamma^2 e^{\Gamma^3}}{\omega K^2}}$$

and assumed that the electron beam has negligible energy spread.

This third-order differential equation can be solved analytically. For the case $\gamma = \gamma_r$ one obtains

$$\tilde{E}_x(z) = A_1 \exp(-i\Gamma z) + A_2 \exp \left( \frac{i + \sqrt{3}}{2} \Gamma z \right) + A_3 \exp \left( \frac{i - \sqrt{3}}{2} \Gamma z \right).$$

The second term exhibits exponential growth as a function of the position $z$ in the undulator. The electric field grows exponentially as $\exp(\frac{\sqrt{3}}{2} \Gamma z)$, the power grows as $\exp(\sqrt{3}\Gamma z)$. The gain parameter $\Gamma$ is related to two parameters which are in widespread use: the Pierce parameter and the power gain length

$$\rho_{pierce} = \frac{\lambda_0 \Gamma}{4\pi} \quad L_g = \frac{1}{\sqrt{3}\Gamma}.$$  \hspace{1cm} (32)

The above calculations, which have been sketched only very briefly, indicate that there is an onset of an “instability”, leading to a progressing microbunching and an exponential increase in radiation power along the undulator. A quantitative treatment requires elaborate numerical simulations (see Fig. 17). Microbunching has been experimentally observed at the 60 $\mu$m FEL Firefly at Stanford University, see Fig. 18. The exponential growth of radiation power and the progressing microbunching in a long undulator are depicted in Fig. 19. One characteristic but quite undesirable feature of a SASE FEL is the presence of fluctuations which are due to the stochastic nature of the initiating undulator radiation. The light pulse energy fluctuates from pulse to pulse, and the same applies for the wavelength and the time structure. As an example I show in Fig. 20 the measured wavelength distribution of several FEL pulses together with the average over 100 pulses. The origin of the large fluctuations is that small statistical fluctuations in the incoming undulator radiation are strongly amplified in the exponential growth region (see Fig. 19). There are so-called “seeding schemes” under development where radiation of the desired wavelength is produced in a short undulator, passed through a monochromator and then used as seed radiation for the FEL process. Here one can expect much higher monochromaticity of the final FEL radiation. In spite of the fluctuations the SASE radiation is highly coherent as demonstrated by the double-slit interference patterns in Fig. 21.

**Bibliography**

K.Wille, The Physics of Particle Accelerators, Oxford University Press 2001


Fig. 18: Observation of microbunching at the 60 $\mu$m FEL Firefly.
Fig. 19: The exponential growth of radiation power as a function of undulator length. The data at \( \lambda = 100 \) nm have been obtained at the SASE FEL of the TESLA Test Facility at DESY. The progressing microbunching is indicated schematically.

Fig. 20: The measured spectra of three pulses in the SASE FEL at DESY. Also shown is the average spectrum of 100 pulses.

Fig. 21: Measured double-slit diffraction patterns at 100 nm FEL wavelength. The separation of the horizontal slits is 0.5 resp. 1 mm.
ELECTRON AND ION SOURCES FOR PARTICLE ACCELERATORS

R. Scrivens
CERN, Geneva, Switzerland

Abstract
A brief introduction to electron and ion sources for particle accelerators is given. Concentrating on the basic processes for the production of these particles, thermionic and photocathode production of electrons and the processes in Penning and electron cyclotron resonance ion sources are covered. Finally, negative ions are discussed.

1 INTRODUCTION
Particle sources are an essential, integral part of an accelerator chain. Although the accelerator designer might like to make a parameter list for a particle source, it is in many cases the availability of the source that defines the parameters and scheme of acceleration.

In this paper, we discuss some of the fundamental physics of electron and ion sources. For electron sources we treat the thermionic and photoelectric modes of production. For ion sources we show briefly the processes inside a Penning and Electron Cyclotron Resonance (ECR) ion source, before discussing the production of negative ions. For the plethora of other ion sources, several hand books describing their modes of operation are available.

2 ELECTRON SOURCES

2.1 Thermionic electron emitters
The energy distribution of electrons inside a solid metal is governed by Fermi–Dirac distribution, which allows only one conduction electron to occupy a given wave function and spin polarity inside the metal. The electron density \( n(E) \) in a small energy range \( dE \) is given by

\[
n(E) dE = \left[ \frac{4 \pi (2m_e)^{3/2}}{h^3} \right] \sqrt{E} \left[ 1 + \exp \left( \frac{E - E_{\text{Fermi}}}{kT} \right) \right]^{-1} dE
\]

where \( m_e \) is the electron mass, \( h \) and \( k \) are the Planck and Boltzmann constants and \( T \) is the temperature. \( E_{\text{Fermi}} \) is then the highest energy level filled for \( T = 0 \) K. As the temperature of the metal is increased, the electrons move to higher energy states, where some are above the metal’ work-function. Approximately integrating the density of electrons above the work-function yields the Richardson–Dushman equation, which gives the current density of electrons available to be extracted from the metal, i.e.

\[
J = A T^2 \exp \left( \frac{-e \phi_{\text{work}}}{kT} \right);
\]

where

\[
A = \frac{4 \pi e m_e k^2}{h^3} \approx 1.2 \times 10^6 \text{ A m}^{-2} \text{ K}^{-2}.
\]
Hence the current available is a strong function of the temperature of the material, with the highest emission from a material with a low work-function. In addition, the factor $A$ is usually below the theoretical value, and is fixed for a material but is not a function of temperature.

The parameters of several common cathode materials are given in Table 1, with a plot of electron emission versus temperature in Fig. 1.

Table 1: Parameters of sample cathode materials [1]

<table>
<thead>
<tr>
<th></th>
<th>$A$ (A cm$^{-2}$ K$^{-2}$)</th>
<th>$\phi_{\text{work}}$ (eV)</th>
<th></th>
<th>$A$ (A cm$^{-2}$ K$^{-2}$)</th>
<th>$\phi_{\text{work}}$ (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>60</td>
<td>4.54</td>
<td>Mixed oxide†</td>
<td>0.01</td>
<td>~1</td>
</tr>
<tr>
<td>Ta</td>
<td>60</td>
<td>4.12</td>
<td>Cs/O/W†</td>
<td>0.003</td>
<td>0.72</td>
</tr>
<tr>
<td>Thoriated W</td>
<td>3</td>
<td>2.63</td>
<td>LaB$_6$</td>
<td>29</td>
<td>2.66</td>
</tr>
<tr>
<td>Cesium</td>
<td>160</td>
<td>1.81</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† Parameters for mixed oxide (SrO/BaO) and CsO on W substrate depend critically on the mixture and activation. Typical values are given.

![Graph showing electron emission as a function of temperature for selected cathode material. NB the Cs curve is only for indication, as the melting point is 300 K.](image)

**Fig. 1:** Electron emission as a function of temperature for selected cathode material. NB the Cs curve is only for indication, as the melting point is 300 K.

From this data we see that the best cathode materials can either be heated to high temperatures (tungsten and tantalum), or have a low work-function. Cesium, with the lowest work-function for a solid element, has a low melting point at 300 K that limits the current available in the solid phase.

The use of surface coatings to lower the work-function of a high-temperature material produces some of the best cathodes. Thoriated tungsten and various oxides of strontium and barium [2] can be used as coatings on W, Ta, or Mo substrates. However, the surface coatings may be eroded by ion bombardment (from ionized residual gas or plasma), limiting the cathode lifetime. Further loss of surface material arises from evaporation at high temperatures.

Lanthanum hexaboride (LaB$_6$), a solid material with a low work-function, remains stable up to 1700 K. LaB$_6$ is reasonably resistant to ion bombardment. Evaporation of lanthanum from the boride frame is slow, even with electron emission rates of 1 A/cm$^2$, solid LaB$_6$ cathodes can have lifetimes of more than one year [3].
2.2 Photocathodes

When a photon with an energy higher than the work-function for a material impinges on the material’s surface, it can cause an electron to be emitted. The maximum photon wavelength that can cause emission can be written in units as $\lambda = \frac{1240}{\phi}$, where $\phi$ is the work-function and $\lambda$ is in nanometres. Excess energy from the photon is carried in the electron’s kinetic energy.

The ratio of the number of electrons emitted per incident photon is called the Quantum Efficiency ($Q_e$), which is a function of the photon energy. For metals, the minimum-energy photons are typically in the ultraviolet energy range [$\lambda$(Cu) $< 267$ nm, $\lambda$(W) $< 275$ nm]. As metals are good reflectors in this region, the probability of the photon ejecting an electron is low, leading to a poor quantum efficiency. However, very low currents of electrons can be produced with photon energies lower than the work-function, by using the thermal tail of electrons between the Fermi energy and the work-function.

As semiconductors have lower conductivity than metals, the photon may penetrate further into the material, and the resulting quantum efficiency is higher. The $Q_e$ values for several semiconductors are listed in Table 2. Some materials [e.g. Bialkali (K$_2$CsSb)] have very high efficiency. However, the Bialkali surface is very unstable and it has been impossible to make operational electron sources with this material. Extensive research has led to Cs$_2$Te becoming the most common semiconductor photocathode material used, for electron guns [4], because of its acceptable stability in real source conditions (where bombardment by positive ions and X-rays is a concern).

Photocathodes now employ fast laser pulses to produce complicated bunch-trains. Picosecond bunches can be formed at the RF frequency repetition rate, considerably easing the following electron-bunching structure system, and making the first acceleration of the electrons by radio-frequency cavities possible.

<table>
<thead>
<tr>
<th>Material</th>
<th>$Q_e$ (%)</th>
<th>$\lambda$ (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu</td>
<td>0.01</td>
<td>267</td>
</tr>
<tr>
<td>K$_2$CsSb</td>
<td>29</td>
<td>590</td>
</tr>
<tr>
<td>Cs$_2$Te</td>
<td>12.4</td>
<td>~350</td>
</tr>
<tr>
<td>GaAs:Cs</td>
<td>17</td>
<td>225</td>
</tr>
</tbody>
</table>

2.3 Child–Langmuir current limit

Electrons are normally emitted from surfaces (by either thermionic or photo-emission) with low energy, and accelerated away from the cathode by an electric field produced by the potential difference between the cathode and anode.

The electric field component produced by the resulting space charge of these electrons reduces the field at the surface of the cathode. If this field falls to zero, no further current can be pulled from the cathode surface. This current density at which the electric field goes to zero is called the Child–Langmuir current limit [5].

It is calculated for a parallel-plate cathode and anode system (that are parallel and extend to infinity) by integrating the one-dimension Poisson equation, with the charge density given a fixed emission current, using the boundary voltages at the cathode ($V = 0$) and anode ($V = V$)

$$\frac{d^2V}{dx^2} = -\frac{\rho}{\varepsilon_0}; \quad J = \rho V; \quad qV = \frac{1}{2}mv^2; \quad V(x = 0) = 0; \quad V(x = d) = V; \quad \frac{dV(x = 0)}{dx} = 0$$
where $\rho(x)$ is the charge density, $d$ is the anode to cathode gap, and $v(x)$, $m$, and $q$ the particle velocity, mass, and charge, respectively. The final boundary condition $[dv(0)/dx = 0]$ allows the electric field at the cathode surface to tend towards zero. This yields the Child–Langmuir current limit ($J_{C-L}$)

$$J_{C-L} = \frac{4}{9} e \left( \frac{2q}{m} \right)^{1/2} \frac{V^{3/2}}{d^2}.$$ 

This limit is only valid for parallel-plate extraction systems. However, the current density obtained is usually a good guide to the implementation limit of most electron guns.

If the voltage applied between an electron-emitting cathode and an extracting anode is insufficient to extract the thermal (or photo-induced) current density available, the electron gun is extraction-limited. Holding the cathode and anode at an equal potential results in no current being drawn from the cathode. By this method, a biased grid that is downstream from the cathode (called a cathode grid) can be used to modulate the electron current, without requiring that the cathode temperature be modified. Fast modulation of the cathode-grid potential can result in an electron-pulse rise time of the order of a few nanoseconds, even for thermionic electron guns.

The Child–Langmuir current density limit is also valid for positive-ion extraction. However, the distance $d$ is not well defined in this case, as the plasma can move in relation to the anode and cathode surfaces. The factor $m^{-1/2}$ results in the current density limit for proton extraction being 45 times lower than for electrons, for a given voltage and distance.

### 3 ION SOURCES

In order to produce an ion beam for an accelerator, ion sources require an ion production region, and an ion extraction system. Ion production normally takes place in a plasma, although ions may be sputtered or desorbed from a surface in some source types.

The plasma conditions and resulting ion production are affected by many different processes, including:

- electron heating
- plasma confinement (electric and magnetic)
- collisions (electron–electron, electron–ion, ion–electron, ion–ion + residual gas)
- atomic processes (ionization, excitation, disassociation, recombination)
- surface physics (coatings + desorption, electron emission).

In addition, the source engineer must take into account thermal effects on the source anodes, cathodes and chambers, as well as considering erosion of material inside the source. Furthermore, a reliable and stable ion production is sought with a small beam emittance that often minimizes power consumption to the high-voltage platform of the source.

Ultimately, the plasma properties must be tuned to maximize the production of the ion of interest (while suppressing other unwanted ions).
3.1 Penning or Philips Ionization Gauge (PIG) ion source

The Penning ion source was popular as an internal source inside cyclotrons in the 1940s, and is based on the Penning discharge, first studied by Maxwell. A schematic diagram of a Penning source is shown in Fig. 2.

The source is based on a cylinder anode, with a biased cathode at each cylinder end (in the hot-cathode version, one cathode is heated and the other is termed an anticathode or reflector). A magnetic field of at least 0.1 T runs parallel to the cylinder axis. During the discharge process, electrons are emitted from the cathode and accelerated into the gas, where, through ionization of the gas atoms and molecules, a plasma is formed. The electrons cannot easily reach the anode because of the magnetic field, and a high ratio will not strike the anticathode. These electrons are then left to oscillate between the two cathodes and can cause further ionization.

The gas in the source is kept at a pressure in the range of $10^{-3}$ mbar to 1 mbar, so that the mean free electron path is of the order of the source dimension. If the pressure is too high, the electrons lose energy by multiple collisions and will not reach an energy level high enough to cause ionization of gas atoms. If the pressure is too low, the electrons will oscillate without colliding with the rest gas, until the high electron space charge forces them to the anode wall.

The ions in the plasma can diffuse to the anode wall, where a slit may be placed to allow these ions to escape the source (and form an ion beam). In an alternative arrangement, the extraction aperture may be placed in the anticathode. The entire source is placed at a high positive potential, with the outlet aperture facing the puller electrode at ground potential. The electric field accelerates the positive ions to form an ion beam.

![Fig. 2: A Penning or PIG ion source. Electron paths are shown (for illustration only), through the plasma region](image-url)

3.2 Electron Cyclotron Resonance Ion Source (ECRIS)

The ECR ion source is based on plasma heating at the electron cyclotron frequency ($f_{ecr}$) in a magnetic field, given by

$$\omega_{ecr} = 2\pi f_{ecr} = \frac{eB}{m}.$$
For electrons in a magnetic field in the range 0.05–1 T, this corresponds to a frequency range of 1.4 GHz to 28 GHz. The availability of commercial magnetrons and klystrons results in most sources working at 2.45, 10, 14.5, 18 and 28 GHz.

In an ECR source, the microwaves are fed into a low-pressure gas immersed in the magnetic field, produced by two solenoids and in most cases a multipole magnet structure. The scheme of the CERN ECR source, with the on-axis magnetic field superimposed, is shown in Fig. 3. This magnetic field configuration means that the centre of the plasma area lies at the minimum of the magnetic field.

Charged particles oscillating in a magnetic field have kinetic energy \( K \) given by the sum of the drift velocity parallel to the magnetic field and the rotational energy \( K = K_\parallel + K_\perp \).

These two energy components can be written as

\[
K_\parallel = \frac{mv_\parallel^2}{2} \quad K_\perp = \mu B = \frac{mv_\perp^2}{2}
\]

hence

\[
v_\parallel^2 = \frac{2}{m} (K - \mu B)
\]

where \( B \) is the magnetic field and \( \mu \) is the magnetic moment. \( K \) and \( \mu \) are constants of motion in an electric-field-free region. As the particle moves into a higher \( B \) field region, the energy is transferred from the parallel direction to the rotating frame. At a certain magnetic field, the parallel velocity will reduce to zero, and the charge particle is reflected back towards the lower magnetic field region.

By this mechanism, charged particles may be stored in a magnetic bottle. In this case, the electrons can be heated by the microwaves to very high energies. The ions may also be confined long enough to undergo many ionizing collisions, resulting in highly charged ions (at GSI Au\(^{24+}\), Pb\(^{27+}\) and U\(^{59+}\) are produced).

**Fig. 3:** ECR4 source at CERN, with indication of the plasma area, and the axial magnetic field
At CERN, the ECR source uses an after-glow increase in intensity that occurs when the microwave heating is switched off. By this method short, millisecond pulses of much higher ion intensity can be produced (Fig. 4).

### 3.3 Negative ions

Negative ions are required for certain acceleration schemes, including charge-exchange injection into synchrotrons.

The binding energy of an additional electron to an atom is termed the electron affinity ($E_{\text{ea}}$). For noble elements, this energy is less than zero and hence stable negative ions cannot be formed. In contrast, halogens have a large electron affinity. The value of the $E_{\text{ea}}$ for the first elements of the periodic table are given in Table 3.

<table>
<thead>
<tr>
<th>Element</th>
<th>$E_{\text{ea}}$ (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>0.7542</td>
</tr>
<tr>
<td>Li</td>
<td>0.6182</td>
</tr>
<tr>
<td>Be</td>
<td>&lt; 0</td>
</tr>
<tr>
<td>B</td>
<td>0.277</td>
</tr>
<tr>
<td>C</td>
<td>1.2629</td>
</tr>
<tr>
<td>N</td>
<td>&lt; 0</td>
</tr>
<tr>
<td>O</td>
<td>1.462</td>
</tr>
<tr>
<td>F</td>
<td>3.399</td>
</tr>
<tr>
<td>Ne</td>
<td>&lt; 0</td>
</tr>
</tbody>
</table>

The production of negative ions from a source is broadly split into two types, surface production and volume production.

In the surface production source, the principle involves the desired ion species (A) being absorbed onto a surface with a low work-function. In this case, the surface is usually a tungsten or molybdenum substrate, coated with cesium or Cs$_2$O, for which the work-function can be lowered to 1.1 eV in the case of Cs/O/W. When the ion of interest is desorbed or sputtered from the surface (by another A or Cs atom), there is a high probability it will carry a donor electron. It is essential to control the Cs layer on the surface, which usually requires an oven containing Cs to be connected to the source.

Volume production of negative ions is possible in a plasma, even though the low binding energy of the extra electron makes the ion very fragile. However, the cross section for electron capture by a vibrationally excited molecule and its subsequent disassociation into a neutral and negative ion (i.e. $\text{H}_2^* + e \rightarrow \text{H}^- + \text{H}$) is much higher than the equivalent destruction cross section. The optimization...
of this process in a source is required, as the excitation of H$_2$ requires electrons of 40 eV, while the maximum cross section for attachment electrons occurs at 0.5–1.0 eV [6].

Negative hydrogen ions have been extracted from many different source types, including Penning, duoplasmatrons, ECR, magnetrons etc. The source must be optimized to produce a region with a warm plasma for molecule excitation, and another with a cooler plasma for the electron attachment. The separation of the two regions can be made with a magnetic filter.

The introduction of a gas with electron donors (for example Cs) can greatly increase the output current of negative ions, even in a volume source.

![Diagram of negative-ion extraction](image)

**Fig. 5:** Scheme of negative-ion extraction, where the electrons are dumped onto an intermediate electrode

The extraction of a negative ion beam is complicated by the fact that electrons will also be pulled from the plasma, with an electron current that can be up to hundreds of times higher than the ion current. This makes a great demand not only for the beam dynamics in the region of extraction, but also requires a high voltage and high-current supply.

The current of electrons pulled from the plasma to the full beam potential can be greatly reduced by applying a perpendicular magnetic field, causing the electrons to be pulled onto a dumping electrode at a potential similar to, but higher than, the source potential. In this way, the total power of electrons can be brought to manageable levels (Fig. 5).

## 4 CONCLUSION

The physics and engineering of electron and ion source is a vast topic only touched on in this paper. The fundamental processes behind thermionic and photocathode electron sources have been addressed. Strong research of photocathodes’ ability to prolong the lifetime at high quantum efficiencies, and optimization of the production of short bunches is ongoing.

Furthermore, the Penning and ECR ion sources have been discussed. The Penning ion source is now a mature technology, while research in ECR sources is strongly pursued in many laboratories, utilizing different frequencies, magnetic field configurations, and methods of introducing non-gaseous ions into the plasma.

Negative hydrogen ions are the particle of choice for the next generation of high-power accelerators. Research into high-current, high-duty-factor, low-emittance H$^-$ sources is highly active.
References


Bibliography


Abstract
The purpose of this note is to serve as an introduction to MAD-X for the novice user. It describes the basic building elements and the most important commands to define a machine and perform the most important calculations. It cannot replace a reference manual for the more advanced and demanding use of MAD-X, but should help starting up the design of a new accelerator lattice and to understand existing MAD-X input. This note is a writeup of a presentation given at a course on optics design at the CERN Accelerator School at DESY Zeuthen, 2003.

1 Introduction
The MAD-X (Methodical Accelerator Design) program is a general purpose accelerator and lattice design program. The main aim of this note is to allow the beginner to understand the main building blocks of MAD-X and to set up and use a basic machine. The most important features are described and discussed. Details and more advanced options must be left to the MAD-X reference manual [1].

This note is a writeup of a presentation given at a course on optics design at the CERN accelerator school [2].

The main objectives of an accelerator design program are:

- Read the elements and their sequence from a file.
- Calculate the optics parameters from a machine description.
- Define and compute (match) the desired properties of such a machine.
- Simulate and correct possible machine imperfections.
- Simulate the beam dynamics in the designed machine.

Both, circular and linear accelerators (or beam lines) are normally handled by such programs.

1.1 What is needed as input to design an accelerator lattice?
The design of a machine requires the definition of the machine and the basic ingredients are:

- Definition of the properties of all machine elements.
- Strength of all active machine elements.
- Position of all machine elements in the machine, i.e. the order in which they appear in the accelerator or beam line.

These definitions are helped by a well adapted and defined input language.

2 MAD-X language
For the definition of a machine and the execution of actions, an input language is used following a well defined syntax and grammar.
2.1 Coordinate system

A circular or linear machine in MAD-X is a sequence of elements placed along the reference orbit which is defined as the path through the ideal magnets of a charged particle with the reference momentum (Fig. 1). This defines the coordinate system \((x, y, s)\) where \(x\) is the transverse coordinate in the bending plane (usually horizontal, positive values to the outside) and \(y\) is the orthogonal transverse coordinate (usually vertical, positive values upward). The local variable \(s\) is the tangent to the reference orbit. The coordinate system is therefore a local system moving along the reference path.

2.2 Conventions

**Units:**

For all computations MAD-X uses SI units, i.e. in particular metre and radian for length and angle.

**Input:**

The MAD-X input consists of a sequence of statements (commands, actions, or declarations etc.). All statements are free format with ",," as separators, can occupy any number of lines and are terminated by a semicolon ";;".

Blank lines have no effect. Statements are not case sensitive with the exception of strings enclosed in " ".

Commands or statements can be given a label which allows them to be used or re-used later on:

LAB01: mycommand;

**Comments:**

A single line (not a statement across several lines !) can be commented by "!" or "/\". A region of lines can be commented by enclosing them between "/*" and "/*/."
2.3 Variables and expressions

Parameter values:
Integer or floating point numbers can be assigned to named parameters and can be used in further declarations or commands, e.g.:

LENDIP = 8.0;

Various names (keywords used by MAD-X) are protected and cannot be used as variables or labels.

The numerical value of an assignment can be replaced by an expression, e.g.:

LNFX = 6912.00/(NCSEL*(4*LBEND + 2*LQUAD));

Expressions:
Parameters and variables can be used in expressions, in particular to define dependent quantities. Standard arithmetic operations and functions such as SQRT(), EXP(), trigonometric functions etc. can be used in the expressions as well as random number generators [1]. For instance,

E.g.: ANGLE = 2.0*PI/NBEND;

can compute the bending angle of a dipole, given the total number of bending magnets (nbend). The constant PI is predefined in MAD-X, together with many other important constants and particle properties.

Deferred expressions:
The usual expressions are evaluated once when the parameter is used, e.g.,

DX = GAUSS()*0.001;

assigns a random number following a Gaussian distribution with a width of 1 mm to the parameter DX. This value is kept in all computations.

A deferred expression is declared by " := " instead of " = " and is evaluated every time the parameter is used, e.g.,

DX := GAUSS()*001;

assigns a different random number everytime the parameter DX is used in the program.

The distinction between normal and deferred expressions becomes important for error assignment and matching.

3 Machine description

3.1 Thick and thin elements
For the calculations, the elements can be defined either as so-called thick lenses with a finite length or as so-called thin lenses with zero length. In the latter case, the effects of an element (e.g., a magnet) on the beam are represented as impulses (kicks) at a fixed value s on the reference orbit. This simplifies the treatment since it allows to treat the machine as a series of linear transformations separated by the "kicks" at the positions of the thin elements. This method is very fast and symplectic by construction and it is therefore best suited for particle tracking.

The disadvantage is the loss of precision when the magnets are very long (compared to the size of the machine) or when fringe fields are important. Part of this precision can be recovered by sub-dividing the magnets into slices, i.e., shorter sections, each representing a thin lens.
3.2 Element definition

Elements are defined using the concept of element classes. All quadrupoles in a machine belong to the class QUADRUPOLE. We can define subclasses with different properties with statements like:

MQL: QUADRUPOLE, L=5.0;
MQS: QUADRUPOLE, L=1.5;

where we define two classes (MQL and MQS) of quadrupoles of different length (thick elements). The definitions can be used to define the real quadrupoles such as:

QFL01: MQL; // Focusing quadrupoles
QFS01: MQS;
QDL01: MQL; // Defocusing quadrupoles
QDS01: MQS;

The quadrupoles defined like this inherit all properties of the class unless they are specified explicitly, in which case they are overwritten. All numerical attributes in a class definition can be expressions.

Dipole magnets can be defined as rectangular (RBEND) or sector (SBEND) bending magnets, e.g.,

MBL: RBEND, L=14.3;
MBS: RBEND, L=5.0;

The length of a rectangular bending magnet is by default the arc length. All details on the definition of bending magnets are found in the reference manual [1].

3.3 Element strength definition

3.3.1 Dipoles

The strength of a bending magnet is specified by the bending angle or alternatively the dipole coefficient \( k_0 \):

\[
k_0 = \frac{1}{p/c} \frac{\delta B_y}{\delta \rho} \text{[in T]} = \frac{1}{\rho} = \frac{\text{angle}}{l} \text{[in rad/m]}
\]

In the latter case, a finite length must be specified.

The definition for a dipole magnet is:

MB001: RBEND, L=14.3, ANGLE=2*PI/1132; //Total number of dipoles is 1132

or, alternatively:

MB001: MBL, ANGLE=2*PI/1132; //Total number of dipoles is 1132

using the defined sub-class.

3.3.2 Quadrupoles

We define a quadrupole by its quadrupole coefficient \( k_1 \), which is defined as:

\[
k_1 = \frac{1}{p/c} \frac{\delta B_y}{\delta x} \text{[in T/m]} = \frac{1}{l \cdot f}
\]

We define quadrupoles as:

QF007: QUADRUPOLE, L=5.0, K1 = +0.00147235;
QD007: QUADRUPOLE, L=5.0, K1 = -0.00147235;

or using sub-classes:

QF007: QFL, K1 = +0.00147235;
QD007: QDL, K1 = -0.00147235;
3.3.3 Sextupoles
Higher order multipoles such as sextupoles we can define as:

\[
SF007: \text{SEXTUPOLE, L=1.4, } K2 = +0.00147235;
\]

with:

\[
k2 = \frac{1}{p/c} \frac{\delta^2 B_y}{\delta^2 x} \quad \text{[in T/m$^2$]}
\]

3.3.4 Orbit correctors
Orbit correction dipoles are identified by the keyword KICKER. The strength of an orbit corrector is the deflection angle (KICK) measured in rad. Valid definitions are:

\[
\begin{align*}
LKICK &= 0.1; \\
MCV01: & \text{VKICKER, L=0.1, } \text{KICK:=KCV01;} \\
MCV02: & \text{VKICKER, L=LKICK, } \text{KICK:=KCV02;} \\
MCV01: & \text{HKICKER, L=LKICK, } \text{KICK:=KCH01;} \\
MC001: & \text{KICKER, L=0.1, } \text{VKICK:=KXV001, } \text{HKICK:=KXH001;}
\end{align*}
\]

The class VKICKER or HKICKER refer to orbit correctors for the vertical and horizontal planes respectively. The attribute KICK refers to the corresponding plane only. The single class KICKER can be used to specify orbit correctors for both planes. In that case, two attributes HKICK and VKICK are needed to separate the functions in the two planes.

In the above example, the correctors and their strengths are given individual names which allows to set them explicitly to independent values. For the standard orbit correction with MAD-X this is however not always necessary. The declarations of the kicks as deferred expressions allows the kicks to be changed explicitly or by the program.

3.4 Multipoles
A special class of elements is defined with the keyword MULTIPOLE. These are general elements of zero length (thin lenses) and can be used with one or more components of any order. All thin elements can be written as multipoles in the form:

\[
\begin{align*}
\text{MPM: MULTIPOLE;} \\
\text{MPLE01: } & \text{MPM, LRAD=0.0, TILT=angle, } \\
& \quad \text{KNL=\{k_{n0L}, k_{n1L}, k_{n2L}, k_{n3L}, \ldots\}, } \\
& \quad \text{KSL=\{k_{s0L}, k_{s1L}, k_{s2L}, k_{s3L}, \ldots\};}
\end{align*}
\]

The components KNL and KSL are the normal and skew components of the multipole multiplied by the relevant magnetic length. Note that the strength definitions are position dependent, therefore leading zeroes must be filled for components that do not exist. The attribute LRAD is a fictitious length, which is only used to compute synchrotron radiation effects. For the computation of lattice functions etc., it can be set to some dummy value.

Using multipoles, a thin quadrupole can be defined as:

\[
\begin{align*}
\text{QFT: } & \text{MPM, LRAD=0, } \text{KNL=\{0, k_{n1L}, 0, 0 \};}
\end{align*}
\]

The thin lens version of a dipole would be written as:

\[
\begin{align*}
\text{MBT: } & \text{MPM, LRAD=0, } \text{KNL=\{k_{n0L}, 0, 0, 0 \};}
\end{align*}
\]
3.5 Markers

The element class MARKER is used to insert an inactive element at a position \( s \) for later use, e.g., as a reference. The syntax is:

\[
\text{START_IP: MARKER, \text{AT} = 1839.872;}
\]

If present in a sequence, the lattice functions are calculated at their positions and they play an important role for matching.

A complete list of keywords and pre-defined element classes is found in the reference manual [1].

3.6 Element position in a SEQUENCE

The representation of the machine is called a sequence. It defines the order in which the elements appear in the accelerator or beam line. In a simple case, a sequence can be defined like:

\[
\text{seq_name: SEQUENCE, \text{REFER=CENTRE, LENGTH=6912.00;}} \quad \ldots \quad \ldots
\]

MQF05 : MQL, \text{AT} = 256.0000;
BPMH05 : MONITOR, \text{AT} = 1.75, \text{FROM=MQF05;}
MCH05 : HKICKER, \text{AT} = 2.10, \text{FROM=MQF05;}
MBL05.002: MBL, \text{AT} = 266.9000;
MBL05.002: MBL, \text{AT} = 278.1000;
MQR05 : MQD, \text{AT} = 288.0000;
BPMV05 : MONITOR, \text{AT} = 1.75, \text{FROM=MQR05;}
MCV05 : VKICKER, \text{AT} = 2.10, \text{FROM=MQR05;}
MBL05.003: MBL, \text{AT} = 297.9000;
MBL05.004: MBL, \text{AT} = 310.1000;
\ldots \quad \ldots
\]

\text{ENDSEQUENCE;}

The keywords SEQUENCE and ENDSEQUENCE define the beginning and end of the definition and the sequence is assigned a name \text{seq_name}.

The statements look familiar and the additional attribute \text{AT} defines the position relative to the beginning of the sequence. A position relative to an existing element can be assigned with the \text{FROM} attribute. The total \text{LENGTH} of the sequence is specified on the header line of the sequence. The positions can be defined at the CENTRE, ENTRY or EXIT of an element, indicated by the \text{REFER} attribute. The names given to the elements must be unique, i.e., must not appear twice in the same sequence.

Several sequences with different names can be defined in the same file.

In the example above we have assigned the position to named elements. A second possibility is to use class names like in:

\[
\ldots
\]

MBL: MBL, \text{AT} = 278.1000;
MQD: MQD, \text{AT} = 288.0000;
BPM: BPM, \text{AT} = 1.75, \text{FROM=MQD05;}
MCV: MCV, \text{AT} = 2.10, \text{FROM=MQD05;}
MBL: MBL, \text{AT} = 297.9000;
MBL: MBL, \text{AT} = 310.1000;
\ldots
\]

assuming BPM, MCV, etc. have been defined as classes before. In this case all elements have the same name which is the name of the class and they cannot be distinguished by name.
Finally, a previously defined sequence can be inserted, allowing the possibility to nest sequences. Example 1 in Appendix A defines a complete machine using the commands already discussed up to now. The dipoles are defined as thin elements whereas the quadrupoles and sextupoles have a finite length.

### 3.7 Using repetitive definition for periodic machines

The sequence of a periodic machine or the periodic part of a machine can be defined using the MAD-X macro commands. After the usual definition of the cell length \( l_{\text{cell}} \), the half length of a quadrupole \( l_{\text{quad2}} \) and the number of cells \( n_{\text{cell}} \), the whole machine can be defined with a `while`-loop:

```plaintext
n = 1;
while (n < n_{\text{cell}}+1) {
    qf: qf, at=(n-1)*l_{\text{cell}}+l_{\text{quad2}};
    lsf: lsf, at=(n-1)*l_{\text{cell}}+l_{\text{quad2}}+2.5;
    ch: ch, at=(n-1)*l_{\text{cell}}+l_{\text{quad2}}+3.1;
    bpm: bpm, at=(n-1)*l_{\text{cell}}+l_{\text{quad2}}+3.2;
    mbps: mbps, at=(n-1)*l_{\text{cell}}+l_{\text{quad2}}+3.5;
    mbps: mbps, at=(n-1)*l_{\text{cell}}+l_{\text{quad2}}+9.9;
    mbps: mbps, at=(n-1)*l_{\text{cell}}+l_{\text{quad2}}+22.10;
    mbps: mbps, at=(n-1)*l_{\text{cell}}+l_{\text{quad2}}+28.50;
    qd: qd, at=(n-1)*l_{\text{cell}}+l_{\text{quad2}}+32.00;
    lsd: lsd, at=(n-1)*l_{\text{cell}}+l_{\text{quad2}}+34.50;
    cv: cv, at=(n-1)*l_{\text{cell}}+l_{\text{quad2}}+35.10;
    bpm: bpm, at=(n-1)*l_{\text{cell}}+l_{\text{quad2}}+35.20;
    mbps: mbps, at=(n-1)*l_{\text{cell}}+l_{\text{quad2}}+35.50;
    mbps: mbps, at=(n-1)*l_{\text{cell}}+l_{\text{quad2}}+41.90;
    mbps: mbps, at=(n-1)*l_{\text{cell}}+l_{\text{quad2}}+54.10;
    mbps: mbps, at=(n-1)*l_{\text{cell}}+l_{\text{quad2}}+60.50;
    n = n + 1;
}
```

The two types of sequence definitions are entirely equivalent.

A complete definition using this technique is given in example 2 in Appendix B. The MAD-X commands and the resulting output are identical to example 1 (Appendix A).

### 4 MAD-X commands

In addition to the statements which are used to define a machine, the MAD-X commands are used to define and execute actions on the machines, e.g. calculations of Twiss functions, I/O of the lattices, particle tracking etc.. An important part of the design procedure is lattice matching, i.e. to vary element parameters to make machine properties (e.g., Twiss functions) assume defined values at specified positions (e.g., interaction points, etc.). A complete description of all MAD-X commands is found in [1]. Here I shall list the most important commands which are necessary to do the basic calculations.

#### 4.1 BEAM command

Some of the MAD-X actions require the knowledge of the beam properties. They are defined with the `BEAM` command:

```plaintext
BEAM, PARTICLE=name, ENERGY=xxx, SEQUENCE=sname;
```

The name of the particle type can be given as well as the particle’s energy. The properties (e.g., mass and charge) of the most important particles are known to MAD-X. Alternatively, the mass and charge can be specified with `MASS =` and `CHARGE =`. When the `SEQUENCE` attribute is given, it will assign the beam properties to the particles at the specified locations.
only to this particular sequence, otherwise to the active sequence. A complete list of all possible beam quantities is found in [1].

4.2 Input definition

MAD-X statements and commands can be given on the standard input or can be read from a file with:

\texttt{CALL,FILE="filename";}

This file can contain one or more sequences, part of a sequence or commands and is inserted at the position of the call.

After a sequence has been read, it can be used with:

\texttt{USE, PERIOD=sname;}

This command will expand the specified sequence, insert the drift spaces and make it active.

4.3 MAD-X actions

MAD-X actions are executed to perform operations on the available machines. To calculate the linear lattice functions (Twiss parameters) around the machine, the action:

\texttt{TWISS;}

must be executed, which operates on the sequence defined in the last USE command. However, a sequence can be specified explicitly on the Twiss command. A summary table is given after execution.

4.4 MAD-X output

The TWISS command can be modified to specify the wanted output:

\texttt{SELECT,FLAG=TWISS,COLUMN=NAME,S,MUX,BETX,MUY,BETY; TWISS,FILE="twiss.out";}

In the SELECT command the lattice functions wanted can be specified before TWISS is executed. The full list of the lattice functions is given in [1]. The lattice functions are written into the file "twiss.out".

The SELECT command can be used to restrict the output to only a range or type of elements:

\texttt{SELECT,FLAG=TWISS,RANGE=beg/end;}

or:

\texttt{SELECT,FLAG=TWISS,PATTERN="Q.*";}

The first will output the lattice functions only within the specified range and the second would restrict the output to all elements starting with the specified pattern in the element name. The SELECT commands act for the desired action (FLAG=) and can be accumulated or overwritten.

4.5 MAD-X graphical output

MAD-X has a built-in graphics package. To plot lattice functions for example, the command sequence:

\texttt{SELECT,FLAG=TWISS; TWISS,FILE="twiss.beta"; PLOT, HAXIS=S, VAXIS=BETX, BETY;}

may be used to plot the horizontal and vertical $\beta$-functions as a function of the position $s$. The RANGE attribute can be used with the PLOT command. An output file and a PostScript file are written simultaneously. For details and all options see [1].
4.6 MAD-X example

In the second part of example 1 the necessary MAD-X commands are given to calculate the lattice functions with the TWISS command, write them onto a file and plot them in postscript format. At the end of the execution of a TWISS command, a summary table is printed:

```
++++++ table: summ

<table>
<thead>
<tr>
<th>length</th>
<th>orbit5</th>
<th>alfa</th>
<th>gammatr</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.9120000e+03</td>
<td>-0.0000000e+00</td>
<td>1.6807003e-03</td>
<td>2.4392418e+01</td>
</tr>
<tr>
<td>q1</td>
<td>dq1</td>
<td>betxmax</td>
<td>dxmax</td>
</tr>
<tr>
<td>2.6580000e+01</td>
<td>-3.3561557e+01</td>
<td>1.0754431e+02</td>
<td>2.5680113e+00</td>
</tr>
<tr>
<td>dxrms</td>
<td>xcomax</td>
<td>xcorms</td>
<td>q2</td>
</tr>
<tr>
<td>1.9304378e+00</td>
<td>0.0000000e+00</td>
<td>0.0000000e+00</td>
<td>2.6620000e+01</td>
</tr>
<tr>
<td>dq2</td>
<td>betymax</td>
<td>dymax</td>
<td>dyrms</td>
</tr>
<tr>
<td>-3.3598479e+01</td>
<td>1.0749730e+02</td>
<td>0.0000000e+00</td>
<td>0.0000000e+00</td>
</tr>
<tr>
<td>ycomax</td>
<td>ycorms</td>
<td>deltap</td>
<td></td>
</tr>
<tr>
<td>0.0000000e+00</td>
<td>0.0000000e+00</td>
<td>0.0000000e+00</td>
<td></td>
</tr>
</tbody>
</table>
```

The main parameters of the lattice are summarized in this table, such as horizontal and vertical tunes (Q1, Q2), chromaticities (DQ1, DQ2), etc.

The lattice functions $\beta_x$ and $D_x$ are plotted by the above command sequence as shown in Fig. 2.

The functions are plotted between the 10th and 16th quadrupole of the class QD as specified in the RANGE attribute.

As requested, the lattice functions are written to a file "twiss.out" and its format is shown in the last part of Appendix A. At the beginning of this file the basic parameters are summarized again.

5 Matching with MAD-X

The adjustment of machine properties, i.e., matching, is a vital part of the design process and a detail description is beyond the scope of this introduction. However, some very basic features will be demonstrated by some examples. A basic tutorial on some matching techniques is found in [3, 4].

5.1 Global matching

Some global machine parameters such as tune or chromaticity can be adjusted by global matching. The following two examples are used as a demonstration:

```
MATCH, SEQUENCE=CASSPS;
  VARY, NAME=KQF, STEP=0.00001;
  VARY, NAME=KQD, STEP=0.00001;
  GLOBAL, SEQUENCE=CASSPS, Q1=26.58;
  GLOBAL, SEQUENCE=CASSPS, Q2=26.62;
  LMDIF, CALLS=10, TOLERANCE=1.0E-21;
ENDMATCH;

MATCH, SEQUENCE=CASSPS;
  VARY, NAME=KSF, STEP=0.00001;
  VARY, NAME=KSD, STEP=0.00001;
  GLOBAL, SEQUENCE=CASSPS, DQ1=0.0;
  GLOBAL, SEQUENCE=CASSPS, DQ2=0.0;
  LMDIF, CALLS=10, TOLERANCE=1.0E-21;
ENDMATCH;
```
Fig. 2: Lattice functions computed and plotted by MAD-X
The matching attributes and commands are enclosed between the MATCH and ENDMATCH statements. The desired sequence can be specified. In the first example the global horizontal and vertical tunes are matched to the desired values. The strengths of the main quadrupoles (KQF and KQD) are varied in the procedure. In the second example the global chromaticities are matched to zero, by variation of the sextupole strengths KSF and KSD. Other attributes control the method used and the quality of the procedure. The new values of the strengths are now associated with the sequence. The following execution of the TWISS command would therefore produce the new parameters.

**NOTE:** the latter is only true when the quadrupole strengths are defined using deferred expressions, i.e.,

```
QF: QUADRUPOLE, K1:=KQF;
```

Otherwise the new strengths KQF and KQD are calculated all right, but not assigned to the elements, i.e., they will not be used in subsequent calculations, e.g., computing lattice functions with the TWISS command.

### 5.2 Local and insertion matching

Probably the most important matching procedures are those which are used to modify the lattice locally, e.g., for inserting non-periodic regions for experiments, collimation systems etc. In order to avoid a distortion of other parts of the machine, the matching must be restricted to the local region and additional constraints must ensure the smooth continuation into the periodic part of the machine.

The example below is a simple matching of a symmetric low $\beta$-insertion using four independent quadrupoles. The matching is restricted to the range between the elements *left* and *right* where the normal lattice parameters are given as constraints. For more details on this example see [4].

```
MATCH, SEQUENCE=CASCELL5, RANGE=LEFT/RIGHT, BETX=28.2, BETY=87.0;
VARY, NAME=KQ1.L, STEP=0.00001;
VARY, NAME=KQ2.L, STEP=0.00001;
VARY, NAME=KQ3.L, STEP=0.00001;
VARY, NAME=KQ4.L, STEP=0.00001;
CONSTRAINT, RANGE=RIGHT, SEQUENCE=CASCELL5, BETX=28.2, BETY=87.0;
CONSTRAINT, RANGE=IP, SEQUENCE=CASCELL5, BETX=10.0, BETY=1.0;
LMDIF, CALLS=100, TOLERANCE=1.0E-21;
ENDMATCH;
```

### 6 Error definition

During the design process of a machine it becomes important to test it against imperfections. For that purpose, alignment and field errors can be assigned to all machine elements. The calculations will take these imperfections into account and correction procedures (e.g. orbit correction) are available in MAD-X to test possible correction strategies.

#### 6.1 Alignment errors

The elements in a machine can be misaligned with the available MAD-X error actions. The command sequence:

```
SELECT,FLAG=ERROR,CLASS=MQ;
EALIGN,DX:=GAUSS()*0.0005;DY:=GAUSS()*0.0002;
```

assigns alignment errors to all quadrupoles belonging to the class MQ with a r.m.s. value of 0.5 mm in the horizontal and 0.2 mm in the vertical plane, both following a Gaussian distribution. In that case again the
use of deferred expressions is of utmost importance. To assign the errors, the program steps through the sequence and for every element of the selected class the corresponding misalignments DX and DY are evaluated each time. Using the standard expression, the misalignments DX and DY are calculated once and all elements of the selected class get the same error. For a complete list of all misalignment options see [1].

6.2 Field errors
The program allows to assign field errors of any order to the machine elements with commands like:

\begin{verbatim}
SELECT, FLAG=ERROR, CLASS=MB;
EFCOMP, RADIUS:=0.017, ORDER:=0,
DKNR:={0,0,GAUSS()*7E-4,GAUSS()*1E-4,0,0},
DKSR:={0,0,GAUSS()*3E-4,GAUSS()*6E-4,0,0},
\end{verbatim}

In this example normal and skew field errors (sextupole and octupole) are assigned to dipole magnets of the class MB. It is possible to assign absolute or relative field errors, the latter normalized to the strength of the corresponding element. The RADIUS (reference for the measurement) and ORDER control this behaviour. For a detailed discussion see [1].

7 Orbit correction
A misaligned machine can be corrected using the MAD-X orbit correction procedures [5]. The input data is taken from the last TWISS table, i.e., TWISS must run before a correction can be executed.

Very basic closed orbit correction statements are of the form:

\begin{verbatim}
CORRECT, PLANE=X, NCORR=20, ERROR=1.0E-04;
\end{verbatim}

or

\begin{verbatim}
CORRECT, MODE=SVD;
\end{verbatim}

For all details and options see [1, 5].

8 Advanced options and commands
MAD-X has many more features and commands for advanced design and evaluation of accelerator lattices. Most prominent are the evaluation of beam parameters (in case of radiation), geometrical survey, tracking and physical and dynamic aperture determination. However, a full description is well beyond the scope of this simplified introduction. To get a flavour, I shall give two examples, one for a simple tracking and another for the advanced use of macros.

8.1 Particle tracking
The example shown below demonstrates particle tracking in MAD-X. It shows the simultaneous tracking of 20 particles in horizontal phase space where the particles are distributed on a circle in $x-px$ phase space. All commands are enclosed between the keywords TRACK and ENDTRACK. The initial coordinates of the particles are assigned with the command START and the tracking is executed with RUN.

Tracking in MAD-X is possible using thin lenses. A lattice defined with thick elements has to be converted to thin lenses with the command MAKETHIN before the tracking can be done. For details on the command MAKETHIN consult the reference manual [1].
The use of tracking may require additional attributes in the BEAM command. The above example shows the power of the input language.

8.2 Use of macros

The power of the MAD-X input language is further enhanced by the use of macros. For illustration, example 2 in Appendix B has been modified. For some cases, it is required that the elements have distinct names, (e.g., where all elements must be treated as separate objects, such as orbit corrections). This can be easily done by editing example 1, where every element is listed on a separate line. Editing the example 2 with the while loop would fail because two elements must not have the same name. Using the macro language, the while loop can be modified like in example 3 in Appendix C. The definition of a cell is now done within the subroutine `inst`. This subroutine takes input parameters from the calling MAD-input and most important, can change the names of the elements, using the input information. The result of this scheme is the same as before, however the orbit correctors, their corresponding strength parameters and the beam position monitors are now numbered sequentially. A MAD-X input file and the corresponding Twiss output are shown in the second part of example 3. The increasing sequence numbers as part of the element names are now clearly visible. In this example the quadrupoles have been misaligned in the two planes following a Gaussian distribution with 0.1 mm and 0.2 mm r.m.s. respectively.

Therefore the horizontal and vertical orbit is distorted and the maximum and r.m.s. values can be found in the Twiss summary table.

9 How to run MAD-X?

MAD-X can be run either interactively or in batch mode.

9.1 Interactive mode

To run MAD-X interactively, one can execute MAD-X and input the commands and statements in the command line of the standard input.

Alternatively, one or more files with commands and statements can be read using the command:

CALL,FILE="filename";
9.2 Batch mode

MAD-X can be run as a background of batch program by redirecting an input file into the MAD-X standard input. For UNIX (LINUX) like:

```bash
madx < inputfile
```

The output is normally send to the standard output, unless it is redirected.

Appendices

The appendices list examples which are referenced in the text. For all cases the sequence definition is given in the first part and a MAD-X input file to use the sequence is displayed in the second part of each example.

A Example 1, Simplest case

A.1 Sequence definition

```plaintext
// define the total length
circum=6912.00;

// define number of cells and therefore cell length
ncell = 108;
lcell = circum/ncell;

// define lengths of elements and half lengths
lquad = 3.085;
lmb = 6.260;
lsex = 1.0;

// forces and other constants;
// element definitions;

// define bending magnet as multipole
mbps: multipole, lrad=dummy, knl={2.0*pi/(8*ncell)};

// define quadrupole and their strengths
qsps: quadrupole, l=lquad;
qf: qsps, k1:=kqf;
qd: qsps, k1:=kqd;
kqf = 1.4631475E-02;
kqd = -1.4643443E-02;

// define sextupoles for chromaticity correction
lsf: sextupole, l=lsex,k2:=ksf;
lsd: sextupole, l=lsex,k2:=ksd;
ksf = 2.0284442E-02;
ksd = -3.8394267E-02;

// define orbit correctors and beam position monitors
bpm: monitor, l=0.1;
ch: hkicker, l=0.1;
cv: vkicker, l=0.1;

cassps: sequence, refer=centre, l = circum;
start_machine: marker, at = 0;
```

qf, at = 1.5425;
lsf, at = 4.0425;
ch, at = 4.6425;
bpm, at = 4.7425;
mbps, at = 5.0425;
mbps, at = 11.4425;
mbps, at = 23.6425;
mbps, at = 30.0425;
qd, at = 33.5425;
lsd, at = 36.0425;
cv, at = 36.6425;
bpm, at = 36.7425;
mbps, at = 37.0425;
mbps, at = 43.4425;
mbps, at = 55.6425;
mbps, at = 62.0425;
qf, at = 65.5425;
lsf, at = 68.0425;
ch, at = 68.6425;
bpm, at = 68.7425;
mbps, at = 69.0425;
mbps, at = 75.4425;
mbps, at = 87.6425;
mbps, at = 94.0425;
qd, at = 97.5425;
lsd, at = 100.0425;
cv, at = 100.6425;
bpm, at = 100.7425;
mbps, at = 101.0425;
mbps, at = 107.4425;
mbps, at = 119.6425;
mbps, at = 126.0425;
qf, at = 129.5425;
lsf, at = 132.0425;
ch, at = 132.6425;
bpm, at = 132.7425;
mbps, at = 133.0425;
mbps, at = 139.4425;
mbps, at = 151.6425;
mbps, at = 158.0425;
qd, at = 161.5425;
lsd, at = 164.0425;
cv, at = 164.6425;
bpm, at = 164.7425;
mbps, at = 165.0425;
mbps, at = 171.4425;

...
A.2 MAD-X directives

TITLE, s='MAD-X test';

// Read input file with machine description
call file="spsall.seq";
option,-echo;

// Define the beam for the machine
Beam, particle = proton, sequence=cassps, energy = 450.0;

// Use the sequence with the name: cassps
use, period=cassps;

// Define the type and amount of output for the action TWISS
select,flag=twiss,column=name,s,x,y,mux,betx,muy,bety,dx,dy;

// Execute the Twiss command to calculate the Twiss parameters
// Compute at the centres of the elements and write to: twiss.out
twiss,centre,file=twiss.out;

// Plot the horizontal and vertical beta function between the
// 10th and 16th occurrence of a defocussing quadrupole
plot, haxis=s, vaxis=x, betx, bety,colour=100,
    range=qd[10]/qd[16];
plot, haxis=s, vaxis=dx, colour=100,
    range=qd[10]/qd[36];
stop;
### A.3 TWISS summary table

<table>
<thead>
<tr>
<th>length</th>
<th>orbit5</th>
<th>alfa</th>
<th>gammatr</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.9120000e+03</td>
<td>-0.0000000e+00</td>
<td>1.6807003e-03</td>
<td>2.4392418e+01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>q1</th>
<th>dq1</th>
<th>betxmax</th>
<th>dxmax</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.6580000e+01</td>
<td>-3.3561557e+01</td>
<td>1.0754431e+02</td>
<td>2.5680113e+00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>dxrms</th>
<th>xcomax</th>
<th>xcorms</th>
<th>q2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.9304378e+00</td>
<td>0.0000000e+00</td>
<td>0.0000000e+00</td>
<td>2.6620000e+01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>dq2</th>
<th>betymax</th>
<th>dymax</th>
<th>dyrms</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3.3598479e+01</td>
<td>1.0749730e+02</td>
<td>0.0000000e+00</td>
<td>0.0000000e+00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ycomax</th>
<th>ycorms</th>
<th>deltap</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0000000e+00</td>
<td>0.0000000e+00</td>
<td>0.0000000e+00</td>
</tr>
</tbody>
</table>

### A.4 TWISS lattice functions written to the file "twiss.out"

```plaintext
@ NAME %05s "TWISS"
@ TYPE %05s "TWISS"
@ SEQUENCE %06s "CASSPS"
@ PARTICLE %06s "PROTON"
@ MASS %le 0.938271998
@ CHARGE %le 1
@ ENERGY %le 450
@ PC %le 449.999021827
@ GAMMA %le 479.605062241
@ KBUNCH %le 1
@ BCURRENT %le 0
@ SIGE %le 0
@ SIGT %le 0
@ NPART %le 0
@ EX %le 1
@ EY %le 1
@ ET %le 1
@ LENGTH %le 6912
@ ALFA %le 0.00168070032886
@ ORBIT5 %le -0
@ GAMMATR %le 24.3924182122
@ Q1 %le 26.58
@ Q2 %le 26.62
@ DQ1 %le -33.5615573373
@ DQ2 %le -33.5984799903
@ DXMAX %le 2.5680113011
@ DYMAX %le 0
@ XCOMAX %le 0
@ YCOMAX %le 0
@ BETXMAX %le 107.544319159
@ BETYMAX %le 107.497305443
@ XCORMS %le 0
@ YCORMS %le 0
@ DXRMS %le 1.93043782638
@ DYRMS %le 0
@ DELTAP %le 0
```
<table>
<thead>
<tr>
<th>NAME</th>
<th>S</th>
<th>BETX</th>
<th>DX</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>CASSPS$START</code></td>
<td>0</td>
<td>103.8655173</td>
<td>2.523441048</td>
</tr>
<tr>
<td><code>START_MACHINE</code></td>
<td>0</td>
<td>103.8655173</td>
<td>2.523441048</td>
</tr>
<tr>
<td><code>QF</code></td>
<td>1.5425</td>
<td>107.5443192</td>
<td>2.568011301</td>
</tr>
<tr>
<td><code>DRIFT_0</code></td>
<td>3.11375</td>
<td>103.7300292</td>
<td>2.521784419</td>
</tr>
<tr>
<td><code>LSF</code></td>
<td>3.6425</td>
<td>101.2568359</td>
<td>2.49136849</td>
</tr>
<tr>
<td><code>DRIFT_1</code></td>
<td>4.5925</td>
<td>96.90195183</td>
<td>2.43657606</td>
</tr>
<tr>
<td><code>MBSPS</code></td>
<td>5.0425</td>
<td>94.87888064</td>
<td>2.410646213</td>
</tr>
<tr>
<td><code>DRIFT_2</code></td>
<td>8.2425</td>
<td>81.22989411</td>
<td>2.249527296</td>
</tr>
<tr>
<td><code>MBSPS</code></td>
<td>11.4425</td>
<td>68.87370326</td>
<td>2.08840838</td>
</tr>
<tr>
<td><code>DRIFT_3</code></td>
<td>17.5425</td>
<td>48.90078374</td>
<td>1.825635993</td>
</tr>
<tr>
<td><code>MBSPS</code></td>
<td>23.6425</td>
<td>33.62561103</td>
<td>1.562863606</td>
</tr>
<tr>
<td><code>DRIFT_2</code></td>
<td>26.8425</td>
<td>27.4909994</td>
<td>1.448286904</td>
</tr>
<tr>
<td><code>MBSPS</code></td>
<td>30.0425</td>
<td>22.64918345</td>
<td>1.333710202</td>
</tr>
<tr>
<td><code>DRIFT_4</code></td>
<td>31.02125</td>
<td>21.42644598</td>
<td>1.305783529</td>
</tr>
<tr>
<td><code>QD</code></td>
<td>33.5425</td>
<td>19.51348873</td>
<td>1.255914454</td>
</tr>
<tr>
<td><code>DRIFT_0</code></td>
<td>35.11375</td>
<td>20.35528979</td>
<td>1.278677181</td>
</tr>
<tr>
<td><code>LSF</code></td>
<td>35.6425</td>
<td>20.93741261</td>
<td>1.293764004</td>
</tr>
<tr>
<td><code>DRIFT_1</code></td>
<td>36.5925</td>
<td>22.07198555</td>
<td>1.320870352</td>
</tr>
<tr>
<td><code>MBSPS</code></td>
<td>37.0425</td>
<td>22.64918345</td>
<td>1.333710202</td>
</tr>
<tr>
<td><code>DRIFT_2</code></td>
<td>40.2425</td>
<td>27.4909994</td>
<td>1.448286904</td>
</tr>
<tr>
<td><code>MBSPS</code></td>
<td>43.4425</td>
<td>33.62561103</td>
<td>1.562863606</td>
</tr>
<tr>
<td><code>DRIFT_3</code></td>
<td>49.5425</td>
<td>48.90078374</td>
<td>1.825635993</td>
</tr>
<tr>
<td><code>MBSPS</code></td>
<td>55.6425</td>
<td>68.87370326</td>
<td>2.08840838</td>
</tr>
<tr>
<td><code>DRIFT_2</code></td>
<td>58.8425</td>
<td>81.22989411</td>
<td>2.249527296</td>
</tr>
<tr>
<td><code>MBSPS</code></td>
<td>62.0425</td>
<td>94.87888064</td>
<td>2.410646213</td>
</tr>
<tr>
<td><code>DRIFT_4</code></td>
<td>63.02125</td>
<td>99.31172846</td>
<td>2.467043631</td>
</tr>
<tr>
<td><code>QF</code></td>
<td>65.5425</td>
<td>107.5443192</td>
<td>2.568011301</td>
</tr>
<tr>
<td><code>DRIFT_0</code></td>
<td>67.11375</td>
<td>103.7300292</td>
<td>2.521784419</td>
</tr>
<tr>
<td><code>LSF</code></td>
<td>67.6425</td>
<td>101.2568359</td>
<td>2.49136849</td>
</tr>
<tr>
<td><code>DRIFT_1</code></td>
<td>68.5925</td>
<td>96.90195183</td>
<td>2.43657606</td>
</tr>
<tr>
<td><code>MBSPS</code></td>
<td>69.0425</td>
<td>94.87888064</td>
<td>2.410646213</td>
</tr>
<tr>
<td><code>DRIFT_2</code></td>
<td>72.2425</td>
<td>81.22989411</td>
<td>2.249527296</td>
</tr>
<tr>
<td><code>MBSPS</code></td>
<td>75.4425</td>
<td>68.87370326</td>
<td>2.08840838</td>
</tr>
<tr>
<td><code>DRIFT_3</code></td>
<td>81.5425</td>
<td>48.90078374</td>
<td>1.825635993</td>
</tr>
<tr>
<td><code>MBSPS</code></td>
<td>87.6425</td>
<td>33.62561103</td>
<td>1.562863606</td>
</tr>
<tr>
<td><code>DRIFT_2</code></td>
<td>90.8425</td>
<td>27.4909994</td>
<td>1.448286904</td>
</tr>
<tr>
<td><code>MBSPS</code></td>
<td>94.0425</td>
<td>22.64918345</td>
<td>1.333710202</td>
</tr>
<tr>
<td><code>DRIFT_4</code></td>
<td>95.02125</td>
<td>21.42644598</td>
<td>1.305783529</td>
</tr>
<tr>
<td><code>QD</code></td>
<td>97.5425</td>
<td>19.51348873</td>
<td>1.255914454</td>
</tr>
<tr>
<td><code>DRIFT_0</code></td>
<td>99.11375</td>
<td>20.35528979</td>
<td>1.278677181</td>
</tr>
<tr>
<td><code>LSF</code></td>
<td>99.6425</td>
<td>20.93741261</td>
<td>1.293764004</td>
</tr>
<tr>
<td><code>DRIFT_1</code></td>
<td>100.5925</td>
<td>22.07198555</td>
<td>1.320870352</td>
</tr>
<tr>
<td><code>MBSPS</code></td>
<td>101.0425</td>
<td>22.64918345</td>
<td>1.333710202</td>
</tr>
<tr>
<td><code>DRIFT_2</code></td>
<td>104.2425</td>
<td>27.4909994</td>
<td>1.448286904</td>
</tr>
<tr>
<td><code>MBSPS</code></td>
<td>107.4425</td>
<td>33.62561103</td>
<td>1.562863606</td>
</tr>
<tr>
<td><code>DRIFT_3</code></td>
<td>113.5425</td>
<td>48.90078374</td>
<td>1.825635993</td>
</tr>
<tr>
<td><code>MBSPS</code></td>
<td>119.6425</td>
<td>68.87370326</td>
<td>2.08840838</td>
</tr>
<tr>
<td><code>DRIFT_2</code></td>
<td>122.8425</td>
<td>81.22989411</td>
<td>2.249527296</td>
</tr>
<tr>
<td><code>MBSPS</code></td>
<td>126.0425</td>
<td>94.87888064</td>
<td>2.410646213</td>
</tr>
<tr>
<td><code>DRIFT_4</code></td>
<td>127.02125</td>
<td>99.31172846</td>
<td>2.467043631</td>
</tr>
</tbody>
</table>
B Example 2, Use of WHILE command

B.1 Sequence definition

// define the total length
circum = 6912.0;

// define number of cells and therefore cell length
ncell = 108;
lcell = circum/ncell;

// define lengths of elements and half lengths
lquad = 3.085;
lquad2 = lquad/2.;
lsex = 1.0;

// forces and other constants;
// element definitions;

// define bending magnet as multipole
mbsp: multipole, lrad=dummy, knl={2.0*pi/(8*ncell)};

// define quadrupole and their strengths
qsp: quadrupole, l=lquad;
qf: qsp, k1:=kqf;
qd: qsp, k1:=kqd;
kqf = 1.4631475E-02;
kqd = -1.4643443E-02;

// define sextupoles for chromaticity correction
lsf: sextupole, l=lsex,k2:=ksf;
lsd: sextupole, l=lsex,k2:=ksd;
ksf = 2.0284442E-02;
ksd = -3.8394267E-02;

// define orbit correctors and beam position monitors
bpm: monitor, l=0.1;
ch: hkicker, l=0.1;
cv: vkicker, l=0.1;

// sequence declaration;
cassps: sequence, refer=centre, l:=circum;
start_machine: marker, at = 0;

// This defines ONE cell, repeat NCELL times
// to get the full machine
// SPS has 8 bending magnets per cell
n = 1;
while (n < ncell+1) {
  qf: qf, at=(n-1)*lcell+lquad2;
  lsf: lsf, at=(n-1)*lcell+lquad2+2.5;
  ch: ch, at=(n-1)*lcell+lquad2+3.1;
}
bpm: bpm, \quad \text{at}=(n-1)\times lcell+lquad2+3.2;
mbsps: mbsps, \quad \text{at}=(n-1)\times lcell+lquad2+3.50;
mbsps: mbsps, \quad \text{at}=(n-1)\times lcell+lquad2+9.90;
mbsps: mbsps, \quad \text{at}=(n-1)\times lcell+lquad2+22.10;
mbsps: mbsps, \quad \text{at}=(n-1)\times lcell+lquad2+28.50;
qd: qd, \quad \text{at}=(n-1)\times lcell+lquad2+32.00;
lsd: lsd, \quad \text{at}=(n-1)\times lcell+lquad2+34.50;
cv: cv, \quad \text{at}=(n-1)\times lcell+lquad2+35.10;
bpm: bpm, \quad \text{at}=(n-1)\times lcell+lquad2+35.20;
mbsps: mbsps, \quad \text{at}=(n-1)\times lcell+lquad2+35.50;
mbsps: mbsps, \quad \text{at}=(n-1)\times lcell+lquad2+41.90;
mbsps: mbsps, \quad \text{at}=(n-1)\times lcell+lquad2+54.10;
mbsps: mbsps, \quad \text{at}=(n-1)\times lcell+lquad2+60.50;

n = n + 1;
}
end_machine: marker at=circum;
endsequence;

C Example 3, Use of MAD-X macros and imperfections

C.1 Sequence definition

// define a subroutine "inst" to insert elements
// with numbering
// all strings "nx" in the macro are replaced by
// the input value of nx.
inst(nx,n,lcell,lquad2): macro = {
  qf: qf, \quad \text{at}=(n-1)\times lcell+lquad2;
  lsf: lsf, \quad \text{at}=(n-1)\times lcell+lquad2+2.5;
  chnx: hkicker,l=0.0,kick=khchnx,
      \quad \text{at}=(n-1)\times lcell+lquad2+3.1;
  bpmhnx: monitor,l=0.0,
      \quad \text{at}=(n-1)\times lcell+lquad2+3.2;
  mbsps: mbsps, \quad \text{at}=(n-1)\times lcell+lquad2+3.50;
  mbsps: mbsps, \quad \text{at}=(n-1)\times lcell+lquad2+9.90;
  mbsps: mbsps, \quad \text{at}=(n-1)\times lcell+lquad2+22.10;
  mbsps: mbsps, \quad \text{at}=(n-1)\times lcell+lquad2+28.50;
  qd: qd, \quad \text{at}=(n-1)\times lcell+lquad2+32.00;
  lsd: lsd, \quad \text{at}=(n-1)\times lcell+lquad2+34.50;
  cvnx: vkicker,l=0.0,kick=kcvnx,
      \quad \text{at}=(n-1)\times lcell+lquad2+35.10;
  bpmvnx: monitor,l=0.0,
      \quad \text{at}=(n-1)\times lcell+lquad2+35.20;
  mbsps: mbsps, \quad \text{at}=(n-1)\times lcell+lquad2+35.50;
  mbsps: mbsps, \quad \text{at}=(n-1)\times lcell+lquad2+41.90;
  mbsps: mbsps, \quad \text{at}=(n-1)\times lcell+lquad2+54.10;
  mbsps: mbsps, \quad \text{at}=(n-1)\times lcell+lquad2+60.50;

  n = n + 1;
}

// define the total length
circum=6912.0;

// define number of cells and therefore cell length
ncell = 108;
lcell = circum/ncell;

// define lengths of elements and half lengths
lquad = 3.085;
lquad2 = lquad/2.;
lquad3 = 0.0;
lmb = 6.260;
lmb2 = lmb/2.;
lsex = 1.0;

// forces and other constants;
// element definitions;

// define bending magnet as multipole
mbps: multipole, lrad=dummy, l=lmb, knl={2.0*pi/(8*ncell)};

// define quadrupole and their strengths
qsps: quadrupole, l=lquad;
qf: qsps, k1:=kqf;
qd: qsps, k1:=kqd;
kqf = 1.4631475E-02;
kqd = -1.4643443E-02;

// define sextupoles for chromaticity correction
lsf: sextupole, l=lsex,k2:=ksf;
lsd: sextupole, l=lsex,k2:=ksd;
ksf = 2.0284442E-02;
ksd = -3.8394267E-02;

// define orbit correctors and beam position monitors
bpm: monitor, l=0.1;
ch: hkicker, l=0.1;
kv: vkicker, l=0.1;

// sequence declaration;
cassps: sequence, refer=centre, l=circum;
start_machine: marker, at = 0;

// This defines ONE cell, repeat NCELL times
// to get the full machine
// SPS has 8 bending magnets per cell
n = 1;
while (n < ncell+1) {
    // here we call the macro, cell number n is argument
    // and used for numbering the elements
    exec inst($n,n,lcell,lquad2);
}
end_machine: marker at=circum;
endsequence;

C.2  MAD-X directives

TITLE, s='MAD-X test';

// Read input file with machine description
// This machine is constructed with macro
// subroutine INST()
call file="spsmac.seq";
option,-echo;
// Define the beam for the machine
Beam, particle = proton, sequence=cassps, energy = 450.0;

// Use the sequence with the name: cassps
use, sequence=cassps;

eoption,add=false,seed=62971100;
select,flag=error,pattern="q.*";
ealign,dx=\text{tgauss(3.0)*1.0e-4},dy=\text{tgauss(3.0)*2.0e-4};
eprint;

// Define the type and amount of output
select,flag=twiss,class=monitor,column=name,s,x,betx;
select,flag=twiss,class=\text{vkicker},column=name,s,x,betx;
select,flag=twiss,class=\text{hkicker},column=name,s,x,betx;

// Execute the Twiss command to calculate the Twiss parameters
// Compute at the centre of the element and write to: twiss.out
twiss,save,centre,file=twiss.out;

stop;

C.3 TWISS summary table


C.4 TWISS lattice functions

@ NAME %05s "TWISS"
@ TYPE %05s "TWISS"
@ SEQUENCE %05s "CASSPS"
@ PARTICLE %05s "PROTON"
@ MASS %le 0.938271998
@ CHARGE %le 1
@ ENERGY %le 450
@ PC %le 449.999021827
@ GAMMA %le 479.605062241
@ KBUNCH %le 1
@ BCURRENT %le 0
<table>
<thead>
<tr>
<th>NAME</th>
<th>S</th>
<th>X</th>
<th>BETX</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;CH1&quot;</td>
<td>4.6425</td>
<td>7.270283965e-05</td>
<td>96.71513054</td>
</tr>
<tr>
<td>&quot;BPMH1&quot;</td>
<td>4.7425</td>
<td>7.074377128e-05</td>
<td>96.26414369</td>
</tr>
<tr>
<td>&quot;CV1&quot;</td>
<td>36.6425</td>
<td>-0.000609759073</td>
<td>22.16791457</td>
</tr>
<tr>
<td>&quot;BPMV1&quot;</td>
<td>36.7425</td>
<td>-0.0006136057601</td>
<td>22.29453352</td>
</tr>
<tr>
<td>&quot;CH2&quot;</td>
<td>68.6425</td>
<td>-0.001619774866</td>
<td>96.60029697</td>
</tr>
<tr>
<td>&quot;BPMH2&quot;</td>
<td>68.7425</td>
<td>-0.001616493006</td>
<td>96.15012782</td>
</tr>
<tr>
<td>&quot;CV2&quot;</td>
<td>100.6425</td>
<td>-0.0006809745368</td>
<td>22.33285114</td>
</tr>
<tr>
<td>&quot;BPMV2&quot;</td>
<td>100.7425</td>
<td>-0.0006809745368</td>
<td>22.33285114</td>
</tr>
<tr>
<td>&quot;CH3&quot;</td>
<td>132.6425</td>
<td>-0.0006702055164</td>
<td>96.69082398</td>
</tr>
<tr>
<td>&quot;BPMH3&quot;</td>
<td>132.7425</td>
<td>-0.0006672287746</td>
<td>96.23936818</td>
</tr>
<tr>
<td>&quot;CV3&quot;</td>
<td>164.6425</td>
<td>0.0002920301383</td>
<td>22.11019753</td>
</tr>
<tr>
<td>&quot;BPMV3&quot;</td>
<td>164.7425</td>
<td>0.0002953017665</td>
<td>22.23679301</td>
</tr>
<tr>
<td>&quot;CH4&quot;</td>
<td>196.6425</td>
<td>0.001157348134</td>
<td>96.6649292</td>
</tr>
<tr>
<td>&quot;BPMH4&quot;</td>
<td>196.7425</td>
<td>0.00115479396</td>
<td>96.21335536</td>
</tr>
<tr>
<td>&quot;CV4&quot;</td>
<td>228.6425</td>
<td>0.000373991881</td>
<td>22.08455422</td>
</tr>
<tr>
<td>&quot;BPMV4&quot;</td>
<td>228.7425</td>
<td>0.0003728320023</td>
<td>22.21117154</td>
</tr>
<tr>
<td>&quot;CH5&quot;</td>
<td>260.6425</td>
<td>1.317974197e-05</td>
<td>96.71965263</td>
</tr>
<tr>
<td>&quot;BPMH5&quot;</td>
<td>260.7425</td>
<td>1.236758651e-05</td>
<td>96.28877918</td>
</tr>
<tr>
<td>&quot;CV5&quot;</td>
<td>292.6425</td>
<td>-0.0002913559575</td>
<td>22.1823912</td>
</tr>
<tr>
<td>&quot;BPMV5&quot;</td>
<td>292.7425</td>
<td>-0.0002936789916</td>
<td>22.30912451</td>
</tr>
<tr>
<td>&quot;CH6&quot;</td>
<td>324.6425</td>
<td>-0.00091513218543</td>
<td>96.62975425</td>
</tr>
<tr>
<td>&quot;BPMH6&quot;</td>
<td>324.7425</td>
<td>-0.00091513218543</td>
<td>96.17889384</td>
</tr>
<tr>
<td>&quot;CV6&quot;</td>
<td>356.6425</td>
<td>-0.004721562615</td>
<td>22.1454883</td>
</tr>
<tr>
<td>&quot;BPMV6&quot;</td>
<td>356.7425</td>
<td>-0.004729356669</td>
<td>22.27221019</td>
</tr>
<tr>
<td>&quot;CH7&quot;</td>
<td>388.6425</td>
<td>-0.0006096356393</td>
<td>96.67763938</td>
</tr>
</tbody>
</table>
References

[1] The MAD-X Home Page, version February 2003,

[2] W. Herr, MAD for pedestrian, Presentation at CERN Accelerator School

[3] W. Herr, Course on optics design,
http://cern.ch/werner.herr/COURSE/.

[4] O. Brüning and W. Herr, Problems and solutions of the exercises in the optics course,

Participants

AUCHMANN, B. CERN, Geneva, CH
BAEHR, J. DESY-Zeuthen, Zeuthen, DE
BELLODI, G. CCLRC, Didcot, GB
BEUTNER, B. DESY, Hamburg, DE
BLAIR, G. Dept of Physics, RHUL, London, GB
BOHNET, I. DESY-Zeuthen, Zeuthen, DE
CALVO, E. CERN, Geneva, CH
CASTRO CARBALLO, M. CERN, Geneva, CH
CROCKFORD, G. CERN, Geneva, CH
DEHN, M. Institut für Kernphysik, Mainz, DE
ELZHOV, A. JINR, Dubna, RU
FABICH, A. CERN, Geneva, CH
FOURIE, D. iThemba LABS, Somerset West, ZA
FRITIOFF, T. CERN, Geneva, CH
GOJSKA, A. The Andrzej Soetan Institute, Otwock-Swierk, PL
HAN, J. DESY-Zeuthen, Zeuthen, DE
HAYES, D. Old Dominion University, Norfolk, US
KAIN, V. CERN, Geneva, CH
KRAMER, D. CERN, Geneva, CH
KRASILNIKOV, M. DESY-Zeuthen, Zeuthen, DE
KVARDAKOV, V. The Budker Institute of Nuclear Physics, Novosibirsk, RU
LEUSCHNER, A. DESY, Hamburg, DE
LUDWIG, M. CERN, Geneva, CH
MASLOVA, M. JINR, Dubna, RU
MOKODUWE, P. iThemba LABS, Somerset West, ZA
MURATORI, B. Daresbury Laboratory, Warrington, GB
OPPELT, A. DESY-Zeuthen, Zeuthen, DE
PAOLUZZI, M. CERN, Geneva, CH
PETROSSIAN, B. DESY-Zeuthen, Zeuthen, DE
RESTA LOPEZ, J. BURJASSOT, ES
RIEMANN, S.  DESY-Zeuthen, Zeuthen, DE
SACHWITZ, F.  DESY-Zeuthen, Zeuthen, DE
SAKUMI, A.  CERN, Geneva, CH
SENOCAK, B.  Siegen University, Siegen, DE
SHARMA, A.  CAT, Indore, IN
STAHL, A.  DESY-Zeuthen, Zeuthen, DE
STANCARI, G.  University of Ferrara, Ferrara, IT
STAYKOV, L.  DESY-Zeuthen, Zeuthen, DE
STEFFEN, B.  DESY, Hamburg, DE
STRODL, T.  GSI, Darmstadt, DE
SUOMINEN, P.  University of Jyvaskyla, Jyvaskylan Yliopisto, FI
TAN, J.  CERN, Geneva, CH
TARVAIENEN, O.  University of Jyvaskyla, Jyvaskylan Yliopisto, FI
URSCHUETZ, P.  CERN, Geneva, CH
WEIS, M.  Institut für Kernphysik, Mainz, DE
WIN, S.  KEK, Ibaraki, JP
WINTER, A.  III. Physikalisches Inst. RWTH Aachen, Aachen, DE