CERN Yellow Reports: School Proceedings
Volume 5/2018

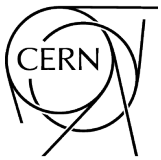
CERN-2018-008-SP



Beam Injection, Extraction and Transfer

Erice, Italy
10 – 19 March 2017

Editor: B. Holzer



CERN Yellow Reports: School Proceedings
Published by CERN, CH-1211 Geneva 23, Switzerland

ISBN 978–92–9083–508–0 (paperback)
ISBN 978–92–9083–509–7 (PDF)
ISSN 2519-8041 (Print)
ISSN 2519-805X (Online)
DOI https://doi.org/10.23730/CYRSP-2018-005

Accepted for publication by the CERN Report Editorial Board (CREB) on 18 December 2018

Available online at https://e-publishing.cern.ch/ and https://cds.cern.ch/

Copyright © CERN, 2018, unless otherwise specified

Creative Commons Attribution 4.0

Knowledge transfer is an integral part of CERN's mission.
CERN publishes this volume Open Access under the Creative Commons Attribution 4.0 license (http://creativecommons.org/licenses/by/4.0/) in order to permit its wide dissemination and use. The submission of a contribution to a CERN Yellow Reports series shall be deemed to constitute the contributor's agreement to this copyright and license statement. Contributors are requested to obtain any clearances that may be necessary for this purpose.

This volume is indexed in: CERN Document Server (CDS), INSPIRE, Scopus.

This report should be cited as:

Proceedings of the CAS–CERN Accelerator School: Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018). https://doi.org/10.23730/CYRSP-2018-005

A contribution in this volume should be cited as:

[Author name(s)], in Proceedings of the CAS–CERN Accelerator School:Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018), pp. [first page] – [lastpage], https://doi.org/10.23730/CYRSP-2018-005. [first page]

Abstract

These proceedings on the special topic of beam injection, extraction and transfer, summarise a series of lectures which was held in the frame of the CERN Accelerator School, CAS, in Erice 2017. Following introductory talks on electromagnetism, relativity and the basics of beam dynamics, different injection and extraction schemes are presented. Special magnetic and electrostatic elements for the case of lepton- and hadron-beams are explained and examples of the technical layout of state of the art kicker and septa designs are discussed. The programme was completed with seminars and lectures on more sophisticated topics related to stripping injection and the special features of resonant extraction, as used e.g. in machines for medical applications.



Preface

The CERN Accelerator School -CAS- has the mission to preserve and disseminate the knowledge accumulated at CERN and elsewhere on particle accelerators and storage rings. In a series of lectures the standard topics usually are covered on an introductory and advanced level in alternating order. In addition, certain topics, relevant for the field, are treated in special schools to follow the technological developments as well as the need for education at places where new facilities were planned or commissioned.

Over the past decades, accelerators and storage rings have been designed and commissioned on an ever increasing level of beam quality and machine performance. New scientific findings went hand in hand with the continuous effort to increase the performance of the particle accelerators.

We have learned to optimise the beam brightness in synchrotron light sources as well as in high energy colliders, and the harvest is impressive: the standard model that proves its solidity more than ever, including the recent discovery of the Higgs particle at LHC in 2012.

A pre-requisite and condition "sine-qua-non" for the successful operation of any particle accelerator is the beam quality, usually expressed in the keyword of emittance. And here, the beam injection, extraction and transfer as a most critical part, deserves special attention. Careful matching of the beam between the pre-accelerator and the synchrotron light source or the collider ring is needed in order to maintain the high quality of the particle beam. Special care has to be applied for the design and operation of the fast acting machine elements, like kickers and septa, as here smallest tolerances already can lead to significant degradation of the beam quality.

Without doubt these machine elements deserve special treatment and so it was only natural to organise a lecture series, dedicated to this special topic in order to bring the expertise in this challenging task together and to train young people. In the present proceedings, a number of lectures on particle beam dynamics and electromagnetism are introducing into the field. They are followed by an overview of the different principles of extraction and injection techniques, for both hadron and lepton beams. In parallel to the theoretical aspects the design principles of the different devices are explained and examples presented for state of the art technical solutions.

We are convinced that these proceedings give a broad overview about the topic with examples from present day machines and we hope that they will help new-comers as well as experienced colleagues to find some useful and interesting information.

We would like to take the opportunity to thank all lecturers for the enormous effort of preparing, presenting, and writing up their topics and the motivation to share their knowledge with the younger colleagues. Very special thanks must go to the enthusiasm and positive feedback of the participants, the students, that attended the school in an impressive number, for their motivation.

As always, the backing of the CERN management, the guidance of the CAS Advisory and Programme Committees and the CERN Publishing Service for their dedication and commitment to the production of this document deserve to be mentioned.

Finally, and last but not least, we want to express our gratitude to our CAS secretary, Barbara Strasser, for her year-long engagement in all issues of the CAS. As always, she ensured that the school was held under optimum conditions. With her passing, we have lost a good friend, a very professional colleague and a precious support for CAS.

Bernhard Holzer, Editor
Organiser and Chair
of the CERN Accelerator School
on Beam Injection, Extraction and Transfer

PROGRAMME FOR BEAM INJECTION, EXTRACTION AND TRANSFER SCHOOL
Erice, Sicily, 10-19 March, 2017

Time	Friday 10 March	Saturday 11 March	Sunday 12 March	Monday 13 March	Tuesday 14 March	Wednesday 15 March	Thursday 16 March	Friday 17 March	Saturday 18 March	Sunday 19 March
09:00		Opening Talks	Overview of Injection & Extraction Techniques	Timing, Synchronisation & Longitudinal Aspects I	Kicker Magnets I		Emittance Preservation	Beam Instrumentation	Transfer Line Design, Matching and Design Tools II	
09:50			B. Goddard	H. Damerau	M. Barnes		V. Kain	P. Forck	W. Bartmann	
10:00	A	Review of Transverse Beam Dynamics I	Injection: Electron Beams	Magnets and Special Magnets	Kicker Magnets II		Septa I	Secondary and Radioactive Beams	Exotic Injection and Extraction Methods	D E P A R T U R E
10:50	R	B. Holzer	M. Aiba	C. Muehle	M. Barnes	E	M. Paraliev	H. Weick	B. Goddard	
11:00	R	COFFEE	COFFEE	COFFEE	COFFEE	X	COFFEE	COFFEE	COFFEE	
11:30	I	Review of Longitudinal Beam Dynamics	Injection: Hadron Beams	Timing, Synchronisation & Longitudinal Aspects II	Single and Multi-Turn Extraction	C	Septa II	Low Energy Beam Transport	Transfer Line Design, Matching and Design Tools III	
12:20	V	F. Tecker	C. Bracco	H. Damerau	M. Fraser	U	M. Paraliev	R. Baartman	W. Bartmann	
12:30	A	LUNCH	LUNCH	LUNCH	LUNCH	R	LUNCH	LUNCH	LUNCH	
14:30	L	Review of Transverse Beam Dynamics II	Phase Space Painting and H-Stripping Injection		Resonant Extraction	S	Optics Measurement Techniques for Transfer Lines I	Transfer Line Design, Matching and Design Tools I	**Seminar** Injection into a Trap	
15:20	D	B. Holzer	C. Bracco	F R E E	P. Bryant	I O N	P. Forck	W. Bartmann	F. Herfurth	
15:30	A	Particle Interactions with Matter	Introduction to Case Study		**Seminar** Extraction Techniques for Medical Beams		Optics Measurement Techniques for Transfer Lines II	Machine Protection and Activation	Case Study Presentations	
16:20	Y	A. Lechner		A F T E R N O O N	P.Bryant		P.Forck	A. Nordt		
16:30		TEA	TEA		TEA		TEA	TEA	TEA	
17:00		Recap. of Electromagnetism	Injection and Extraction in Cyclotrons		Case Study		Case Study	Case Study Preparation	Case Study Presentation	
17:50		U. VanRienen	M. Seidel							
18:00		Recap. Of Relativity and Space Charge	Case Study		Case Study		Case Study	Case Study Preparation	Outlook: Physics beyond the Standard Model	
18:50		M. Ferrario								
20:00		DINNER	DINNER	DINNER	DINNER		DINNER	Special Dinner	DINNER	

Contents

Preface
B. Holzer v

Transverse Beam Dynamics
B. Holzer 1

Longitudinal Beam Dynamics—Recap
F. Tecker 31

Particle Interactions with Matter
A. Lechner 47

Recapitulation of Electromagnetism
T. Flisgen et al. 69

Space Charge Mitigation
M. Ferrario 89

Introduction to Beam Transfer
B. Goddard 109

Injection: Electron Beams
M. Aiba 121

Injection: Hadron Beams
C. Bracco 131

Phase Space Painting and Hn^- Stripping Injection
C. Bracco 141

Injection and Extraction in Cyclotrons
M. Seidel 151

Timing, Synchronization and Longitudinal Aspects
H. Damerau 163

Magnets and Special Magnets
C. Muehle 193

Kicker Systems
M. J. Barnes 229

Fast Extraction: Single and Multiturn
M. A. Fraser 285

Resonant Extraction
P. J. Bryant 315

Emittance Preservation
V. Kain 345

Septa
M. Paraliev 363

Measurement Techniques for Transfer Lines and Beam Instrumentation
P. Forck 395

Low-Energy Beam Transport Design
R. Baartman 491

Transfer Lines
W. Bartmann 507

Machine Protection and Activation
A. Nordt et al. 533
Exotic Injection and Extraction Methods
B. Goddard 553
List of Participants 567

Proceedings of the CAS–CERN Accelerator School: Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018)

Transverse Beam Dynamics

B.J. Holzer
CERN, Geneva, Switzerland

Abstract
This paper gives an overview of transverse dynamics in particle accelerators with—following the emphasis of the school—special focus on transfer lines and their application for beam injection and extraction. We will restrict ourselves to a bare introduction to the basic concepts that are needed to understand the boundary conditions to be met when we are trying to transfer a beam of charged particles from one accelerator to the next.

Keywords
Accelerator physics; transverse dynamics; transfer line; injection; extraction.

1 Introduction

If we contemplate a little on the topic of this school, we get the impression that we have to treat the problem of transferring particles from one accelerator to the next, including the quite tricky business of injecting or extracting the beam into (or out of), for example, a storage ring. Now while this involves some beautiful technical concepts, such as special magnets, kickers, and septa, the overall picture is quite a bit more than just pushing a button and getting off to beddy-byes and that's it. We have to optimize the system in such a way as to get the ideal orbit position in both transverse planes (we call them x and y for the horizontal and vertical amplitudes) and we have to get the ideal corresponding angles, x' and y'. But in addition, we have to take care that the properties of the complete particle ensemble are maintained. In particular, the beam quality—we call it the *emittance* and we will talk about it in quite some detail—has to be preserved in order not to destroy the particle density distribution during the beam transfer, where fast-acting fields are applied. Unlike any other field changes in an accelerator, the injection and extraction elements are non-adiabatic; this means that their impact on the beam is *fast* and special care has to be taken.

In this sense, this paper on transverse beam dynamics, together with the accompanying papers on longitudinal dynamics, defines the background and the language for the particular courses in this school. To be very clear from the beginning: the language that we will use, concerning transverse beam dynamics, is usually applied and valid only for periodic structures; namely synchrotrons and storage rings. The problems that are studied there are typically of the form:

> *How do we define the orbit of the particles in an accelerator, how do we bend the particles onto a ring, and how do we manage to keep them on—or close to—a trajectory with deviations from this ideal path of only a fraction of a millimetre?*

While strictly (i.e., mathematically) spoken, this language is valid for these periodic cases only, it is so powerful and elegant that we also like to use it in linear accelerators (even linear colliders, in some cases) and in transfer lines. Thus, we will use concepts that have been developed for 'periodic structures', i.e., where the situation seen by the particle repeats itself after a certain time or distance. The reason lies at the bot of the mathematics involved. Now, while this condition is easily fulfilled in a circular accelerator, namely in the case of a synchrotron, a transfer line does not have such a built-in periodicity. The particles are passing through the structure only once and that's it. However, we still like to make use of the expressions derived for the periodic machines, as they present a nifty tool for designing the accelerator as well as expressing the most relevant beam parameters. Even more, the injection and

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-005.1

extraction conditions have to meet the optical parameters of the connected accelerators, which are indeed periodic.

So transferring a particle beam from one accelerator to another one means the following.

- We need a system of fast-acting magnetic or electric fields that create bending forces to inject or extract the beam (or a number of bunches).
- We have to install magnetic bending fields to guide the particles in the transfer line from one place to the next.
- We have to use focusing forces (quadrupole lenses) to keep the particle ensemble close together and match the optical conditions defined by the pre- and post-accelerator at both ends of the line.
- Finally, we have to take care of the timing, i.e., the right moment of the beam transfer. We have to trigger our transfer elements in such a way that the particles, running down the transfer line, will arrive in the new machine (usually) in the centre of the new RF bucket. This last item, however, is beyond our scope here and will be treated in the following paper on longitudinal dynamics.

2 Transverse beam dynamics

The transverse beam dynamics of charged particles in an accelerator describes the movement of single particles under the influence of external transverse bending and focusing fields. It includes the detailed arrangement (for example, their positions in the machine and their strengths) of the accelerator magnets used to obtain well-defined, predictable parameters of the stored particle beam, and it describes methods to optimize the trajectories of single particles, as well as the dimensions of the beam, considered as an ensemble of many particles. A detailed treatment of this field in full mathematical detail, including sophisticated lattice optimizations, such as the right choice of the basic lattice cells and the design of dispersion suppressors or chromaticity compensation schemes, is beyond the scope of this overview. For further reading and for more detailed descriptions, we therefore refer to the more complete explanations in Refs. [1–3]. For the time being, we will just give a basic introduction to the topic and explain with –more or less hand-waving—how the trick goes.

2.1 Geometry of the ring

In general, magnetic fields are used in circular accelerators to provide the bending force and to focus the particle beam. In principle, the use of electrostatic fields would also be possible, 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

$$\mathbf{F} = q \cdot (\mathbf{E} + \mathbf{v} \times \mathbf{B}). \tag{1}$$

For high-energy particle beams, the velocity v is close to the speed of light and so represents a nice amplification factor whenever we apply a magnetic field. As a consequence, it is much more convenient to use magnetic fields to bend and focus the particles.

Therefore, neglecting electric fields for the moment, we write the Lorentz force and the centrifugal force on the particle on its circular path as

$$F_{\text{Lorentz}} = e \cdot v \cdot B \,, \tag{2}$$

$$F_{\text{centrifugal}} = \frac{\gamma m_0 v^2}{\rho} \,. \tag{3}$$

Assuming an idealized homogeneous dipole magnet along the particle orbit, having pure vertical field lines, we define the condition for a perfect circular orbit as equality between these two forces. This yields

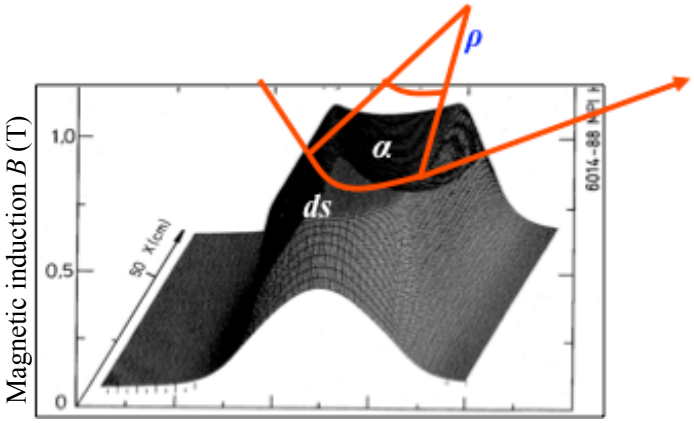


Fig. 1: Field map of a storage ring dipole magnet, and schematic path of a particle

the following condition for the idealized ring:

$$\frac{p}{e} = B \cdot \rho \,, \tag{4}$$

where we are referring to protons and have accordingly set $q = e$. This condition relates the so-called beam rigidity $B\rho$ to the momentum of a particle that can be carried, e.g., in a storage ring, and it ultimately defines, for a given magnetic field, the bending angle that we can achieve if a dipole field is applied to bend the beam in a transfer line or to extract it from a pre-accelerator. In reality, instead of having a continuous dipole field, the structure of our accelerator or transfer line (we call it the *lattice*) will consist of several dipole magnets, usually powered in series to define the geometry of the ring.

Now, for completeness, we would like to add two more drops of information. Clearly enough, the same argument that we used for the effect of magnetic fields is also true in the case of electrostatic fields. Again, the effect depends on the inverse of the momentum and we get e.g., for the bending angle in an electric field

$$\theta_{\mathrm{el}} = \frac{El}{p\beta} \,.$$

The second comment to make here is that as soon as we talk about *low*-energy particles, it might turn out that magnetic fields are not the ideal choice. They will have a small effect on the beam as soon as the particle velocity is low. Then the accelerator components might better be electrostatic devices instead of magnets; they are also usually smaller and could be the preferred choice if the available space is limited.

Coming back to the more common situation of magnets as bending and focusing elements, we sketch the trajectory of a single particle in Fig. 1. In the free space outside the dipole magnet, the particle trajectory follows a straight line. As soon as the particle enters the magnet, it is bent onto a circular path until it leaves the magnet at the other side.

The overall effect of the main bending (or 'dipole') magnets in the ring is to define a more or less circular path, which we call the 'design orbit'. By definition, this design orbit has to be a closed loop, and so the main dipole magnets in the ring have to define a full bending angle of exactly 2π. If α denotes the bending angle of a single magnet, then

$$\alpha = \frac{\mathrm{d}s}{\rho} = \frac{B\,\mathrm{d}s}{B \cdot \rho} \,. \tag{5}$$

We therefore require that, integrating over all dipole magnets, we get

$$\frac{\int B\,\mathrm{d}s}{B \cdot \rho} = 2\pi \,. \tag{6}$$



Fig. 2: Superconducting dipole magnet in the LHC storage ring

Thus, a storage ring or synchrotron is not a 'ring' in the true sense of the word but more a polygon, where 'poly' means the discrete number of dipole magnets installed in the 'ring'.

In the extreme case of the LHC, the dipole field has been pushed to the highest achievable values: 1232 superconducting dipole magnets, each 15 m long, define the geometry of the ring (or better 1232-gon, whatever the Greek expression for that might be), and, via Eq. (6), they determine the maximum momentum for the stored proton beam. Using these equations, for a maximum momentum $p = 7$ TeV/c, we obtain a required magnetic field of

$$B = \frac{2\pi \cdot 7000 \cdot 10^9 \text{ eV}}{1232 \cdot 15 \text{ m} \cdot 2.99792 \cdot 10^8 \text{ m s}^{-1}} , \tag{7}$$

or

$$B = 8.33 \text{ T} , \tag{8}$$

to bend the beams. For convenience, we have expressed the particle momentum in units of GeV/c here. Figure 2 shows a photograph of one of the LHC dipole magnets, built with superconducting NbTi filaments, which are operated at a temperature $T = 1.9$ K.

2.2 Focusing properties

In addition to the main bending magnets that guide the beam onto a closed orbit, focusing fields are needed to keep the particles close together. In modern storage rings and light sources, we have to keep more than 10^{12} particles in the machine, distributed over a number of bunches, and these particles have to be focused to keep their trajectories close to the design orbit. Furthermore, they are stored in the machine for many hours, and a carefully designed focusing structure is needed to maintain the necessary beam size in the ring and guarantee stability of the transverse motion.

Following classical mechanics, linear restoring forces are used, just as in the case of a harmonic pendulum. Quadrupole magnets provide the corresponding field property: they create a magnetic field that depends linearly on the amplitude of the particle, i.e., the distance of the particle from the design orbit:

$$B_x = -g \cdot y, \qquad B_y = -g \cdot x . \tag{9}$$

The constant g is called the gradient of the magnetic field and it characterizes the focusing strength of the quadrupole lens in both transverse planes. The minus sign is a convention that follows the fact that for a positive amplitude, the field configuration of a focusing quadrupole will lead to a Lorentz force that reduces this amplitude, according to Fig. 3. As in the case of the dipole field, the quadrupole gradient

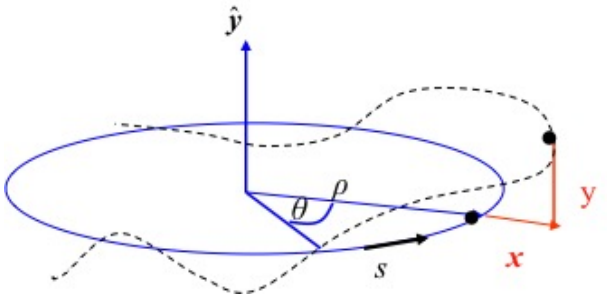


Fig. 3: Co-ordinate system used in particle beam dynamics; the longitudinal co-ordinate s moves around the ring with the particle considered.

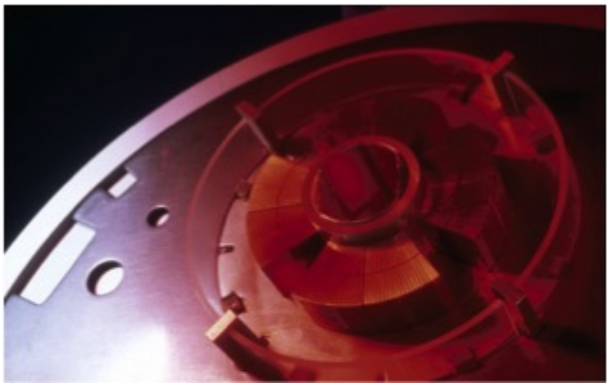


Fig. 4: Superconducting quadrupole magnet in the LHC storage ring

is usually normalized to the particle momentum to obtain expressions that are valid for any particle momentum or energy. This normalized gradient is denoted k and defined as

$$k = \frac{g}{p/e} = \frac{g}{B\rho} \,. \tag{10}$$

The technical layout of such a quadrupole—again we use the LHC design as an example—is depicted in Fig. 4. As in the case of the dipoles, the LHC quadrupole magnets were built using superconducting technology to achieve the highest possible focusing forces.

Now that we have defined the two basic building blocks of a storage ring, we need to arrange them in a so-called magnet lattice and optimize the field strengths in such a way as to obtain the required beam parameters. An example to show how such a magnet lattice looks like in a real storage ring is given in Fig. 5. This photograph shows the dipole (orange) and quadrupole (red) magnets in the TSR storage ring in Heidelberg [4]. Eight dipoles are used to bend the beam into a 'circle', and the quadrupole lenses between them provide the focusing to keep the particles within the aperture limits of the vacuum chamber.

A general design principle of modern synchrotrons and storage rings should be pointed out here. In general, these machines are built following a so-called separate-function scheme: every magnet is designed and optimized for a certain task, such as bending, focusing, or chromatic correction. We separate the magnets in the design according to the job they are supposed to do; only in rare cases is a combined-function scheme chosen nowadays, where different magnet properties are combined in one piece of hardware. To express this principle mathematically, we use the general Taylor expansion of the normalized magnetic field,

$$\frac{B(x)}{p/e} = \frac{1}{\rho} + k \cdot x + \frac{1}{2!} m x^2 + \frac{1}{3!} n x^3 + \cdots \,, \tag{11}$$

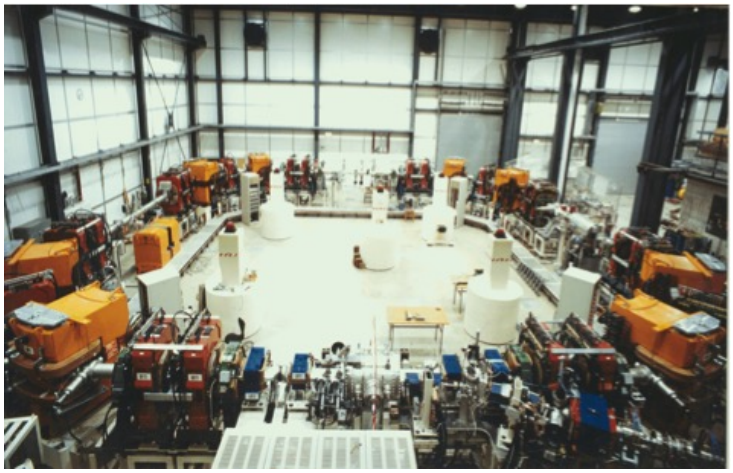


Fig. 5: The TSR storage ring, Heidelberg, is a typical example of a separate-function strongly focusing storage ring [4].

where $1/\rho$ refers to the normalized dipole component, k to the quadrupole contribution, and, n, m, etc., to the field contributions from the higher-order multipoles. Following these arguments, for the moment we take into account only constant (dipole) or linear (quadrupole) terms. The higher-order contributions to the field will be treated later as (hopefully) small perturbations.

Under these assumptions, we can derive—in linear approximation—the equation of motion of the transverse particle movement. We start with a general expression for the radial acceleration, known from classical mechanics (see, e.g., Ref. [5]):

$$a_r = \frac{\mathrm{d}^2\rho}{\mathrm{d}t^2} - \rho\left(\frac{\mathrm{d}\theta}{\mathrm{d}t}\right)^2 . \tag{12}$$

The first term refers to an explicit change in the bending radius, and the second to the centrifugal acceleration. Referring to our co-ordinate system, and replacing the ideal radius ρ with $\rho+x$ for the general case (Fig. 3), we obtain the relation for the balance between the radial force and the counteracting Lorentz force:

$$F = m\frac{\mathrm{d}^2}{\mathrm{d}t^2}(x+\rho) - \frac{mv^2}{x+\rho} = evB . \tag{13}$$

On the right-hand side of the equation, we take only linear terms of the magnetic field into account,

$$B_y = B_0 + x\frac{\mathrm{d}B_y}{\mathrm{d}x} , \tag{14}$$

and for convenience we replace the independent variable t with the co-ordinate s,

$$x' = \frac{\mathrm{d}x}{\mathrm{d}s} = \frac{\mathrm{d}x}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}s} , \tag{15}$$

'Convenience' in this context means that we are more interested in the amplitude x and angle x' of the particle trajectory and therefore prefer the derivative of x with respect to s. Thus, we obtain an expression for the particle trajectories under the influence of the focusing properties of the quadrupole and dipole fields in the ring, described by a differential equation. This equation is derived in its full beauty elsewhere [3], so we shall just state it here:

$$x'' - x\cdot\left(k - \frac{1}{\rho^2}\right) = 0 , \tag{16}$$

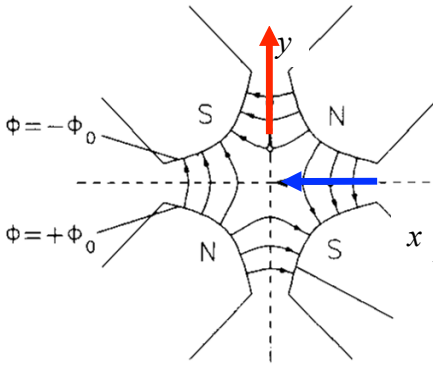


Fig. 6: Field configuration in a quadrupole magnet and the direction of the focusing and defocusing forces in the horizontal and vertical planes.

where k is the normalized gradient introduced in Eq. (11) and the $1/\rho^2$ term represents so-called weak focusing, which is a property of the bending magnets. Depending on the actual sign of k, the quadrupole will focus (negative sign) or defocus (positive sign) the beam in the corresponding plane. The particles will now follow the 'circular' path defined by the dipole fields and, while they are running around the machine on this path, they will, in addition, undergo harmonic oscillations in both transverse planes. The situation is shown schematically in Fig. 3. An ideal particle will follow the design orbit represented by the circle in the diagram. Any other particle will perform transverse oscillations under the influence of the external focusing fields, and the amplitude of these oscillations will ultimately define the beam size. To be brief, and referring to the horizontal plane for a moment, we can make the statement that: *'under the influence of the focusing fields from the quadrupoles k and dipoles $1/\rho^2$, the transverse movement of the particles inside the single lattice elements looks like a harmonic oscillation'*.

Unlike the case of a classical harmonic oscillator, however, the equations of motion in the horizontal and vertical planes differ somewhat. Assuming a horizontal focusing magnet, the equation of motion is as shown in Eq. (16). In the vertical plane, however, because of the orientation of the field lines and thus—in the end—by Maxwell's equations, the forces instead have a defocusing effect. Also, the weak focusing term disappears in general:

$$y'' + y \cdot k = 0 \,. \tag{17}$$

The principal problem arising from the different directions of the Lorentz force in the two transverse planes of a quadrupole field is sketched in Fig. 6. As a consequence, to overcome this uncomfortable situation, we have to explicitly introduce quadrupole lenses that focus the beam in the horizontal and vertical directions in some alternating order. It is the task of the machine designer to find an adequate solution to this problem and to define a magnet pattern that will provide an overall focusing effect in both transverse planes. The rest is easy, in the sense of A. Wolski's statement: "...in principle, there are only two steps in the analysis of any dynamical system. The first step is to write down the equations of motion; and the second step is to solve them" [6].

Now, following closely the example of the classical harmonic oscillator, we can write down the solutions of the equations of motion. For simplicity, we focus on the horizontal plane; a 'focusing' magnet is therefore focusing in this horizontal plane and at the same time defocusing in the vertical plane. Starting with the initial conditions for the particle amplitude x_0 and angle x_0' in front of the magnet element, we obtain the following relations for the trajectory inside the magnet:

$$x(s) = x_0 \cdot \cos\left(\sqrt{|K|}\, s\right) + x_0' \cdot \frac{1}{\sqrt{|K|}} \sin\left(\sqrt{|K|}\, s\right), \tag{18}$$

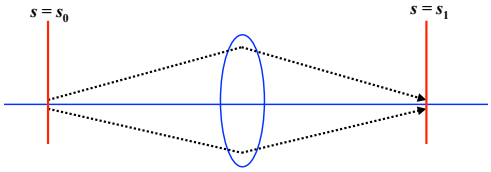


Fig. 7: Effect of a focusing quadrupole magnet

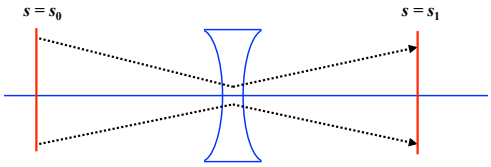


Fig. 8: Effect of a defocusing quadrupole magnet

$$x'(s) = -x_0 \cdot \sqrt{|K|}\sin\left(\sqrt{|K|}\,s\right) + x'_0 \cdot \cos\left(\sqrt{|K|}\,s\right) . \tag{19}$$

Here, the parameter K combines the quadrupole gradient and the weak focusing effect: $K := 1/\rho^2 - k$. Usually, these two equations are combined into a more elegant and convenient matrix form,

$$\begin{pmatrix} x \\ x' \end{pmatrix}_s = \mathbf{M}_{\rm foc} \begin{pmatrix} x \\ x' \end{pmatrix}_0 , \tag{20}$$

where the matrix $\mathbf{M}_{\rm foc}$ contains all the relevant information about the magnet element:

$$\mathbf{M}_{\rm foc} = \begin{pmatrix} \cos\left(\sqrt{|K|}\,s\right) & \frac{1}{\sqrt{|K|}}\sin\left(\sqrt{|K|}\,s\right) \\ -\sqrt{|K|}\sin\left(\sqrt{|K|}\,s\right) & \cos\left(\sqrt{|K|}\,s\right) \end{pmatrix} . \tag{21}$$

The situation is illustrated in Fig. 7.

In the case of a defocusing magnet (or to be quite clear, also, in the case of the vertical plane, of a horizontal focusing magnet), we obtain analogously that

$$\begin{pmatrix} x \\ x' \end{pmatrix}_s = \mathbf{M}_{\rm defoc} \begin{pmatrix} x \\ x' \end{pmatrix}_0 , \tag{22}$$

with

$$\mathbf{M}_{\rm defoc} = \begin{pmatrix} \cosh\left(\sqrt{|K|}\,s\right) & \frac{1}{\sqrt{|K|}}\sinh\left(\sqrt{|K|}\,s\right) \\ \sqrt{|K|}\sinh\left(\sqrt{|K|}\,s\right) & \cosh\left(\sqrt{|K|}\,s\right) \end{pmatrix} ; \tag{23}$$

see Fig. 8.

For completeness, we also include the case of a field-free drift. With $K = 0$, we obtain

$$\mathbf{M}_{\rm drift} = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} . \tag{24}$$

This matrix formalism allows us to combine the elements of a storage ring in an elegant way, and so it is straightforward to calculate particle trajectories. In this context, we would like to emphasize a few points:

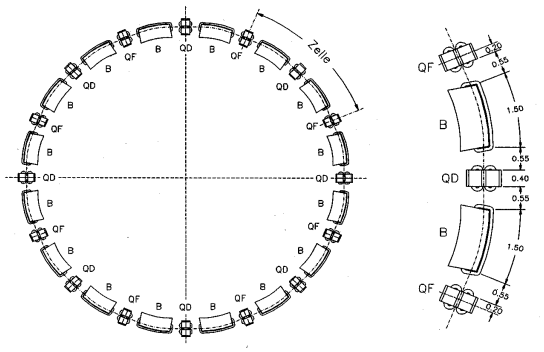


Fig. 9: Simple periodic chain of bending magnets (B) and focusing (QF) or defocusing (QD) quadrupoles forming the basic structure of a storage ring. Courtesy, K. Wille [1].

- a certain quadrupole lens will always have two opposing effects: focusing in one plane and defocusing in the other;
- et vice versa (*'and the other way round'*, for the non-Latin-speaking community);
- we therefore like to combine these effects in a common '4×4 matrix';
- in linear approximation and without explicit coupling fields, such as roll angles of the quadrupoles or solenoids, the motion in the two transverse planes is uncoupled. An amplitude in the horizontal direction, e.g., will not have any influence on the vertical motion and therefore the corresponding non-diagonal elements of the matrix $M_{1,3}$, $M_{1,4}$, etc., are zero.

It is therefore convenient to describe this simultaneous effect in the two planes in a single 4×4 matrix and define a vector for both transverse amplitudes and angles:

$$\begin{pmatrix} x \\ x' \\ y \\ y' \end{pmatrix}_s = \begin{pmatrix} \cos\left(\sqrt{|K|}\,s\right) & \frac{1}{\sqrt{|K|}}\sin\left(\sqrt{|K|}\,s\right) & 0 & 0 \\ -\sqrt{|K|}\sin\left(\sqrt{|K|}\,s\right) & \cos\left(\sqrt{|K|}\,s\right) & 0 & 0 \\ 0 & 0 & \cosh\left(\sqrt{|K|}\,s\right) & \frac{1}{\sqrt{|K|}}\sinh\left(\sqrt{|K|}\,s\right) \\ 0 & 0 & \sqrt{|K|}\sinh\left(\sqrt{|K|}\,s\right) & \cosh\left(\sqrt{|K|}\,s\right) \end{pmatrix} \cdot \begin{pmatrix} x \\ x' \\ y \\ y' \end{pmatrix} \tag{25}$$

As an example of a larger structure, we consider the simple case of an alternating focusing and defocusing lattice, a so-called FODO lattice, see Fig. 9.

As we know the properties of each and every element in the accelerator, we can construct the corresponding matrices and calculate, step by step, the amplitude and angle of a single-particle trajectory around the ring. Even more conveniently, we can multiply out the different matrices and, given initial conditions x_0 and x_0' at a certain position in the storage ring, obtain directly the trajectory at any location in the ring:

$$\mathbf{M}_{\text{total}} = \mathbf{M}_{\text{foc}} \cdot \mathbf{M}_{\text{drift}} \cdot \mathbf{M}_{\text{dipole}} \cdot \mathbf{M}_{\text{drift}} \cdot \mathbf{M}_{\text{defoc}} \cdots . \tag{26}$$

The trajectory thus obtained is shown schematically in Fig. 10.

We have to point out the following facts in this context.

- At each moment, which means inside each lattice element, the trajectory is part of a harmonic oscillation.
- However, because of the different restoring or defocusing forces, the solution will look different at each location.
- In the linear approximation that we have used in this context, all particles experience the same external fields, and their trajectories will differ only because of their different initial conditions.

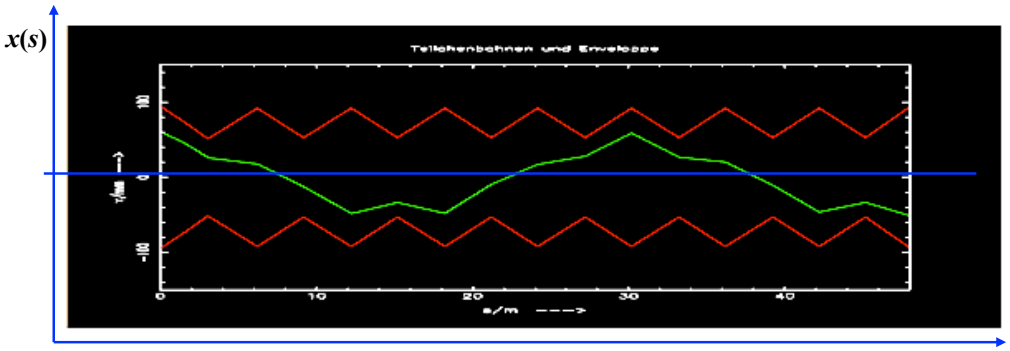


Fig. 10: Calculated particle trajectory in a simple storage ring

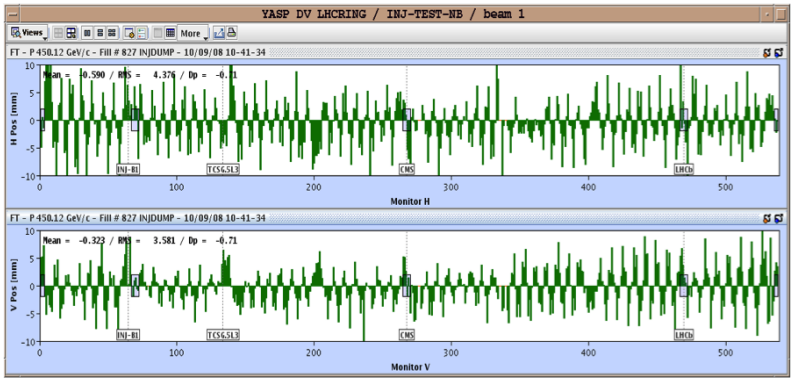


Fig. 11: Beam position measured at the injection septum and after the first turn around the machine during one of the very first beam injections into the LHC storage ring.

– There seems to be an overall oscillation in both transverse planes while the particle is travelling around the ring. Its amplitude stays well within the boundaries set by the vacuum chamber, and its frequency in the example of Fig. 10 is roughly 1.4 transverse oscillations per revolution, which corresponds to the eigenfrequency of the particle under the influence of the external fields.

Coming closer to a real, existing machine, Fig. 11 shows an orbit, measured during one of the first injections into the LHC storage ring. The horizontal oscillations are plotted in the upper half of the figure and the vertical oscillations in the lower half, on a scale of ± 10 mm. Each histogram bar indicates the value recorded during the first turn of the beam by a beam position monitor at a certain location in the ring; the orbit oscillations are clearly visible. During these first injections, a beam screen had been introduced right after the injection point. In Fig. 12, the spot of the injected beam is clearly visible as well as the one after the first turn. In both transverse planes, these spots are not yet lying on top of each other and so the orbit is not yet closed. Here we have to emphasize very clearly that until this moment the storage ring 'LHC' is a simple transfer line. The conditions of horizontal and vertical orbit amplitude at the injection point and after one turn are not the same. While this seems to have a small effect on the orbit, it makes a large difference in the mathematical treatment and for the matter of resonances that can amplify orbit fluctuations. We will come back to this problem in a later section. However, in the case of Fig. 12, after a straightforward orbit correction, the closed orbit condition can be achieved and we finally obtain what we call a 'stored beam'.

By counting (or, better, fitting) the number of oscillations in both transverse planes, we obtain, in the case of the LHC, values of

$$Q_x = 64.31, \qquad Q_y = 59.32. \tag{27}$$

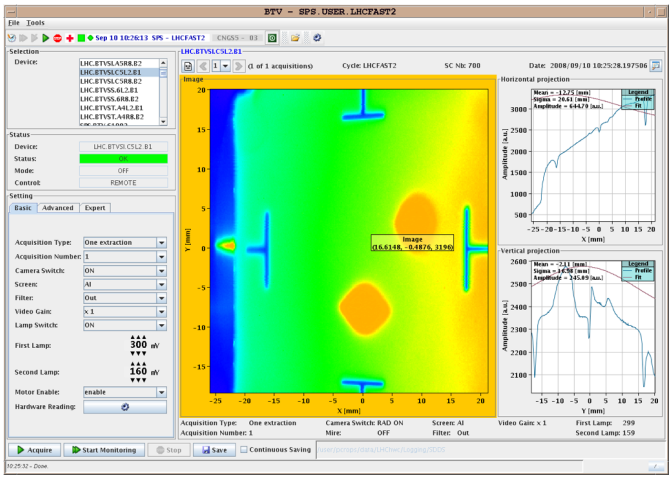


Fig. 12: Measured orbit of the first turn in LHC during the commissioning of the machine. The beam screen is located immediately after the injection septum and shows the spot at injection and after one full turn around the machine.

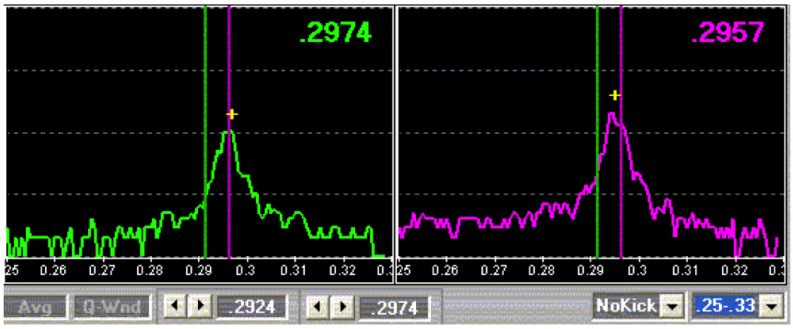


Fig. 13: Tune signal of proton beam in storage ring (HERA-p)

These values, which describe the eigenfrequencies of the particles, are called the horizontal and vertical *tunes*, respectively. Knowing the revolution frequency, we can easily calculate the transverse oscillation frequencies, which correspond to these tune values and which, for this type of machine, usually lie in the range of some hundreds of kilohertz.

As the tune characterizes the particle oscillations under the influence of all external fields, it is one of the most important parameters of a storage ring. Therefore, it is usually displayed and controlled at all times by the control system of such a machine. As an example, Fig. 13 shows the tune diagram of the HERA proton ring [7]; this was obtained via a Fourier analysis of a spectrum measured from the signal of the complete particle ensemble. The peaks indicate the two tunes in the horizontal and vertical planes of the machine; in a sufficiently linear machine, a fairly narrow spectrum is obtained.

Briefly referring back to Fig. 10, the question is what the trajectory of the particle will look like in the second turn, or the third, or after an arbitrary number of turns. Now, as we are dealing with a circular machine, the amplitude x and angle x' at the end of the first turn will be the initial conditions for the second turn, and so on. After many turns, the overlapping trajectories begin to form a pattern, such as that shown in Fig. 14, which indeed looks like a beam that here and there has a larger or a smaller size but still remains well-defined in its amplitude by the external focusing forces.

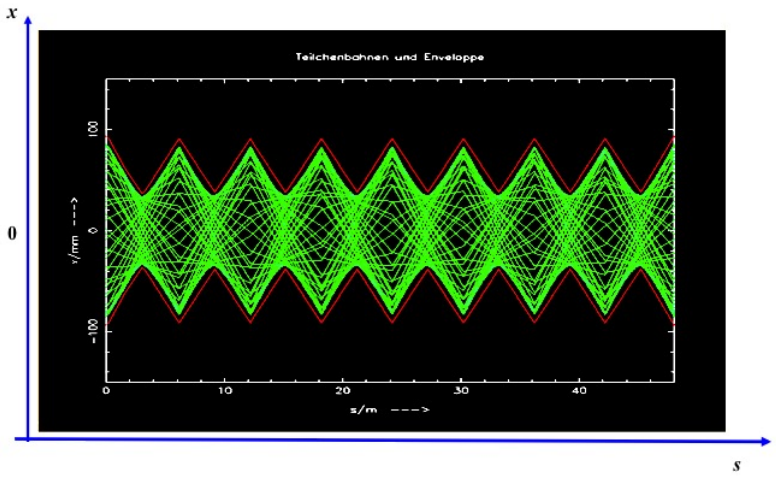


Fig. 14: Many single-particle trajectories together form a pattern that corresponds to the beam size in the ring

3 The Twiss parameters α, β, and γ

As explained in the last section, repeating the calculations that led to the orbit of the first turn will result in a large number of single-particle trajectories that overlap in some way and form the beam envelope. Figure 14 shows the result for 50 turns. Clearly, as soon as we are talking about many turns or many particles, the use of the single-trajectory approach is quite limited and we need a description of the beam as an ensemble of many particles. Fortunately, in the case of periodic conditions in the accelerator, there is another way to describe the particle trajectories and, in many cases, it is more convenient than the aforementioned formalism. It is important to note that, in a circular accelerator, the focusing elements are necessarily periodic in the orbit co-ordinate s after one revolution. Furthermore, storage ring lattices have an internal periodicity in most cases: they are often constructed, at least partly, from sequences in which identical magnetic structures, the lattice cells, are repeated several times in the ring and lead to periodically repeated focusing properties. In this case, the equation of motion can now be written in a slightly different form:

$$x''(s) - k(s) \cdot x(s) = 0 \,, \tag{28}$$

where, for simplicity, we refer to a pure quadrupole magnet and so the $1/\rho^2$ term does not appear. The main issue, however, is that unlike the previous treatment, the focusing parameters (or restoring forces) are no longer constant but are functions of the co-ordinate s. However, they are periodic in the sense that, at least after one full turn, they repeat themselves, i.e., $k(s + L) = k(s)$, leading to the so-called Hill differential equation (28). Following Floquet's theorem [2], the solution of this equation can be written in its general form as

$$x(s) = \sqrt{\varepsilon}\sqrt{\beta(s)}\cos(\psi(s) - \phi), \tag{29}$$

where ψ is the phase of the oscillation, ϕ is its initial condition, and ε is a characteristic parameter of a single particle or, if we are considering a complete beam, of the ensemble of particles. Taking the derivative with respect to s, we get the trajectory angle x':

$$x'(s) = \sqrt{\varepsilon}\sqrt{\beta(s)}\left(\frac{1}{2}\beta'(s)\cos\left(\psi(s) - \phi\right) + \sin\left(\psi(s)\right)\right) \,. \tag{30}$$

The position and angle of the transverse oscillation of a particle at a point s are given by the value of a special amplitude function, the β-function, at that location; ε and ϕ are constants of any particular trajectory. The β-function depends in a rather complicated manner on the overall focusing properties of the storage ring. It cannot be calculated directly by an analytical approach, but instead must be either

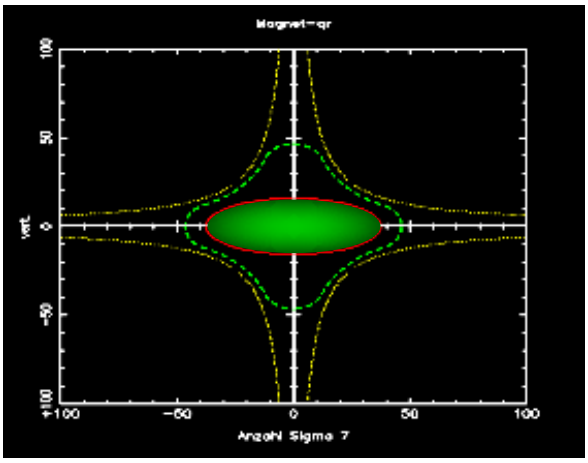


Fig. 15: Transverse beam shape inside a quadrupole magnet: plotted are 7σ of a Gaussian particle density distribution inside the vacuum chamber and magnet aperture.

determined numerically or deduced from properties of the single-element matrices (see, e.g., Ref. [3]). In any case, like the lattice itself, it has to fulfil the periodicity condition

$$\beta(s+L) = \beta(s)\,. \tag{31}$$

Inserting the solution (Eq. (29)) into the Hill equation and rearranging slightly, we get

$$\psi(s) = \int_0^s \frac{\mathrm{d}s}{\beta(s)}\,, \tag{32}$$

which describes the phase advance of the oscillation. It should be emphasized that ψ depends on the particle's oscillation amplitude. At locations where β reaches large values, i.e., the beam has a large transverse dimension, the corresponding phase advance is small; conversely, at locations where we create a small β in the lattice, we obtain a large phase advance. In the context of Fig. 10, we introduced the tune as the number of oscillations per turn, which is nothing other than the overall phase advance of the transverse oscillation per revolution in units of 2π. So, by integrating Eq. (32) around the ring, we get for the tune, the expression

$$Q = \frac{1}{2\pi} \oint \frac{\mathrm{d}s}{\beta(s)}\,. \tag{33}$$

The practical significance of the β-function is shown in Figs. 14 and 15. Whereas in Fig. 14 the single-particle trajectories are plotted turn by turn, Fig. 15 shows schematically a section through the transverse shape of the beam and indicates the beam size inside the vacuum chamber. The hyperbolic profile of the pole shoes of the quadrupole lens is sketched as a yellow dashed line, and the envelope of the overlapping trajectories, given by $\hat{x} = \sqrt{\varepsilon\beta(s)}$, is marked in red and used to define the beam size in the sense of a Gaussian density distribution.

3.1 β, ε, and the phase space ellipses

Although the β-function is a somewhat abstract parameter that results from all focusing and defocusing elements in the ring, the integration constant ε has a well-defined physical interpretation. Given the solution of Hill's equation, Eq. (29), and its derivative, Eq. (30), we can transform the first equation to

$$\cos\big(\psi(s)\big) = \frac{x(s)}{\sqrt{\varepsilon\beta(s)}} \tag{34}$$

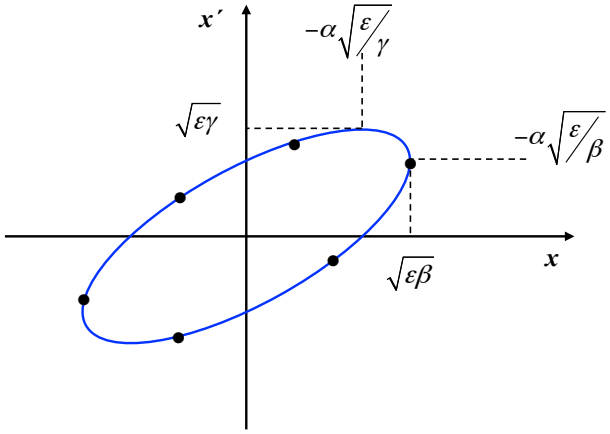


Fig. 16: Ellipse in (x, x') phase space

and insert the expression into Eq. (30) to get an expression for the integration constant ε:

$$\varepsilon = \gamma(s)x^2(s) + 2\alpha x(s)x'(s) + \beta(s)x'^2(s) \; . \tag{35}$$

Here we have followed the usual convention in the literature and introduced the two parameters

$$\alpha(s) = -\frac{1}{2}\beta'(s) \tag{36}$$

and

$$\gamma(s) = \frac{1+\alpha^2(s)}{\beta(s)} \; . \tag{37}$$

On having a closer look at Eq. (35), we realize that we obtain for ε a parametric representation of an ellipse in the (x, x') 'phase space'. Now the mathematical integration constant ε gains physical meaning. In fact, ε describes the space occupied by the particle in the transverse (x, x') phase space (simplified here to a two-dimensional space). More specifically, the area in the (x, x') space that is covered by the particle is given by

$$A = \pi \cdot \varepsilon \; , \tag{38}$$

and, as long as we consider only conservative forces acting on the particle, this area is constant according to Liouville's theorem. Here, we take these facts as given, but we should point out that, as a direct consequence, the so-called emittance ε cannot be influenced by any external fields; it is a property of the beam, and we have to take it as given and act with caution in order not to dilute it.

Here a little side remark is appropriate—at least for the theoretical purists: the real phase space, as we know it from classical mechanics, refers to the canonical conjugate variables, x and p_x. In beam dynamics, however, we keep it a bit simpler and draw the dynamics of our particles in an x–x' co-ordinate system. To point out the difference and to make things clear, sometimes the x–x' space is therefore called the 'trace-space'. However this expression is not (yet) fully accepted in the literature. Still, there are some remarkable consequences; we will come back to this point in the next section.

To be more descriptive, and following the usual textbook treatment of accelerators, we can draw the ellipse of the particle's transverse motion in phase space; see, for example, Fig. 16. The shape and orientation of the ellipse are determined by the optics function β and its derivative, $\alpha = -\frac{1}{2}\beta'$, and so change as a function of the position s in the ring; the area covered in phase space, however, is constant throughout the machine.

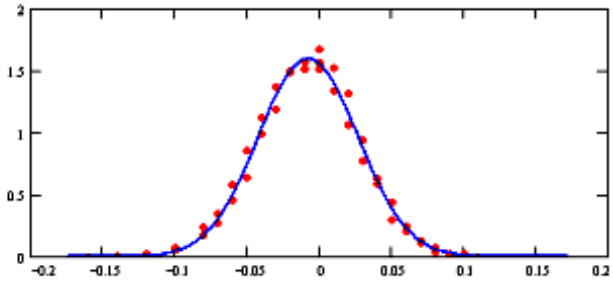


Fig. 17: Transverse particle distribution in a storage ring. The dots correspond to the measurement; the line is a Gaussian fit. A particle at 1σ from the beam centre is used to represent the beam size.

In Fig. 16, expressions for the dependence of the beam size and divergence and, as a consequence, the shape and orientation of the phase space ellipse are included. For the sake of simplicity, we shall not derive these expressions here; instead, see Ref. [3].

Referring once more to the issue of beam injection, it is clear now that in order to obtain a perfect beam transfer into a storage ring, it is *not* sufficient to correct the orbit, which means to get the origin of the phase space ellipse right. In addition, we have to optimize the beam optics of the transfer line in such a way that at the end we obtain the optics functions α and β that correspond to the lattice of the ring. Only then will a dilution of the emittance be avoided. This is valid for both transverse planes.

To complete the picture, we are now plotting, for a given position s in the ring, the co-ordinates x and x' of a single particle, turn by turn; doing so, we obtain the phase space co-ordinates of the particle, as shown in Fig. 16 (marked as dots in the figure). These co-ordinates follow the form of the ellipse, and the particle performs, from one turn to the next, a number of revolutions in phase space that corresponds to its tune. We have already emphasized that, as long as only conservative forces are considered (i.e., no interaction between the particles in a bunch, no collisions with remaining gas molecules, no radiation effects, etc.), the size of the ellipse in (x, x') space is constant and can be considered a quality factor of a single particle. Large areas in (x, x') space mean large amplitudes and angles of the transverse particle motion, and we would consider this as meaning a low particle 'quality'.

Finally, let us now talk a little more about the beam as an ensemble of many (typically 10^{11}) particles. Referring to Eq. (29), at a given position in the ring, the beam size is defined by the emittance ε and the amplitude function β. Thus, at a certain moment in time, the cosine term in Eq. (29) will be equal to one and the amplitude of the trajectory will reach its maximum value. Now, if we consider a particle at one standard deviation (σ) of the transverse density distribution, then by using the emittance of this reference particle we can calculate the size of the complete beam, in the sense that the complete area (within one sigma) of all particles in the (x, x') phase space is surrounded (and so defined) by our one-sigma candidate. Thus, the value $\sqrt{\varepsilon \cdot \beta(s)}$ defines the one-sigma beam size in the transverse plane.

An example of such a particle density distribution is shown in Fig. 17. The dots correspond to the measured values of the particle distribution at the collision point and the blue curve represents a Gaussian fit. The emittance (usually referred to as 'Courant–Snyder invariant') of the single particle at 1σ from the centre is used as representative of the emittance of the complete beam ensemble.

It is the task of the lattice designer to establish beam optics that guarantee—for a given beam emittance—values of the β-function that lead to tolerable beam sizes at every location in the machine. As an example, we shall use the values for the LHC proton beam (Fig. 18). In the periodic pattern of the arc, the β-function is equal to 180 m and the emittance ε at the flat-top energy is roughly 5×10^{-10} rad m. The resulting typical beam size is, therefore, 0.3 mm. Now, clearly, we would not design a vacuum aperture for the machine based on a one-sigma beam size; typically, an aperture requirement corresponding to 12σ is a good rule to guarantee a sufficient beam lifetime, allowing for tolerances arising from magnet

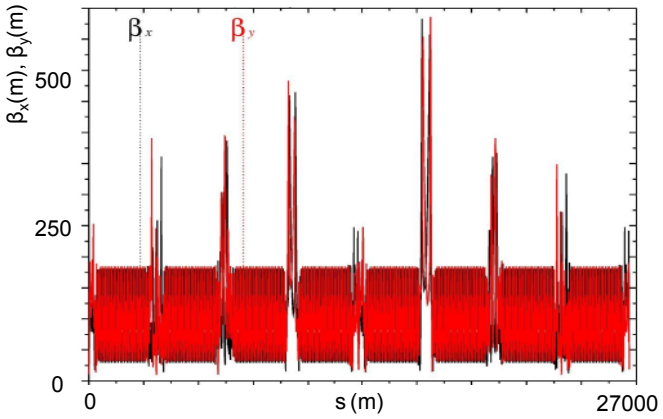


Fig. 18: LHC beam optics at injection energy: owing to the larger beam emittance at low energy, the beta-function has to be limited to values of about 600 m in the straight sections, while in the arc the periodic solution reaches a level of about 180 m.

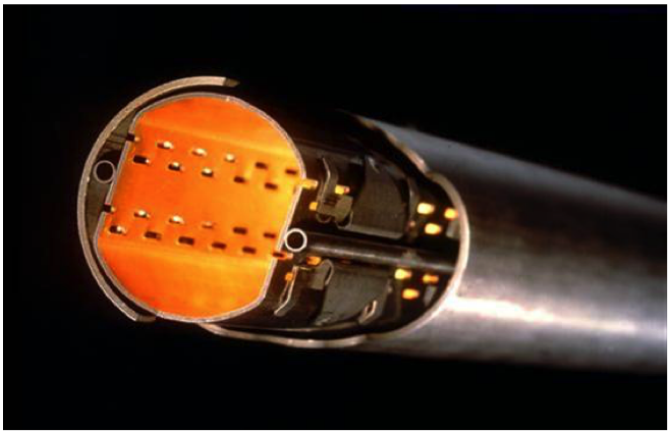


Fig. 19: LHC vacuum chamber with beam screen to shield the bore of the superconducting magnet from synchrotron radiation.

misalignment, optics errors, orbit fluctuations, and operational flexibility. In Fig. 19, part of the LHC vacuum chamber is shown, including the beam screen used to protect the cold bore from synchrotron radiation; its aperture limit corresponds to at least 18σ of the beam size.

3.2 Adiabatic shrinking

The definition of the beam emittance described in the previous section bears a certain problem. Strictly speaking, Liouville's theorem states that—given conservative forces—the particle density in the phase space x, p_x is constant. Now in accelerator physics we are talking about particle amplitudes and angles, x, x'; in the literature, such a co-ordinate system, defined by these variables, is sometimes called the *trace space* to make a clear distinction from the *phase space* of classical mechanics.

The main issue and important consequence for us is related to the particle acceleration. The angle x' of a particle is given by the ratio between longitudinal and transverse momentum:

$$x' = \frac{\mathrm{d}x}{\mathrm{d}s} = \frac{\mathrm{d}x}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}s} = \frac{p_x}{p_s} \propto \frac{1}{m_0 c\beta\gamma} \ , \tag{39}$$

where we express the relativistic momentum as a function of the rest mass m_0, the relativistic parameter $\beta = v/c$ and the Lorentz factor $\gamma = 1/(\sqrt{1-\beta^2})$. Here we have to pause for a moment and ask the

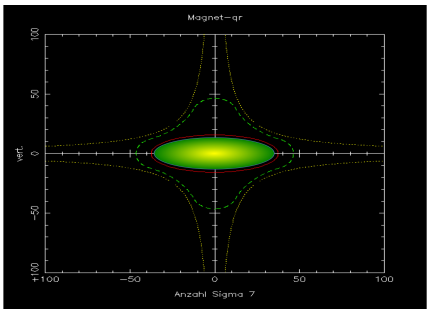


Fig. 20: Beam envelope of HERA proton ring at 40 GeV injection energy. The plot refers to 7σ at the mini-beta quadrupoles.

reader to accept our apologies. There is something called an amplitude- or β-function in beam dynamics, and now suddenly we have in addition the relative speed of the particle v/c, which is also called β. We hope that there is not too much confusion and in the end ... it is not our fault, as it was invented by our colleagues Albert and Ernest [8]. And we can understand their trouble in finding an adequate letter, as the Greek alphabet has only three letters anyway.

Back to our problem: p_s describes the longitudinal component of the particle's momentum; it is this longitudinal component that increases during the acceleration. Now, Liouville's theorem states that, for the canonical conjugate variables x and p_x, the phase space area is constant:

$$\int p_x \mathrm{d}x = \text{constant} \tag{40}$$

and we will not argue about that.

However, in the rather sloppy interpretation that we have used until now, we refer to a co-ordinate system x, x' and so in reality we get

$$\int x' \mathrm{d}x = \frac{\int p_x}{p_s} \mathrm{d}x \tag{41}$$

and as, during acceleration, the longitudinal momentum is obviously increasing, our x–x' ellipse will shrink proportionally to $1/\beta\gamma$. We conclude, therefore, that the beam emittance and, as a consequence, the beam dimension in both transverse planes will shrink during acceleration and this is indeed what we observe. As a consequence, a proton beam in a synchrotron or any particle beam in a linac will have the largest emittance at injection energy and it is here, where the beam optics—expressed as the beta-function—will have to be optimized for sufficient free aperture. The effect can be quite impressive: in Fig. 20, the 7σ envelope of a proton beam is shown inside the vacuum chamber (dashed line) of a mini-beta quadrupole magnet. The figure shows the situation at 40 GeV injection energy. In Fig. 21, for the same beam optics, and at the same location again, the 7σ envelope is shown, but now at a flat-top energy of 920 GeV. Owing to the much higher energy, the beam size is smaller by a factor of $\sqrt{920/40} = \sqrt{23}$ and the beam lifetime was considerably increased on the energy ramp of this machine.

As a direct consequence, we conclude that beam optics that lead to large beta-functions in the ring only can be applied at highest energy, where, owing to the reduced emittance, the overall beam size will still be limited. As an example, we refer again to the LHC situation. In direct comparison with the low-energy optics shown in Fig. 18, we now present the optics that are applied for high-energy collisions, in Fig. 22. Here, we can afford values of β of up to 4.5 km.

The design of the injection and extraction elements (kickers and septa), as well as the magnets for the transfer line in between, will therefore strongly depend on the beam energy considered; not only because of the trivial effect of the beam rigidity, discussed here, but also because of the emittance that determines the beam size and thus the aperture need of the injection and extraction elements.

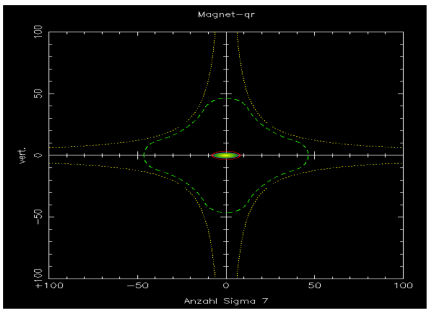


Fig. 21: Beam envelope of HERA proton ring at 920 GeV flat-top energy. The location and number of σ correspond to Fig. 20.

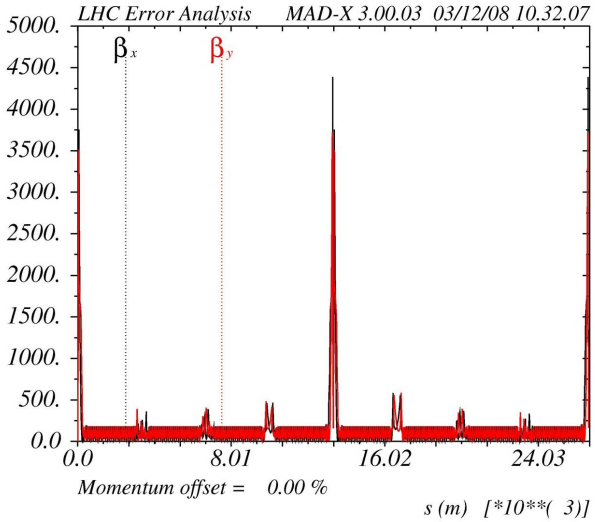


Fig. 22: LHC beam optics at high energy: owing to the small beam emittance at high energy, large values of the amplitude function β can be accepted; compare with the situation at injection energy, shown in Fig.18.

For completeness, we have to mention that as soon as synchrotron light effects must be considered, the situation changes drastically. In electron synchrotrons, the beam dynamics is determined by the equilibrium between synchrotron radiation damping and excitation due to the emitted photon quanta. Therefore, we observe a completely different behaviour; namely, a quadratic increase of the emittance with energy, $\varepsilon \propto \gamma^2$. We will summarize these facts in a later paper in this course.

4 Dispersion

So far, we have treated the beam and the equation of motion as a mono-energetic problem. Unfortunately, life is not so easy, and in the case of a real beam we have to deal with a considerable distribution of the particles with respect to energy or momentum. A typical value is

$$\frac{\Delta p}{p} \approx 1.0 \cdot 10^{-3}. \tag{42}$$

This momentum spread leads to several effects concerning the bending of the dipole magnets and the focusing strength of the quadrupoles. It turns out that the equation of motion, which has been a homogeneous differential equation until now, acquires a non-vanishing term on the right-hand side.

4.1 Dispersive effects

Replacing the ideal momentum p in Eq. (10) with $p_0 + \Delta p$, we obtain in approximation of small Δp, instead of Eq. (16):

$$x'' + x \cdot \left(\frac{1}{\rho^2} - k \right) = \frac{\Delta p}{p} \cdot \frac{1}{\rho} \, . \tag{43}$$

The general solution of our now inhomogeneous differential equation is, therefore, the sum of the solution of the homogeneous equation of motion and a particular solution of the inhomogeneous equation:

$$x(s) = x_\beta(s) + x_\mathrm{i}(s) \, . \tag{44}$$

Here, x_β is the solution that we have discussed up to now (the index β reminds us that these so-called betatron oscillations were observed for the first time in a machine called a *betatron*) and x_i is an additional contribution that still has to be determined. For convenience, we usually normalize this second term and define a special function, the so-called dispersion:

$$D(s) = \frac{x_\mathrm{i}(s)}{\Delta p / p} \, . \tag{45}$$

It describes the dependence of the additional amplitude of the transverse oscillation on the momentum error of the particle. It clearly fulfils the condition

$$x_\mathrm{i}''(s) + K(s) \cdot x_\mathrm{i}(s) = \frac{1}{\rho} \cdot \frac{\Delta p}{p} \, . \tag{46}$$

As before, we have combined the weak and strong focusing effects in the parameter $K := (1/\rho^2) - k$. The dispersion function is defined by the magnet lattice and is usually calculated by optics programs in the context of the calculation of the usual optical parameters. Analytically, it can be determined for single elements via the expression

$$D(s) = S(s) \cdot \int \frac{1}{\rho(\bar{s})} C(\bar{s}) \, \mathrm{d}\bar{s} - C(s) \cdot \int \frac{1}{\rho(\bar{s})} S(\bar{s}) \, \mathrm{d}\bar{s} \, , \tag{47}$$

where $S(s)$ and $C(s)$ correspond to the sine-like and cosine-like elements of the single-element matrices or of the corresponding product matrix if several elements are considered in the lattice.

Although this all sounds somewhat theoretical, we would like to stress that typical values for the beam size and dispersive effect in the case of a high-energy storage ring are

$$x_\beta \approx \text{1–2 mm}, \qquad D(s) \approx \text{1–2 m} \, . \tag{48}$$

Thus, for a typical momentum spread of $\Delta p/p = 1 \cdot 10^{-3}$, we obtain an additional contribution to the beam size from the dispersion function that is of the same order as that from the betatron oscillations, x_β. An example of a high-energy beam optics system including the dispersion function is shown in Fig. 23. It should be pointed out that the dispersion describes the special orbit that an ideal particle would have in the absence of betatron oscillations ($x_\beta = x'_\beta = 0$) for a momentum deviation of $\Delta p/p = 1$. In any case, it describes 'just another particle orbit' and so it is subject to the focusing forces of the lattice elements, as seen in the figure.

5 Transformation of the Twiss parameters α, β, γ

"Once more unto the breach, dear friends," [9].

While it is straightforward to develop the rules for the transformation of the trajectory amplitudes and angles, (x, x'), via the single-element matrices of the lattice elements (Eq. (25)), a similar formulation

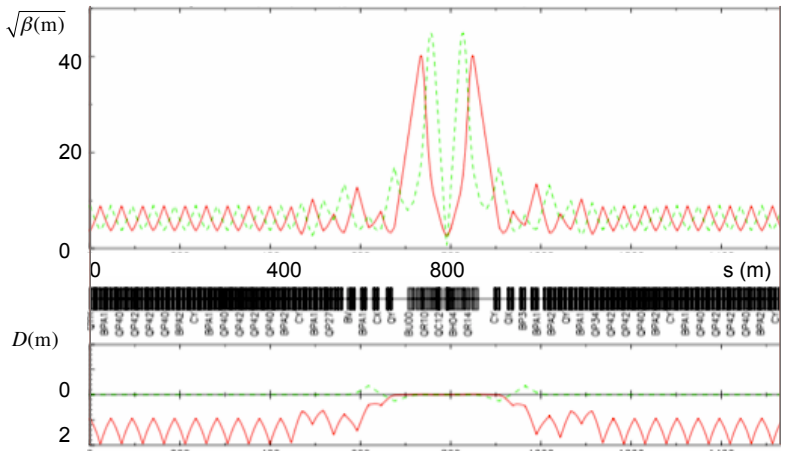


Fig. 23: β-function (top) and dispersion (bottom) of typical high-energy collider ring

can be deduced for the optical functions, α, β, and γ. The derivation is closely related to the fact that—for a given energy—the beam emittance ε is constant.

We are starting again from the usual transformation of a trajectory amplitude and angle between two locations in the lattice

$$\begin{pmatrix} x \\ x' \end{pmatrix}_{s2} = \mathbf{M} \begin{pmatrix} x \\ x' \end{pmatrix}_{s1} , \tag{49}$$

where the matrix $\mathbf{M}$ describes the focusing properties of a single lattice element, or in the case of several lattice elements, it represents the product matrix, as described in Eq. (26). In the general case of a sequence of lattice elements, we write

$$\mathbf{M} = \begin{pmatrix} C & S \\ C' & S' \end{pmatrix} , \tag{50}$$

where the matrix elements $C, S, \ldots$ refer to the elements of the product matrix or, in the trivial case, of, e.g., a single focusing quadrupole, are the usual descriptions that we introduced before:

$$\mathbf{M}_{\text{foc}} = \begin{pmatrix} C & S \\ C' & S' \end{pmatrix} = \begin{pmatrix} \cos\left(\sqrt{|K|}\, s\right) & \frac{1}{\sqrt{|K|}} \sin\left(\sqrt{|K|}\, s\right) \\ -\sqrt{|K|} \sin\left(\sqrt{|K|}\, s\right) & \cos\left(\sqrt{|K|}\, s\right) \end{pmatrix} . \tag{51}$$

Now we consider two locations, s_1 and s_2, in the storage ring, as shown schematically in Fig. 24. At both positions, the emittance can be expressed as a function of the Twiss parameters at these positions:

$$\varepsilon = \gamma_1 x^2(s_1) + 2\alpha_1 x(s_1) x'(s_1) + \beta_1 x'^2(s_1) , \tag{52}$$

$$\varepsilon = \gamma_2 x^2(s_2) + 2\alpha_2 x(s_2) x'(s_2) + \beta_2 x'^2(s_2) , \tag{53}$$

keeping in mind that the numerical value of the emittance at both positions has to be the same, as long as Mr. Liouville's theorem is fulfilled.

Knowing the amplitude and angle of the trajectory at position s_2, we can deduce these values at position s_1:

$$\begin{pmatrix} x \\ x' \end{pmatrix}_{s_1} = \mathbf{M}^{-1} \begin{pmatrix} x \\ x' \end{pmatrix}_{s_2} , \tag{54}$$

where the matrix from s_2 to s_1 is the inverse transformation matrix,

$$\mathbf{M}^{-1} = \begin{pmatrix} S' & -S \\ -C' & C \end{pmatrix} \tag{55}$$

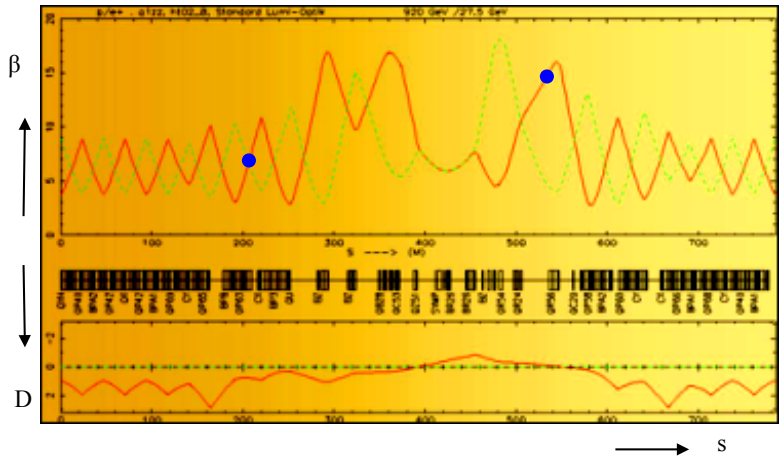


Fig. 24: The optical functions at two positions in a ring are related to each other via the constant beam emittance

and we used the fact that, for all matrices in a storage ring, the determinant has to be equal to one:

$$\mathrm{Det}(M) = CS' - SC' = 1 \ . \tag{56}$$

Thus, we can write for our trajectory co-ordinates:

$$x_1 = S'x_2 - Sx'_2 \,, \tag{57}$$
$$x'_1 = -C'x_2 + Cx'_2 \ . \tag{58}$$

Inserting these into Eq. (53), we express the emittance at position s_1 as a function of the trajectory co-ordinates at position s_2:

$$\varepsilon = \beta_1(Cx'_2 - C'x_2)^2 + 2\alpha_1(S'x_2 - Sx'_2)(Cx'_2 - C'x_2) + \gamma_1(S'x_2 - Sx'_2)^2 \ . \tag{59}$$

Sorting via x and x' and comparing the coefficients, we can finally relate the Twiss functions between the two locations in the ring:

$$\begin{aligned}
\beta(s_2) &= C^2\beta(s_1) - 2SC\alpha(s_1) + S^2\gamma(s_1) \,, \\
\alpha(s_2) &= -CC'\beta(s_1) + (SC' + S'C)\alpha(s_1) - SS'\gamma(s_1) \,, \\
\gamma(s_2) &= C'^2\beta(s_1) - 2S'C'\alpha(s_1) + S'^2\gamma(s_1) \,.
\end{aligned} \tag{60}$$

Once more—for the sake of elegance in our notation—we prefer to combine these relations in matrix form and get

$$\begin{pmatrix} \beta \\ \alpha \\ \gamma \end{pmatrix}_{s_2} = \begin{pmatrix} C^2 & -2SC & S^2 \\ -CC' & SC' + CS' & -SS' \\ C'^2 & -2S'C' & S'^2 \end{pmatrix} \cdot \begin{pmatrix} \beta \\ \alpha \\ \gamma \end{pmatrix}_{s_1} \ . \tag{61}$$

So if we once know, by calculation or measurement, the optical functions at one position in the ring, we can determine them via the single-element matrices in the lattice at any other location. This statement holds equally for rings and for transfer lines.

However—and here we have to be careful—there is something special for transfer lines and linear accelerators that we have to mention; unfortunately, it is quite an uncomfortable item. Unlike rings, be they synchrotrons or storage rings, the optical functions α, β, and γ are not defined in a non-periodic

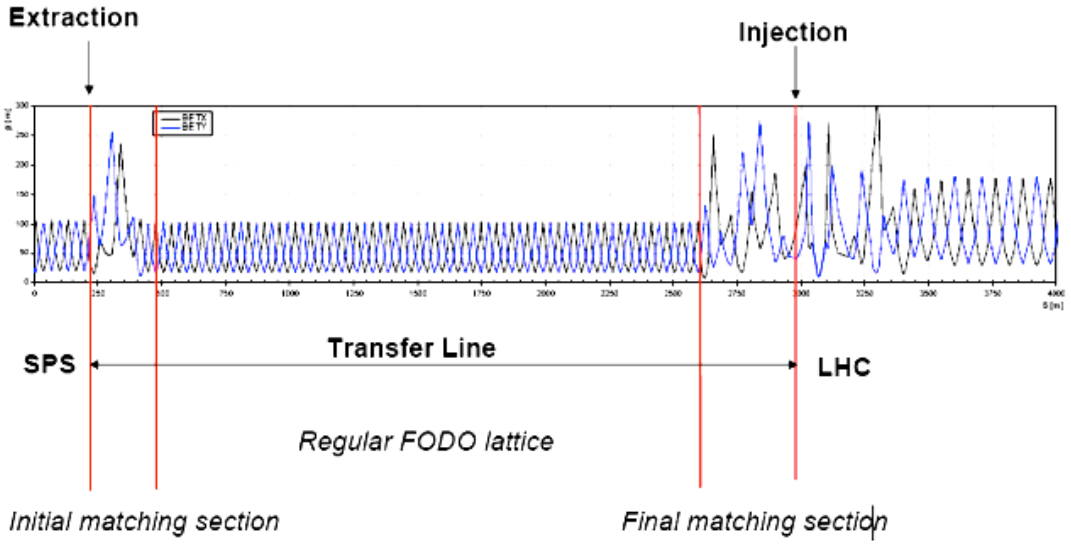


Fig. 25: Optics of the transfer line between the SPS and the HC



Fig. 26: The beam size must be measured at three locations in the transfer line

structure. Hill's equation (Eq. (28)) requires a periodic system and so does its solution. Still, this description is very powerful and so our colleagues from the non-periodic world love to use our language. However, we should be aware of this issue. In Eq. (60), we have learnt how to transform the optical functions from one position in the lattice to another, knowing the focusing elements in between. This means that if we know, e.g., the Twiss functions at the beginning of a transfer line, we can calculate them through the complete linear structure. An example is given in Fig. 25.

In the case of the example in Fig. 25, the optical functions at the beginning of the structure are defined, i.e., uniquely determined by the periodicity of the pre-accelerator, the SPS synchrotron. At the beginning and the end of the transfer line, special matching sections have been introduced to transform the periodic beta-functions from the SPS lattice onto the transfer line structure and from here to the LHC cells. Usually, such a matching section leads for a moment to a somehow increased beam amplitude (owing to the distorted beta-functions) and we have to take care to limit the aperture needs to reasonable values.

More complicated however is the case where a circular pre-accelerator does not exist. In such a case, we have to 'guess' the initial values of α, β, and γ, or, better, we have to measure them at the initial position before we can apply Eq. (60). To do so, we have to perform three measurements at three different locations in the line (see Fig. 26):

$$x(s) = a\sqrt{\beta(s)}\cos\big(\psi(s) - \phi\big)$$
$$x'(s) = a\sqrt{\beta(s)}\big(-\sin(\psi(s) - \phi)\big)\psi' + \frac{\beta'(s)}{2\sqrt{\beta}}a\cos(\psi(s) - \phi)$$

Here we follow a beautiful explanation from Phil Bryant [10]; this is the way the trick goes.

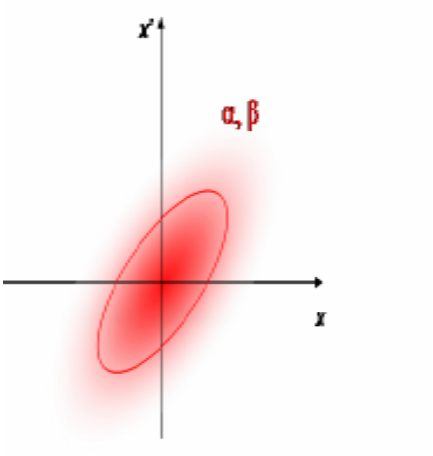


Fig. 27: Example of the transverse beam size measured in a transfer line

Assume that we measure the particle distribution, which might look like the example in Fig. 27.

Fitting a reasonable ellipse to the distribution, we can deduce the standard deviation of the, hopefully, nicely Gaussian-distributed particle density (...OK, for other than Gaussian distributions, it will also work more or less). So we obtain

$$\sigma = \sqrt{\varepsilon\beta} \tag{62}$$

and, as the beam emittance is constant, we can write—referring to each of the three positions s_0, s_1, and s_2—

$$\varepsilon = \frac{\sigma_0^2}{\beta_0} = \frac{\sigma_1^2}{\beta_1} = \frac{\sigma_2^2}{\beta_2} \tag{63}$$

Now, from Eq. (61), we know how the beta-functions transform through the lattice and so we can express β_1 and β_2 as a function of the initial value β_0:

$$\beta_1 = C_1^2\beta_0 - 2C_1S_1\alpha_0 + \frac{S_1^2}{\beta_0}\left(1+\alpha_0^2\right) \tag{64}$$

$$\beta_2 = C_2^2\beta_0 - 2C_2S_2\alpha_0 + \frac{S_2^2}{\beta_0}\left(1+\alpha_0^2\right) \tag{65}$$

Here, C_1, S_1, C_2, and S_2 describe the matrix elements of the product matrix from point s_0 to s_1 and from s_0 to s_2, respectively. Using this information, we can determine both α_0 and β_0:

$$\alpha_0 = \frac{1}{2}\beta_0\Gamma\,, \tag{66}$$

$$\beta_0 = \frac{1}{\sqrt{\left(\frac{\sigma_1}{\sigma_0^2}\right)/S_1^2 - (C_1/S_1)^2 + (C_1/S_1)\Gamma - \Gamma^2/4}}\,, \tag{67}$$

where we introduced the parameter

$$\Gamma = \frac{(\sigma_2/\sigma_0)^2/S_2^2 - (\sigma_1/\sigma_0)^2/S_1^2 - (C_2/S_2)^2 + (C_1/S_1)^2}{C_1/S_1 - C_2/S_2}\,,$$

et c'est ça.

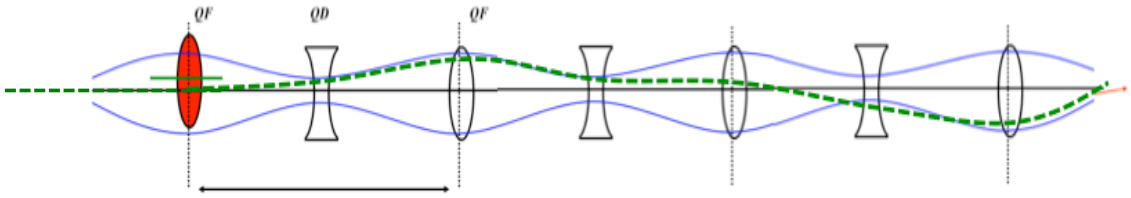


Fig. 28: Effect of a misaligned quadrupole in a transfer line. QD, defocusing quadrupole; QF, focusing quadrupole

6 Dipole errors and quadrupole misalignment

The design orbit, and thus the geometry of the ring or transfer line, is defined by the strength and arrangement of the dipole magnets. Under the influence of imperfections in the dipole field and (transverse) misalignment of the quadrupole magnets, unwanted deflection fields ('kicks') are created that influence this orbit. If these distortions are small enough (and hopefully they are), we will still obtain an orbit that is not too far away from the design. It is this 'reference orbit' that acts as a reference (hic) for the single-particle trajectories (i.e., the betatron oscillations) and we have to take care that it does not differ too much from the design.

A special issue, however, arises from the fact that in a ring this 'reference orbit' has, by definition, to be closed. While in a transfer line the effect of external distortions is somehow straightforward, in a periodic situation we have to be a bit more careful. For simplicity, we assume a small error and describe the effect as a sudden change of the particle's angle x',

$$\Delta\theta = \Delta x' = \frac{\mathrm{d}l}{\rho} = \frac{\int \delta B \mathrm{d}l}{B\rho} \ , \tag{68}$$

where we have again normalized the field error by the beam rigidity to obtain the deflection angle x'.

In a misaligned quadrupole, we get exactly the same problem. An offset Δx in the presence of a field gradient g leads to an effective dipole field that deflects the beam:

$$\Delta\theta = \frac{\int \Delta B \mathrm{d}l}{B\rho} = \frac{\int \Delta x g \mathrm{d}l}{B\rho} \ . \tag{69}$$

For a transfer line, the resulting effect is trivial. Assuming a short, localized deflection, the amplitude of the particle is not affected by the field, but the angle x' is and so we can write:

$$x_\mathrm{f} = x_\mathrm{i} = 0 \, , \tag{70}$$

$$x'_\mathrm{f} = x'_\mathrm{i} + \Delta x' = x'_\mathrm{i} + \frac{\int \delta B \mathrm{d}l}{B\rho} \ , \tag{71}$$

where the indices 'i' and 'f' refer to the position just in front of the kick and straight after. From this moment on, the originally ideal trajectory will be transformed through the lattice elements in the usual way:

$$\begin{pmatrix} x \\ x' \end{pmatrix}_s = \mathbf{M} \begin{pmatrix} x \\ x' \end{pmatrix}_0 \ , \tag{72}$$

which is shown qualitatively for a transfer line in Fig. 28.

In a circular machine, things get a bit more complicated. As we talk about a ring, the periodic boundary conditions after one turn have to be taken into account. Mathematically, we express this fact by:

$$x(s+L) = x(s), \qquad x'(s+L) + \mathrm{d}x' = x'(s) \ . \tag{73}$$

The kick creates an orbit deflection that travels (i.e., oscillates) around the complete ring and leads—unlike the situation in a transfer line—to an offset even at the origin of the kick itself, which is unchanged

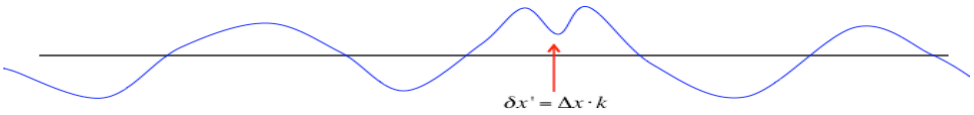


Fig. 29: Effect of a misaligned quadrupole in a ring: owing to the periodicity condition, the orbit has to close upon itself.

turn after turn; we still get a closed orbit. However, as the trajectory has to close upon itself, the angle x' after one turn has to take the distortion $\mathrm{d}x'$ into account (Fig. 29).

Starting from the general solution of Hill's equation, and using the periodicity condition, we write for the amplitude:

$$x(s) = a\sqrt{\beta(s)}\cos\left(\psi(s) - \phi\right) \tag{74}$$

$$x(s+L) = x(s) \tag{75}$$

$$a\sqrt{\beta(s+L)}\cos\left(\psi(s) + 2\pi Q - \phi\right) = a\sqrt{\beta(s)}\cos\left(\psi(s) - \phi\right) \tag{76}$$

The amplitude factor a will be determined by the periodicity conditions and, clearly enough, the phase advance per turn increases by

$$\psi(s+L) = 2\pi Q \ .$$

As the amplitude function β is periodic, by definition, we obtain as a first condition:

$$\cos(2\pi Q - \phi) = \cos(-\phi) = \cos(\phi)\,, \tag{77}$$

$$\phi = \pi Q \ . \tag{78}$$

The boundary condition set by the amplitude fixes the initial condition for the phase ϕ. Following the same arguments, but now for the angle x', we get

$$x(s) = a\sqrt{\beta(s)}\cos\left(\psi(s) - \phi\right), \tag{79}$$

$$x'(s) = a\sqrt{\beta(s)}\big(-\sin(\psi(s) - \phi)\big)\psi' + \frac{\beta'(s)}{2\sqrt{\beta}}a\cos(\psi(s) - \phi)\,, \tag{80}$$

and writing $\delta x'$ for the local kick due to the field distortion,

$$x'(s+L) + \delta x' = x'(s)\,,$$

we get

$$-a\frac{1}{\sqrt{\beta(\tilde{s}+L)}}\sin(2\pi Q - \phi) + \frac{\beta'(\tilde{s}+L)}{2\beta(\tilde{s}+L)}\sqrt{\beta(\tilde{s}+L)}a\cos(2\pi Q - \phi) + \delta x'$$
$$= -a\frac{1}{\sqrt{\beta(\tilde{s})}}\sin(-\phi) + \frac{\beta'(\tilde{s})}{2\beta(\tilde{s})}\sqrt{\beta(\tilde{s})}a\cos(-\phi)\,.$$

Here we have explicitly written $\tilde{s}$ for the position of the dipole field error, to emphasize that, e.g., the optical functions are to be taken at this position. Knowing that, from the periodicity condition we derived,

$$\beta(s+L) = \beta(s), \qquad \phi = \pi Q\,, \tag{81}$$

we can solve for the amplitude a and get

$$a = \delta x'\sqrt{\beta(\tilde{s})}\frac{1}{2\sin(\pi Q)} \ . \tag{82}$$

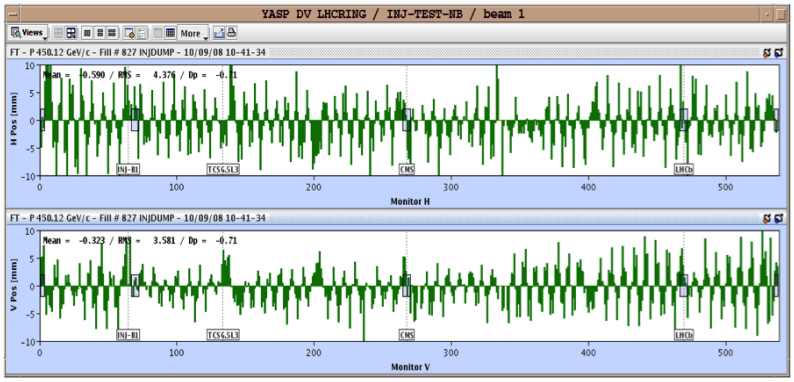


Fig. 30: Measured orbit in LHC during commissioning of the machine

Inserting this equation into Eq. 80, we get the final result that the amplitude of the closed orbit under the influence of a single dipole distortion (or quadrupole misalignment) is given by

$$x(s) = \delta x' \sqrt{\beta(\tilde{s})}\sqrt{\beta(s)}\frac{\cos(\psi(s)-\phi)}{2\sin(\pi Q)} \ . \tag{83}$$

We conclude that the distorted orbit depends on the kick strength, the local β- function at the location $\tilde{s}$ of the distortion, and the β- function at the observation point s. In addition, there is a resonance denominator, which will amplify any external orbit distortion, if the tune in the corresponding plane is on—or close to—an integer value. In such a case, the particle amplitude will increase ad infinitum and lead very quickly to particle losses; so better watch your tune!

For completeness: if we do not refer to a special starting point and express the orbit distortion as normalized dipole strength, $1/\rho(\tilde{s})$, we get the general expression

$$x(s) = \frac{\sqrt{\beta(s)}}{2\sin(\pi Q)} \oint \sqrt{\beta(\tilde{s})}\frac{1}{\rho(\tilde{s})}\cos\big(|\psi(\tilde{s})-\psi(s)|-\pi Q\big)\mathrm{d}\tilde{s} \ . \tag{84}$$

We would not like to close this section without showing a real example of orbits for both cases, the closed orbit in a storage ring and the orbit in a transfer line or linacs. We have already seen an example of the first case in Fig. 11, which we plot here once more, for simplicity (Fig. 30). It shows the closed orbit of the LHC storage ring during the start-up phase of the machine, where considerable amplitudes in both planes were observed. The tune of the machine was set to $Q_x = 64.31$, so sufficiently away from resonance conditions. Still, the alignment tolerances of the magnets of $\Delta x \approx \Delta y \approx 150$ µm caused a considerable orbit distortion of up to 10 mm.

The next plot, Fig. 31, refers to the situation in a transfer line (observed at the HERA collider at DESY). While in the first part of the structure the oscillations are well corrected and small, suddenly a strong trajectory fluctuation is created, owing to a misaligned quadrupole lens in the middle of the lattice. As the transfer line is not closed upon itself, the observed orbit develops according to Eqs. (71) and (72).

6.1 Emittance in electron rings or linacs

There is a special issue about electron beams that should not be forgotten: these particles are so light in weight (or, to be scientifically more correct, their Lorentz factor gamma is, even for moderate particle energies, already so large) that they emit radiation whenever accelerated or bent. This so-called synchrotron radiation has a strong influence on the beam dynamics, see, e.g., Ref. [11]. Summarizing these facts, we can state the following.

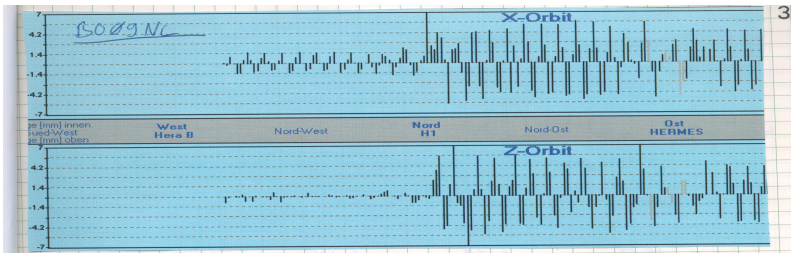


Fig. 31: Measured orbit in a beam transfer line, including the effect of a quadrupole lens that is misaligned in both planes.

– The power of the radiated synchrotron light depends on the energy of the particle ($\gamma = E/m_0c^2$) and the bending radius ρ of the trajectory under the influence of the field acting on it,

$$\Delta P = \frac{e^2 c}{6\pi\varepsilon_0}\frac{\gamma^4}{\rho^2} \ . \tag{85}$$

– The energy loss per turn in a circular machine is given by

$$\Delta E = \frac{e^2}{3\varepsilon_0}\frac{\gamma^4}{\rho} \ . \tag{86}$$

– The critical energy $E_{\mathrm{crit}} = \hbar * \omega$ of the emitted radiation is given by

$$\omega_{\mathrm{c}} = \frac{3c}{2}\frac{\gamma^3}{\rho} \ . \tag{87}$$

The damping effect of the light emission and the quantum effect of the emitted photons leads to an equilibrium emittance of the beam that is given by

$$\varepsilon_{x_0} = \frac{C_q E^2}{J_0}\frac{\langle H\rangle_{\mathrm{mag}}}{\rho} \ , \tag{88}$$

where J_x is the so-called damping partition number, which is determined by the lattice and is usually close to $J_x = 1$. The *H*-function describes the influence of the optical parameters α, β, and γ and dispersion D:

$$\langle H\rangle = \gamma D^2 + 2\alpha D D' + \beta D'^2 \, ; \tag{89}$$

C_q is a constant that we usually introduce to make our equations more compact. It is given by

$$C_q = \frac{55}{32\sqrt{3}}\frac{\hbar c}{(m_{\mathrm{e}}c^2)^3} \tag{90}$$

and for electrons it has the numerical value

$$C_q = 1.468 \times 10^{-6}\left[\frac{\mathrm{m}}{\mathrm{GeV}^2}\right] \ . \tag{91}$$

Now, while this clearly has a strong impact on the magnet strength and the design of the lattice, it also affects the orbit correction scheme. Any (!) external field, including off-centre quadrupoles, and including the effect of orbit corrector dipoles, will influence the beam emittance. In the quest for the highest brilliance of the emitted light and so for smallest possible beam emittances, therefore, these effects must be taken into account.

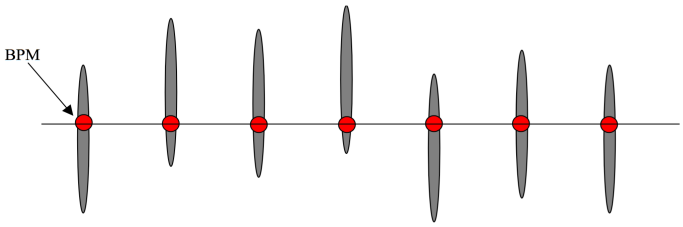


Fig. 32: Lattice with misaligned quadrupoles and perfect beam position monitors (BPMs)

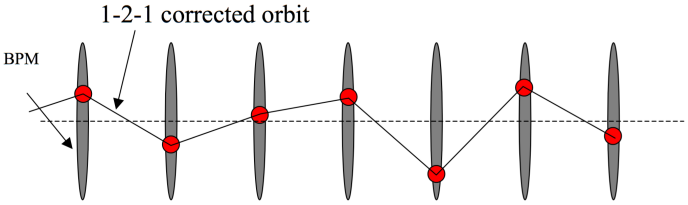


Fig. 33: Lattice with perfectly aligned quadrupoles but offsets in the reading of the beam position monitors (BPMs). While the actual orbit is perfect, the beam position monitor readings simulate an orbit distortion.

Any deflecting field will change (i.e. create) additional beam emittance and we have to be careful when it comes to orbit corrections. Assume the two extreme cases, which are shown schematically in Figs. 32 and 33.

In the case of Fig. 32, an orbit correction algorithm will tell us exactly what to do. Each misaligned quadrupole will lead to an orbit defection, which can and should be corrected by the corrector dipole next to it. The result will be a nearly perfect compensation of the quadrupole offsets and a nicely small beam emittance.

Consider, however, the case of Fig. 33. Here, the quadrupoles are perfectly aligned and it is the beam position monitor system that causes the trouble. Nobody is perfect and so even beam position monitors can have some reading errors that lead to artificial beam position offsets. A straightforward approach, as explained, will reduce the beam position monitor readings, but in reality lead to a distorted orbit and so create additional dispersion in the machine and lead to increased emittances. Especially in the vertical plane this effect is most serious, as vertical bending fields will not usually be present in the machine and the emittance in this plane should be minimized. Special techniques are needed and have been developed to avoid such a problem. Dispersion-free steering methods are widely used [12, 13]; instead of correcting the orbit (i.e., the beam position monitor readings), we concentrate directly on the dispersion that can be measured in the machine and we power our corrector magnets in such a way that the dispersion is minimized around the storage ring.

In the commissioning phase of an accelerator, and during daily operation, such orbit oscillations should be reduced as much as possible. At first sight, they eat up free aperture and can easily lead to particle losses at the vacuum chamber wall. Beyond that, the beam will be off-centre in the magnet elements and especially in the presence of sextupole magnets; this will create optics problems (tune shift and beta beat) and, in the case of vertical offsets, will create coupling. Talking for a moment again about electron beams, any offset in a magnetic field of a quadrupole or higher-order magnet will immediately create synchrotron radiation and unwanted spurious dispersion with its serious drawback for the beam emittance. In light sources and modern lepton colliders, where we seek for the smallest vertical emittance,

any vertical deflecting fields should be avoided; here, people use a series of orbit correction tools that are powered 'on-beam' in feedback loops to keep the orbit as 'golden' as possible.

Moreover, during the actual injection or extraction process, the effect of steering errors, as well as the effect of optics mismatch, will lead to emittance growth that can be large enough to spoil the beam quality. A summary of these problems and their impact on the beam is actually the topic of this school and will be presented in several of the next papers of these proceedings.

7 Conclusion

The topic of injection, extraction, and beam transfer is one of the most critical issues in the lifetime of a particle beam, and therefore in the design of an accelerator complex. The fast-changing fields of the lattice elements involved usually represent non-adiabatic changes that—if handled in an improper way—can easily destroy the beam quality. A decent optimization of the design and the quality of the hardware involved, and the possibility of measuring the relevant beam parameters (and optimizing them accordingly), is a must. The optimization of a typical beam injection, e.g., will therefore involve a careful steering of the orbit amplitude and angle in both transverse planes, matching of the optics (horizontal and vertical α and β functions), and their measurement. Similar care has to be taken in the longitudinal plane, which will be discussed in another paper. Here, we have to optimize and control the bunch length and the phase and the energy of the injected particle bunches.

This can be tedious work, but it pays off as soon as it comes to beam quality, background reduction, emittance dilution, and, for a collider, finally, the luminosity that we can get out of the machine.

And so, in the end, and after all, we want to conclude this brief paper with a word of consolation:

Usually it works.

References

[1] K. Wille, *The Physics of Particle Accelerators* (Oxford University Press, Oxford, 2000).

[2] J. Rossbach and P. Schmüser, Basic course on accelerator optics, edited by S. Turner, CERN-1994-001, (CERN, Geneva, 1994), pp. 17–88, http://dx.doi.org/10.5170/CERN-1994-001.17

[3] B.J. Holzer, Lattice design in high-energy particle accelerators, edited by W. Herr, CERN-2014-009, (CERN, Geneva, 2014).

[4] E. Jaeschke *et al.*, The Heidelberg test storage ring for heavy ions TSR, Proc. EPAC, Rome, 1988, edited by S. Tazzari (World Scientific, Teaneck, 1988), p. 365.

[5] H. Goldstein, *Klassische Mechanik* (Akademische Verlag-Gesellschaft, Wiesbaden, 1981).

[6] A. Wolski, *Beam Dynamics in High Energy Particle Accelerators* (Imperial College Press, London, 2014), p. 59.

[7] HERA Design Team, HERA: a proposal for a large electron proton colliding beam facility at DESY, DESY-HERA-81/10 (DESY, Hamburg, 1981).

[8] E.D. Courant and H.S. Snyder, *Ann. Phys.* **6** (1958) 1.

[9] W. Shakespeare, *Henry V*, Act 3, Scene 1, Line 1.

[10] P.J. Bryant, Proc. CAS–CERN Accelerator School, 5th General Accelerator Physics Course, Jyväskylä, Finland, 1992, CERN-1994-001 (CERN, Geneva, 1994), pp. 17–88. http://dx.doi.org/10.5170/CERN-1994-001.17

[11] L. Rivkin, these proceedings

[12] T.O. Raubenheimer and R.D. Ruth, *Nucl. Instrum. Methods* **A302** (1991) 191.

[13] R. Assmann *et al.*, Emittance optimization with dispersion free steering at LEP, CERN-SL-2000-078-OP, (CERN, Geneva, 2000).

Bibliography

B.J. Holzer, Proc. CAS–CERN Accelerator School, Course on Superconductivity for Accelerators, Erice, Italy, 2013, CERN-2014-005

F. Tecker, Longitudinal beam dynamics, in Proceedings of the CAS-CERN Accelerator School: Advanced Accelerator Physics Course, Trondheim, Norway, 18–29 August 2013, edited by W. Herr, CERN-2014-009 (CERN, Geneva, 2014), p. 1–21, https://doi.org/10.5170/CERN-2014-009.1.

C. Bernardini, *Phys. Persp.* **6** (2004) 156.

LHC design report, CERN-2004-003 (CERN, Geneva, 2004), https://doi.org/10.5170/CERN-2004-003-V-1, https://doi.org/10.5170/CERN-2004-003-V-2, https://doi.org/10.5170/CERN-2004-003-V-3.

P.J. Bryant, A brief history and review of accelerators, in Proceedings of the CAS-CERN Accelerator School: 5th General Accelerator Physics Course, Jyväskylä, Finland, 7–18 September 1992, edited by S. Turner, CERN-1994-001 (CERN, Geneva, 1994), p. 1-16, https://doi.org/10.5170/CERN-1994-001.1.

N. Walker, Emittance preservation, CAS–CERN Accelerator School, 5th Intermediate Accelerator Physics Course, Daresbury, UK, 2007.

Proceedings of the CAS–CERN Accelerator School: Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018)

Longitudinal Beam Dynamics—Recap

F. Tecker
CERN, Geneva, Switzerland

Abstract
The paper gives a very brief summary of longitudinal beam dynamics for both linear and circular accelerators. After discussing synchronism conditions in linacs, it focuses on particle motion in synchrotrons. It summarizes the equations of motion, discusses phase-space matching during beam transfer, and introduces the Hamiltonian of longitudinal motion.

Keywords
Longitudinal beam dynamics; synchrotron motion; synchrotron oscillations; longitudinal phase space; Hamiltonian.

1 Introduction

The force $\vec{F}$ on a charged particle with a charge e is given by the Newton–Lorentz force:

$$\vec{F} = \frac{\mathrm{d}\vec{p}}{\mathrm{d}t} = e\left(\vec{E} + \vec{v} \times \vec{B}\right) . \tag{1}$$

The second term is always perpendicular to the direction of motion, so it does not give any longitudinal acceleration and it does not increase the energy of the particle. Hence, the acceleration has to be done by an electric field $\vec{E}$. To accelerate the particle, the field needs to have a component in the direction of motion of the particle. If we assume the field and the acceleration to be along the z direction, Eq. (1) becomes

$$\frac{\mathrm{d}p}{\mathrm{d}t} = eE_z . \tag{2}$$

The total energy E of a particle is the sum of the rest energy E_0 and the kinetic energy W:

$$E = E_0 + W . \tag{3}$$

In relativistic dynamics, the total energy E and the momentum p are linked by

$$E^2 = E_0^2 + p^2c^2 , \tag{4}$$

from which it follows that

$$\mathrm{d}E = v\,\mathrm{d}p . \tag{5}$$

The rate of energy gain per unit length of acceleration (along the z direction) is then given by

$$\frac{\mathrm{d}E}{\mathrm{d}z} = v\frac{\mathrm{d}p}{\mathrm{d}z} = \frac{\mathrm{d}p}{\mathrm{d}t} = eE_z \tag{6}$$

and the (kinetic) energy gained from the field along the z path follows from $\mathrm{d}W = \mathrm{d}E = eE_z\,\mathrm{d}z$:

$$W = e\int E_z\,\mathrm{d}z = eV , \tag{7}$$

where V is just an electric potential.

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-005.31

The accelerating system will depend on the evolution of the particle velocity, which depends strongly on the type of particle. The velocity is given by

$$v = \beta c = c\sqrt{1 - \frac{1}{\gamma^2}}\,, \tag{8}$$

with the relativistic gamma factor of $\gamma = E/E_0$, the total energy E divided by the rest energy E_0. Electrons reach a constant velocity (close to the speed of light) at relatively low energies of a few megaelectronvolts, while heavy particles reach a constant velocity only at very high energies. As a consequence, one needs different types of resonator, optimized for different velocities. In particular, this requires an acceleration system that remains synchronized with the particles during their acceleration. For instance, when the revolution frequency in a synchrotron varies, the radio frequency (RF) will have to change correspondingly.

2 Phase stability and energy-phase oscillation in a linac

Several phase conventions exist in the literature (see Fig. 1):

- mainly for circular accelerators, the origin of time is taken at the zero crossing with positive slope;
- mainly for linear accelerators, the origin of time is taken at the positive crest of the RF voltage.

In the following, I will stick to the former case of the positive zero crossing for both linear and circular cases.

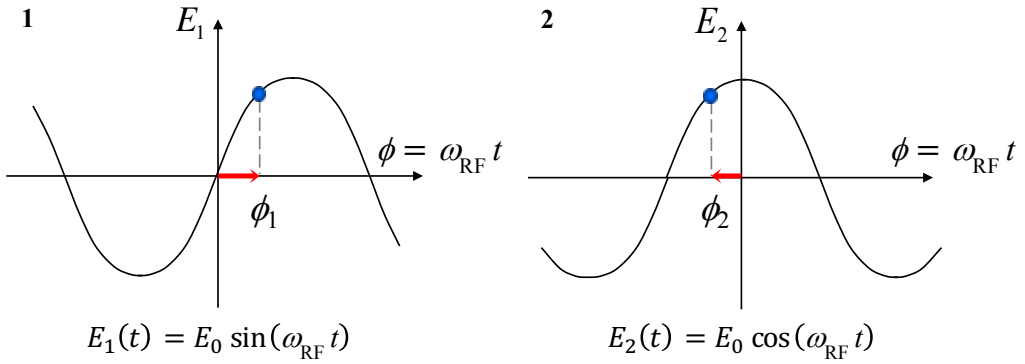


Fig. 1: Common phase conventions: (1) the origin of time is taken at the zero crossing with positive slope; (2) the origin of time is taken at the positive crest of the RF voltage.

Let us consider an Alvarez structure, a succession of drift tubes in a single resonant tank connected to a RF oscillator that changes the potential of the tubes (see Fig. 2). A particle arriving in a gap at the time of an accelerating field will gain energy and will be further accelerated. Inside the tubes, the particle is shielded from the outside field. If the polarity of the field is identical in the next gap, the particle is again accelerated in this gap. This leads to the synchronism condition that the distance L between the gaps must fulfil:

$$L = vT\,, \tag{9}$$

where $v = \beta\, c$ is the particle velocity and T the period of the RF oscillator. As the particle velocity increases, the drift spaces must get longer. It is clear that this arrangement cannot accelerate a continuous beam. Only a certain phase range will be accelerated and the beam must be bunched.

By design, the energy gain for a particle passing through the structure at a certain RF phase ϕ_s is such that the particle reaches the next gap with the same phase ϕ_s. Then the energy gain in the following

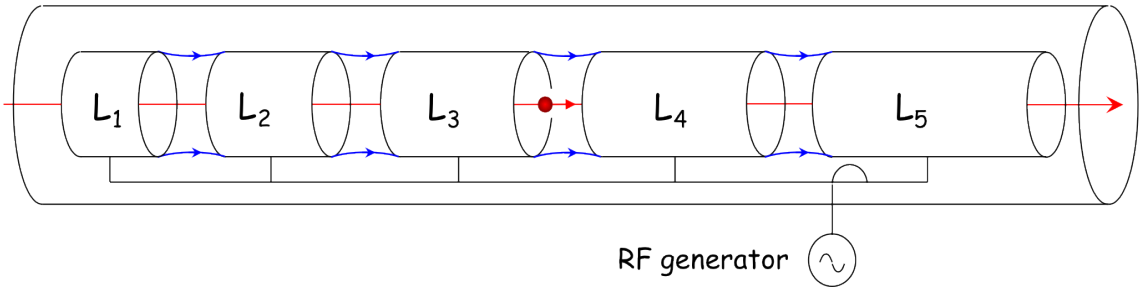


Fig. 2: Alvarez-type accelerating structure

gap will again be the same, and the particle will pass all gaps at this phase ϕ_s, which is called the 'synchronous phase'. So, the energy gain is $eV_s = e\hat{V}\sin\phi_s$. This is illustrated in Fig. 3 by the points P_1 and P_3.

A particle N_1, which arrives in a gap earlier than P_1, will gain less energy and its velocity will be smaller, so that it will take more time to travel through the drift tube. In the next gap, it will appear closer to particle P_1. The effect is opposite for particle M_1, which will gain more energy and reduce its delay compared with P_1. So, the points P_1, P_3, etc. are stable points for the acceleration, since particles positioned slightly away from them will experience forces that will reduce their deviation. On the contrary, it can be seen that point P_2 is an unstable point, in the sense that particles slightly away from this point will deviate even more in the next gaps.

Thus, for stability of the longitudinal oscillation, the particle needs to be on the rising slope of the RF field to have a restoring force towards the stable phase.

To study longitudinal motion, it is convenient to use variables that give the phase and energy relative to the synchronous particle (denoted by the subscript 's'):

$$\varphi = \phi - \phi_s\,, \tag{10}$$

$$w = E - E_s = W - W_s\,. \tag{11}$$

The accelerating field can be simply described by

$$E_z = E_0 \sin(\omega t)\,. \tag{12}$$

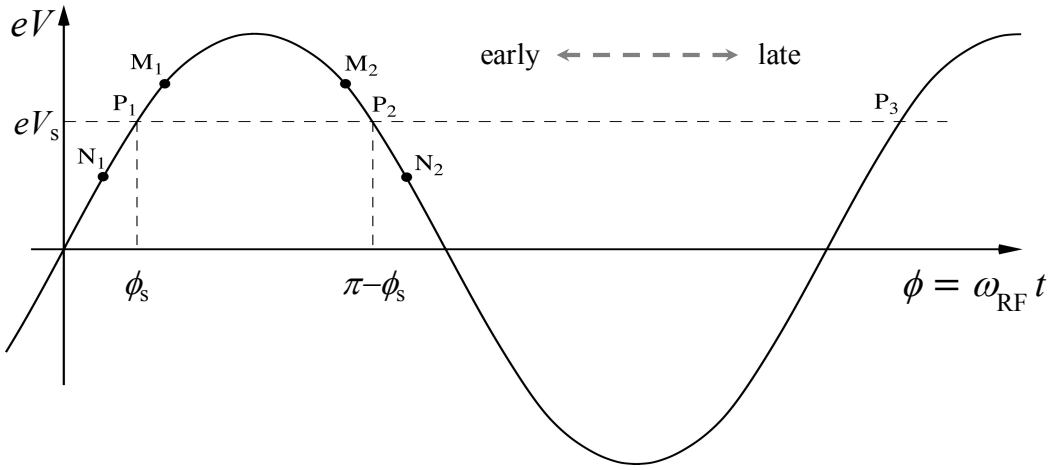


Fig. 3: Energy gain as a function of particle phase

The rate of energy gain for the synchronous particle is given by

$$\frac{\mathrm{d}E_\mathrm{s}}{\mathrm{d}z} = \frac{\mathrm{d}p_\mathrm{s}}{\mathrm{d}t} = eE_0 \sin\phi_\mathrm{s} \tag{13}$$

and, for a non-synchronous particle (for small φ),

$$\frac{\mathrm{d}w}{\mathrm{d}z} = eE_0 \left[\sin(\phi_\mathrm{s}+\varphi) - \sin\phi_\mathrm{s}\right] \approx eE_0 \cos\phi_\mathrm{s}\,\varphi\,. \tag{14}$$

The rate of change of the phase with respect to the synchronous particle is, for small deviations,

$$\frac{\mathrm{d}\varphi}{\mathrm{d}z} = \omega_\mathrm{RF}\left[\frac{\mathrm{d}t}{\mathrm{d}z} - \left(\frac{\mathrm{d}t}{\mathrm{d}z}\right)_\mathrm{s}\right] = \omega_\mathrm{RF}\left(\frac{1}{v} - \frac{1}{v_\mathrm{s}}\right) \approx -\frac{\omega_\mathrm{RF}}{v_\mathrm{s}^2}(v - v_\mathrm{s})\,. \tag{15}$$

Using $\mathrm{d}\gamma = \gamma^3\beta\,\mathrm{d}\beta$, w becomes, in the vicinity of the synchronous particle,

$$w = E - E_\mathrm{s} = m_0c^2(\gamma - \gamma_\mathrm{s}) = m_0c^2\,\mathrm{d}\gamma = m_0c^2\gamma_\mathrm{s}^3\beta_\mathrm{s}\,\mathrm{d}\beta = m_0\gamma_\mathrm{s}^3 v_\mathrm{s}(v - v_\mathrm{s})\,, \tag{16}$$

which leads to

$$\frac{\mathrm{d}\varphi}{\mathrm{d}z} = -\frac{\omega_\mathrm{RF}}{m_0 v_\mathrm{s}^3\gamma_\mathrm{s}^3} w\,. \tag{17}$$

Combining the two first-order equations (Eqs. (14) and (17)) into a second-order equation gives the equation of a harmonic oscillator with the angular frequency Ω_s:

$$\frac{\mathrm{d}^2\varphi}{\mathrm{d}z^2} + \Omega_\mathrm{s}^2\varphi = 0 \qquad \text{with} \qquad \Omega_\mathrm{s}^2 = \frac{eE_0\omega_\mathrm{RF}\cos\phi_\mathrm{s}}{m_0 v_\mathrm{s}^3\gamma_\mathrm{s}^3}\,. \tag{18}$$

Stable harmonic oscillations imply that $\Omega_\mathrm{s}^2 > 0$ and real, which means that $\cos\phi_\mathrm{s} > 0$. Since acceleration means that $\sin\phi_\mathrm{s} > 0$, it follows that the stable phase region for acceleration in the linac is

$$0 < \phi_\mathrm{s} < \frac{\pi}{2}\,, \tag{19}$$

which confirms what we have seen before in our discussion about the restoring force towards the stable phase.

From Eq. (18), it is also clear that the oscillation frequency decreases strongly with increasing velocity (and relativistic gamma factor) of the particle. For highly relativistic particles, the velocity change is negligible, so there is practically no change in the particle phase, and the bunch distribution is no longer changing.

3 Synchrotron

A synchrotron (see Fig. 4) is a circular accelerator in which the nominal particle trajectory is kept at a constant physical radius by variation of both the magnetic field and the RF, to follow the energy variation. In this way, the aperture of the vacuum chamber and the magnets can be kept small.

The RF needs to be synchronous to the revolution frequency. To achieve synchronism, the synchronous particle needs to arrive at the cavity again after one turn with the same phase. This implies that the angular RF, $\omega_\mathrm{RF} = 2\pi f_\mathrm{RF}$, must be an integer multiple of the angular revolution frequency ω:

$$\omega_\mathrm{RF} = h\,\omega\,, \tag{20}$$

where h is an integer and is called the *harmonic number*. As a consequence, the number of stable synchronous particle locations equals the harmonic number h. They are equidistantly spaced around

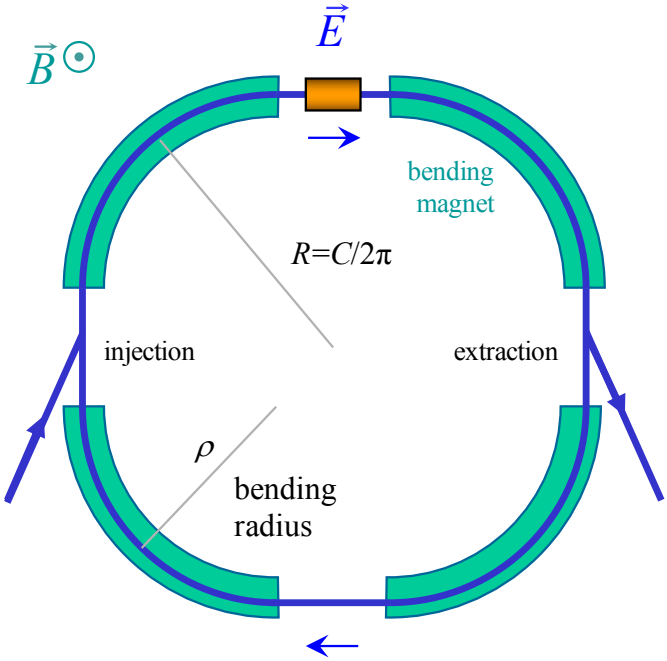


Fig. 4: Synchrotron

the circumference of the accelerator. All synchronous particles will have the same nominal energy and will follow the nominal trajectory.

Energy ramping is obtained by varying the magnetic field, while following the change of the revolution frequency with a change in the RF. The time derivative of the momentum,

$$p = eB\rho\,, \tag{21}$$

yields (when keeping the bending radius ρ constant)

$$\frac{\mathrm{d}p}{\mathrm{d}t} = e\rho\dot{B}\,. \tag{22}$$

For one turn in the synchrotron, this results in

$$(\Delta p)_{\mathrm{turn}} = e\rho\dot{B}T_{\mathrm{r}} = \frac{2\pi e\rho R\dot{B}}{v}\,, \tag{23}$$

where $R = L/2\pi$ is the physical radius of the machine.

Since $E^2 = E_0^2 + p^2c^2$, it follows that $\Delta E = v\Delta p$, so that

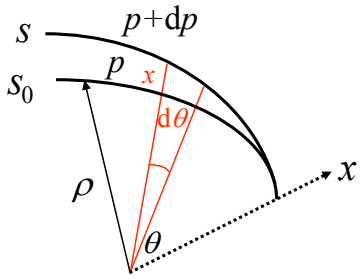
$$(\Delta E)_{\mathrm{turn}} = (\Delta W)_{\mathrm{s}} = 2\pi e\rho R\dot{B} = e\hat{V}\sin\phi_{\mathrm{s}}\,. \tag{24}$$

From this relation it can be seen that the stable phase for the synchronous particle changes during the acceleration, when the magnetic field B changes, as

$$\sin\phi_{\mathrm{s}} = 2\pi\rho R\frac{\dot{B}}{\hat{V}_{\mathrm{RF}}} \qquad \text{or} \qquad \phi_{\mathrm{s}} = \arcsin\left(2\pi\rho R\frac{\dot{B}}{\hat{V}_{\mathrm{RF}}}\right)\,. \tag{25}$$

As mentioned previously, the RF has to follow the change of revolution frequency and will increase during acceleration as

$$\frac{f_{\mathrm{RF}}}{h} = f_{\mathrm{r}} = \frac{v(t)}{2\pi R_{\mathrm{s}}} = \frac{1}{2\pi}\frac{ec^2}{E_{\mathrm{s}}(t)}\frac{\rho}{R_{\mathrm{s}}}B(t)\,. \tag{26}$$



$$\begin{aligned} \mathrm{d}s_0 &= \rho\,\mathrm{d}\theta\,, \\ \mathrm{d}s &= (\rho + x)\,\mathrm{d}\theta\,. \end{aligned}$$

Fig. 5: Orbit length change

Since $E^2 = E_0^2 + p^2c^2$, the RF must follow the variation of the B field with the law

$$\frac{f_{\mathrm{RF}}}{h} = \frac{c}{2\pi R_{\mathrm{s}}} \left\{ \frac{B(t)^2}{(m_0c^2/ec\rho)^2 + B(t)^2} \right\}^{1/2} . \tag{27}$$

This asymptotically tends towards $f_{\mathrm{r}} \to c/(2\pi R_{\mathrm{s}})$ when $v \to c$ and B becomes large compared with $m_0c^2/(ec\rho)$.

3.1 Dispersion effects in a synchrotron

If a particle is slightly shifted in momentum, it will have a different velocity and also a different orbit and orbit length. We can define two parameters.

– The *momentum compaction factor* α_{c}, which is the relative change in orbit length with momentum:

$$\alpha_{\mathrm{c}} = \frac{\Delta L/L}{\Delta p/p}\,. \tag{28}$$

– The *slip factor* η, which is the relative change in revolution frequency with momentum (the slip factor is sometimes also defined in the literature with the opposite sign):

$$\eta = \frac{\Delta f_{\mathrm{r}}/f_{\mathrm{r}}}{\Delta p/p}\,. \tag{29}$$

Let us consider the change in orbit length (see Fig. 5). The relative elementary pathlength difference for a particle with a momentum $p + \mathrm{d}p$ is

$$\frac{\mathrm{d}l}{\mathrm{d}s_0} = \frac{\mathrm{d}s - \mathrm{d}s_0}{\mathrm{d}s_0} = \frac{x}{\rho} = \frac{D_x}{\rho}\frac{\mathrm{d}p}{p}\,, \tag{30}$$

where $D_x = \mathrm{d}x/(\mathrm{d}p/p)$ is the *dispersion function* from the transverse beam optics.

This leads to a total change in the circumference L of

$$\mathrm{d}L = \int_C \mathrm{d}l = \int \frac{x}{\rho}\,\mathrm{d}s_0 = \int \frac{D_x}{\rho}\frac{\mathrm{d}p}{p}\,\mathrm{d}s_0\,, \tag{31}$$

so that

$$\alpha_{\mathrm{c}} = \frac{1}{L}\int \frac{D_x}{\rho}\,\mathrm{d}s_0\,. \tag{32}$$

Since $\rho = \infty$ in the straight sections, we get

$$\alpha_{\mathrm{c}} = \frac{\langle D_x \rangle_{\mathrm{m}}}{R}\,, \tag{33}$$

where the average $\langle \quad \rangle_\mathrm{m}$ is considered over the bending magnets only.

Given that the revolution frequency is $f_\mathrm{r} = \beta c/2\pi R$, the relative change is (using the definition of the momentum compaction factor)

$$\frac{\mathrm{d}f_\mathrm{r}}{f_\mathrm{r}} = \frac{\mathrm{d}\beta}{\beta} - \frac{\mathrm{d}R}{R} = \frac{\mathrm{d}\beta}{\beta} - \alpha_\mathrm{c}\frac{\mathrm{d}p}{p}\,, \tag{34}$$

$$p = mv = \beta\gamma\frac{E_0}{c} \Rightarrow \frac{\mathrm{d}p}{p} = \frac{\mathrm{d}\beta}{\beta} + \frac{\mathrm{d}(1-\beta^2)^{-1/2}}{(1-\beta^2)^{-1/2}} = \underbrace{\left(1-\beta^2\right)^{-1}}_{\gamma^2}\frac{\mathrm{d}\beta}{\beta}\,. \tag{35}$$

Thus, the relative change in revolution frequency is given by

$$\frac{\mathrm{d}f_\mathrm{r}}{f_\mathrm{r}} = \left(\frac{1}{\gamma^2} - \alpha_\mathrm{c}\right)\frac{\mathrm{d}p}{p}\,, \tag{36}$$

which means that the slip factor η is given by

$$\eta = \frac{1}{\gamma^2} - \alpha_\mathrm{c}\,. \tag{37}$$

Obviously, there is one energy with a given γ_tr for which η becomes zero, meaning that there is no change of the revolution frequency for particles with a small momentum deviation. This energy is a property of the transverse lattice, with

$$\gamma_\mathrm{tr} = \frac{1}{\sqrt{\alpha_\mathrm{c}}}\,. \tag{38}$$

From the definition of η, it is clear that an increase in momentum gives the following.

- *Below transition* energy ($\eta > 0$): a higher revolution frequency. The increase of the velocity of the particle is the dominating effect.
- *Above transition* energy ($\eta < 0$): a lower revolution frequency. The particle has a velocity close to the speed of light; this velocity does not change significantly any more. Thus, here the effect of the longer pathlength dominates (for the most common case of transverse lattices with a positive momentum compaction factor, $\alpha_\mathrm{c} > 0$).

At transition, the velocity change and the pathlength change with momentum compensate each other, so the revolution frequency there is independent of the momentum deviation. As a consequence, the longitudinal oscillation stops and the particles in the bunch will not change their phase. Particles that are not at the synchronous phase will get the same non-nominal energy gain each turn and will accumulate an energy error that will increase the longitudinal emittance and can lead to a loss of the particle due to dispersive effects. Thus, transition has to be passed quickly to minimize the emittance increase and the losses.

Electron synchrotrons do not need to cross transition. Owing to the relatively small rest mass of the electron, the relativistic gamma factor is so large that the injection energy is already greater than the transition energy. Hence, the electrons will stay above transition during the whole acceleration cycle.

Since the change of revolution frequency with momentum is opposite below and above transition, this completely changes the range for stable oscillations. As we have seen in the linac case (see Fig. 3), the oscillation is stable for a particle that is on the rising slope of the RF field when we are below transition. Above transition, the oscillation is unstable and the stable region for oscillations is on the falling slope (see Fig. 6). A particle that arrives too early (M_2) will get more energy, and the revolution time will increase, owing to the predominant effect of the longer path. Thus, it will arrive later on the

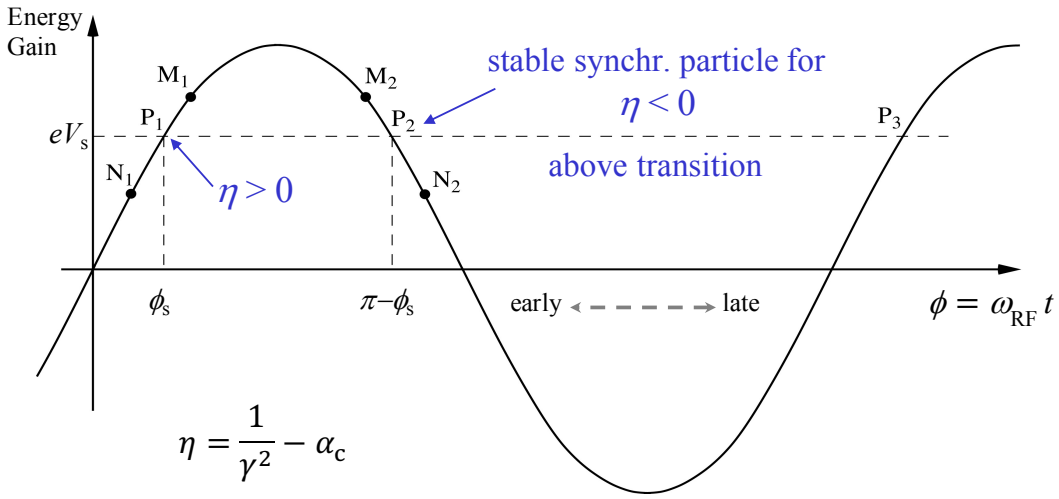


Fig. 6: Energy gain as a function of particle phase. Oscillations are stable around the synchronous phase particle P_1 below transition and around the synchronous phase particle P_2 above transition.

next turn, closer to the synchronous phase. Similarly, a particle that arrives late (N_2) will gain less energy and travel a shorter orbit, also moving towards the synchronous phase.

Crossing transition during acceleration makes the previous stable synchronous phase unstable. Below transition, it is stable on the rising slope of the RF; above transition, the synchronous phase is on the falling slope. Consequently, the RF system needs to make a rapid change of the RF phase when crossing transition energy; a 'phase jump', as indicated in Fig. 7. Otherwise the particles in the bunch get dispersed, have a wrong energy gain, and eventually get lost. A method to improve transition crossing is to change the transverse optics when the energy almost reaches γ_{tr} for optics with a larger momentum compaction factor α_{c}. Hence, γ_{tr} is decreasing at the same time as the energy is increasing, and the time with an energy close to transition can be reduced.

3.2 Equations of longitudinal motion in a synchrotron

As previously done for the linac, we want to look at the oscillations with respect to the synchronous particle and we express the variables with respect to the synchronous particle, as shown in Fig. 8.

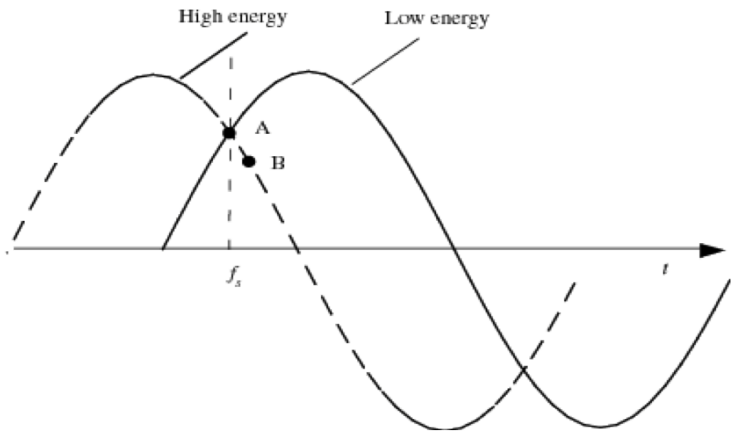


Fig. 7: The synchronous phase changes from rising to falling slopes as the energy crosses transition [1]

particle RF phase:	$\Delta\phi = \phi - \phi_\mathrm{s}$,
particle momentum:	$\Delta p = p - p_\mathrm{s}$,
particle energy:	$\Delta E = E - E_\mathrm{s}$,
azimuth angle:	$\Delta\theta = \theta - \theta_\mathrm{s}$.

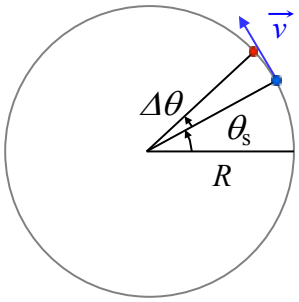


Fig. 8: Variables with respect to synchronous particle

Since the RF is a multiple of the revolution frequency, the RF phase $\Delta\phi$ changes as

$$f_\mathrm{RF} = h\, f_\mathrm{r} \qquad \Rightarrow \qquad \Delta\phi = -h\,\Delta\theta \qquad \text{with} \qquad \theta = \int \omega_\mathrm{r}\mathrm{d}t\,. \tag{39}$$

The minus sign for the RF phases originates from the fact that a particle that is ahead arrives earlier, so at a smaller RF phase.

For a given particle with respect to the reference particle, the change in angular revolution frequency is

$$\Delta\omega_\mathrm{r} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\Delta\theta\right) = -\frac{1}{h}\frac{\mathrm{d}}{\mathrm{d}t}\left(\Delta\phi\right) = -\frac{1}{h}\frac{\mathrm{d}\phi}{\mathrm{d}t}\,. \tag{40}$$

Since

$$\eta = \frac{p_\mathrm{s}}{\omega_\mathrm{rs}}\left(\frac{\mathrm{d}\omega_\mathrm{r}}{\mathrm{dp}}\right)_\mathrm{s}\,,$$

$E^2 = E_0^2 + p^2c^2$, and $\Delta E = v_\mathrm{s}\Delta p = \omega_\mathrm{rs}R_\mathrm{s}\Delta p$, one gets the first-order equation

$$\frac{\Delta E}{\omega_\mathrm{rs}} = -\frac{p_\mathrm{s}R_\mathrm{s}}{h\eta\omega_\mathrm{rs}}\frac{\mathrm{d}(\Delta\phi)}{\mathrm{d}t} = -\frac{p_\mathrm{s}R_\mathrm{s}}{h\eta\omega_\mathrm{rs}}\dot{\phi}\,. \tag{41}$$

The second first-order equation follows from the energy gain of a particle:

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\omega_\mathrm{r}}{2\pi}\,e\hat{V}\sin\phi\,, \tag{42}$$

$$2\pi\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\Delta E}{\omega_\mathrm{rs}}\right) = e\hat{V}\left(\sin\phi - \sin\phi_\mathrm{s}\right). \tag{43}$$

Deriving and combining the two first-order equations (Eqs. (41) and (43)) leads to

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{R_\mathrm{s}p_\mathrm{s}}{h\eta\omega_\mathrm{rs}}\frac{\mathrm{d}\phi}{\mathrm{d}t}\right] + \frac{e\hat{V}}{2\pi}\left(\sin\phi - \sin\phi_\mathrm{s}\right) = 0\,. \tag{44}$$

This second-order equation is non-linear and the parameters within the bracket are, in general, slowly varying in time.

When we assume constant parameters $R_\mathrm{s}, p_\mathrm{s}, \omega_\mathrm{s}$, and η, we get

$$\ddot{\phi} + \frac{\Omega_\mathrm{s}^2}{\cos\phi_\mathrm{s}}\left(\sin\phi - \sin\phi_\mathrm{s}\right) = 0 \qquad \text{with} \qquad \Omega_\mathrm{s}^2 = \frac{h\eta\omega_\mathrm{rs}e\hat{V}\cos\phi_\mathrm{s}}{2\pi R_\mathrm{s}p_\mathrm{s}} \tag{45}$$

and, for small phase deviations from the the synchronous particle,

$$\sin\phi - \sin\phi_\mathrm{s} = \sin(\phi_\mathrm{s} + \Delta\phi) - \sin\phi_\mathrm{s} \approxeq \cos\phi_\mathrm{s}\Delta\phi\,, \tag{46}$$

so that we end up with the equation of a harmonic oscillator:

$$\ddot{\phi} + \Omega_{\mathrm{s}}^2 \Delta\phi = 0\,, \tag{47}$$

where Ω_{s} is the synchrotron angular frequency.

Stability is obtained when Ω_{s} is real so that Ω_{s}^2 is positive. Since most terms in Ω_{s}^2 are positive, this reduces to

$$\eta \cos\phi_{\mathrm{s}} > 0 \tag{48}$$

and the stable region for the synchronous phase depends on the energy with respect to the transition energy, as we have seen from our argument before. The conditions for stability are summarized in Fig. 9.

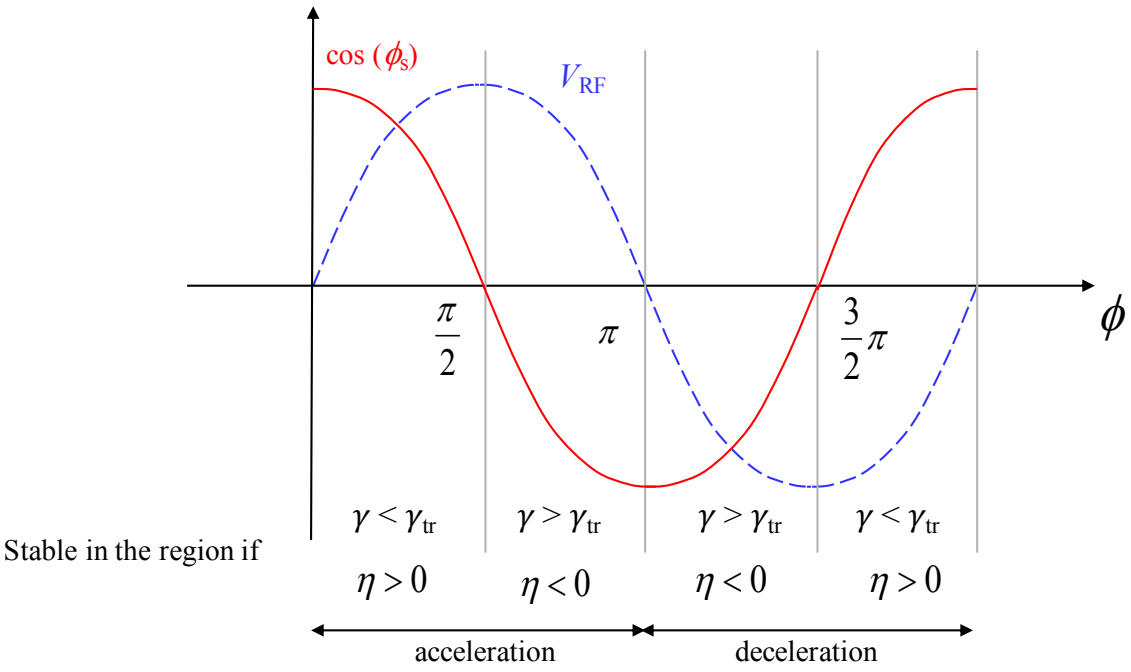


Fig. 9: Stability regions as a function of particle phase, depending on the energy with respect to transition

The *synchrotron tune* ν_{s} is defined as

$$\nu_{\mathrm{s}} = \Omega_{\mathrm{s}}/\omega_{\mathrm{r}}\,, \tag{49}$$

and corresponds to the number of synchrotron oscillations per revolution in the synchrotron. It is generally $\nu_{\mathrm{s}} \ll 1$, as it typically takes of the order of several hundreds of turns to complete one synchrotron oscillation.

For larger phase (or energy) deviations from the synchronous particle, we can multiply Eq. (45) by $\dot{\phi}$ and integrate it, getting an invariant of motion:

$$\frac{\dot{\phi}^2}{2} - \frac{\Omega_{\mathrm{s}}^2}{\cos\phi_{\mathrm{s}}}\left(\cos\phi + \phi\sin\phi_{\mathrm{s}}\right) = I\,, \tag{50}$$

which, for small amplitudes for $\Delta\phi$, reduces to

$$\frac{\dot{\phi}^2}{2} + \Omega_{\mathrm{s}}^2\,\frac{(\Delta\phi)^2}{2} = I'\,. \tag{51}$$

Similar equations exist for the second variable $\Delta E \propto \mathrm{d}\phi/\mathrm{d}t$.

As we have seen before, the restoring force goes to zero when ϕ reaches $\pi - \phi_\mathrm{s}$ and becomes non-restoring beyond (both below and above transition); see Fig. 10. Hence, $\pi - \phi_\mathrm{s}$ is an extreme amplitude for a stable motion, which has a closed trajectory in phase space. This phase-space trajectory separates the region of stable motion from the unstable region. It is called the *separatrix*. The area within this separatrix is called the *RF bucket*, which corresponds to the maximum acceptance in phase space for a stable motion.

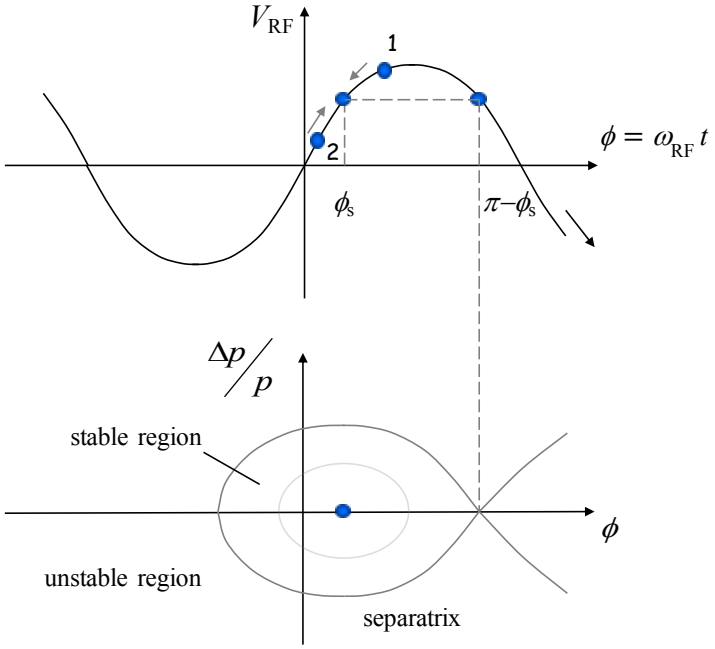


Fig. 10: Top: RF voltage as a function of particle phase. Bottom: phase-space picture

Since we have found an invariant of motion, we can write the equation for the separatrix by calculating it at the phase $\phi = \pi - \phi_\mathrm{s}$, where $\dot{\phi} = 0$:

$$\frac{\dot{\phi}^2}{2} - \frac{\Omega_\mathrm{s}^2}{\cos\phi_\mathrm{s}} \left(\cos\phi + \phi \sin\phi_\mathrm{s}\right) = -\frac{\Omega_\mathrm{s}^2}{\cos\phi_\mathrm{s}} \left(\cos(\pi - \phi_\mathrm{s}) + (\pi - \phi_\mathrm{s}) \sin\phi_\mathrm{s}\right) . \tag{52}$$

From this, we can calculate the second value, ϕ_m, where the separatrix crosses the horizontal axis, which is the other extreme phase for stable motion:

$$\cos\phi_\mathrm{m} + \phi_\mathrm{m} \sin\phi_\mathrm{s} = \cos(\pi - \phi_\mathrm{s}) + (\pi - \phi_\mathrm{s}) \sin\phi_\mathrm{s} . \tag{53}$$

It can be seen from the equation of motion that $\dot{\phi}$ reaches an extreme when $\ddot{\phi} = 0$, corresponding to $\phi = \phi_\mathrm{s}$. Putting this value into Eq. (52) gives

$$\dot{\phi}_\mathrm{max}^2 = 2\,\Omega_\mathrm{s}^2 \left[2 + (2\phi_\mathrm{s} - \pi) \tan\phi_\mathrm{s}\right] , \tag{54}$$

which translates into an acceptance in energy

$$\left(\frac{\Delta E}{E_\mathrm{s}}\right)_\mathrm{max} = \pm\beta \sqrt{-\frac{e\hat{V}}{\pi h \eta E_\mathrm{s}} G(\phi_\mathrm{s})} , \tag{55}$$

where

$$G(\phi_\mathrm{s}) = 2\cos\phi_\mathrm{s} + (2\phi_\mathrm{s} - \pi) \sin\phi_\mathrm{s} . \tag{56}$$

This *RF acceptance* depends strongly on ϕ_s and plays an important role in the capture at injection and the stored beam lifetime. The maximum energy acceptance in the bucket depends on the square root of the available RF voltage, $\hat{V}_\mathrm{RF}$. The phase extension of the bucket is a maximum for $\phi_\mathrm{s} = 0^\circ$ or 180°. As the synchronous phase approaches 90°, the bucket size becomes smaller, as illustrated in Fig. 11.

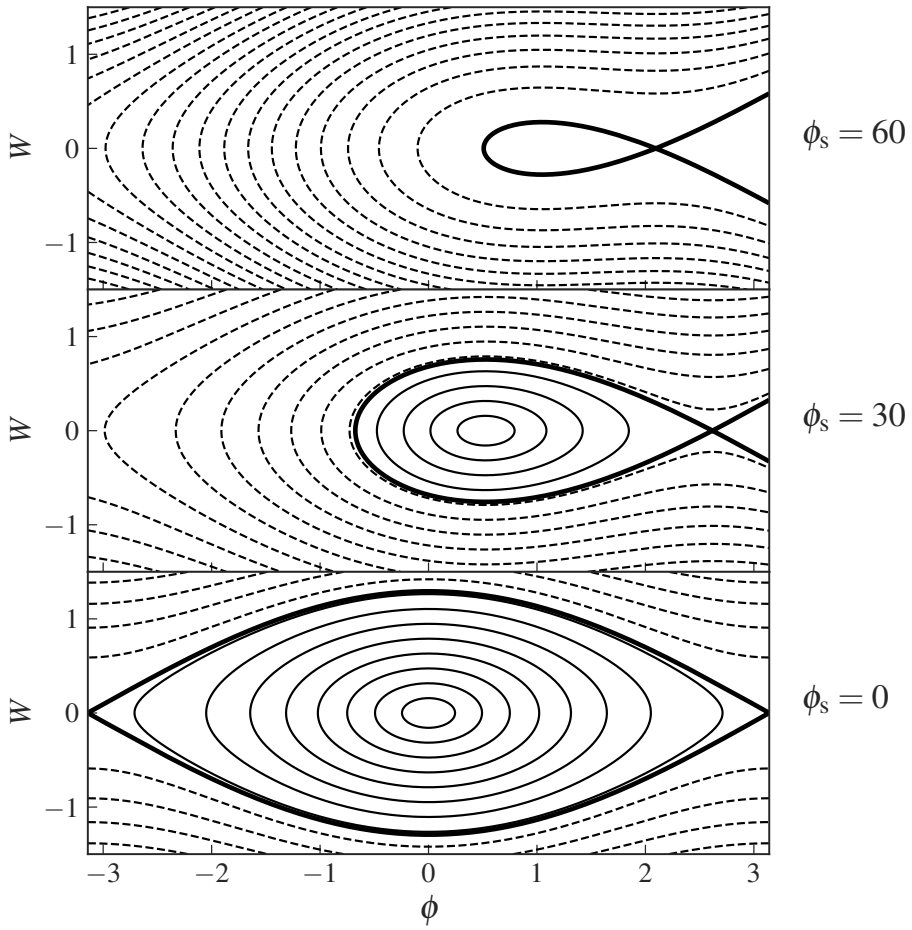


Fig. 11: Phase-space plots for different synchronous phase angles ϕ_s for otherwise identical parameters: thin solid lines represent stable trajectories in phase space, dashed lines represent unstable trajectories, and the thick solid line is the separatrix.

3.3 Stationary bucket

In the case of the stationary bucket, we have no acceleration and $\sin\phi_\mathrm{s} = 0$, so that $\phi_\mathrm{s} = 0$ or π. The equation of the separatrix for $\phi_\mathrm{s} = \pi$ (above transition) simplifies to

$$\frac{\dot{\phi}^2}{2} + \Omega_\mathrm{s}^2 \cos\phi = \Omega_\mathrm{s}^2 \qquad \text{or} \qquad \frac{\dot{\phi}^2}{2} = 2\,\Omega_\mathrm{s}^2 \sin^2\frac{\phi}{2}\,. \tag{57}$$

At this point, it is convenient to introduce a new variable W to replace the phase derivative $\dot{\phi}$,

$$W = \frac{\Delta E}{\omega_\mathrm{rs}} = -\frac{p_\mathrm{s} R_\mathrm{s}}{h\eta\omega_\mathrm{rs}}\dot{\phi}\,, \tag{58}$$

where ω_rs is the revolution frequency of the synchronous particle. As we see later, this new variable is canonical. Different choices of canonical variables are possible and lead to slightly different equations, as, for example, in Ref. [2].

Introducing Ω_s^2 from Eq. (45) leads to the following equation for the separatrix:

$$W = \pm\frac{C}{\pi c}\sqrt{\frac{-e\hat{V}E_\mathrm{s}}{2\pi h\eta}}\,\sin\frac{\phi}{2} = \pm W_\mathrm{bk}\sin\frac{\phi}{2} \qquad \text{with} \qquad W_\mathrm{bk} = \frac{C}{\pi c}\sqrt{\frac{-e\hat{V}E_\mathrm{s}}{2\pi h\eta}}\,. \tag{59}$$

Setting $\phi = \pi$ in the previous equation shows that W_bk is the maximum height of the bucket, which results in the maximum energy acceptance:

$$\Delta E_\mathrm{max} = \omega_\mathrm{rs} W_\mathrm{bk} = \beta_\mathrm{s}\sqrt{2\frac{-e\hat{V}_\mathrm{RF}E_\mathrm{s}}{\pi h\eta}}\,. \tag{60}$$

The bucket area is

$$A_\mathrm{bk} = 2\int_0^{2\pi} W\,\mathrm{d}\phi\,. \tag{61}$$

With $\int_0^{2\pi}\sin(\phi/2)\,\mathrm{d}\phi = 4$, one gets

$$A_\mathrm{bk} = 8W_\mathrm{bk} = 8\frac{C}{\pi c}\sqrt{\frac{-e\hat{V}E_\mathrm{s}}{2\pi h\eta}}\,. \tag{62}$$

3.4 Bunch matching into the stationary bucket

We can describe the motion of a particle inside the separatrix of the stationary bucket starting from the invariant of motion from Eq. (50) and setting $\phi_\mathrm{s} = \pi$:

$$\frac{\dot{\phi}^2}{2} + \Omega_\mathrm{s}^2\,\cos\phi = I\,. \tag{63}$$

The points ϕ_m and $2\pi - \phi_\mathrm{m}$ where the trajectory crosses the horizontal axis are symmetrical with respect to $\phi_\mathrm{s} = \pi$ (see Fig. 12). We can calculate the invariant for $\phi = \phi_\mathrm{m}$ and get

$$\frac{\dot{\phi}^2}{2} + \Omega_\mathrm{s}^2\,\cos\phi = \Omega_\mathrm{s}^2\,\cos\phi_\mathrm{m}, \tag{64}$$

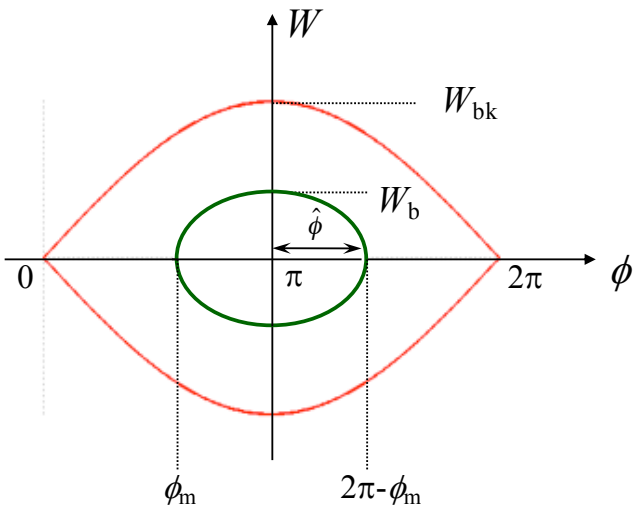


Fig. 12: Phase-space plot for the separatrix of the stationary bucket and a trajectory inside

$$\dot{\phi} = \pm\Omega_\mathrm{s}\sqrt{2\,(\cos\phi_\mathrm{m} - \cos\phi)}, \tag{65}$$

$$W = \pm W_\mathrm{bk}\sqrt{\cos^2\frac{\phi_\mathrm{m}}{2} - \cos^2\frac{\phi}{2}} \qquad \left(\text{using} \quad \cos\phi = 2\cos^2\frac{\phi}{2} - 1\right). \tag{66}$$

Setting $\phi = \pi$ in the previous equation allows us to calculate the bunch height W_b:

$$W_\mathrm{b} = W_\mathrm{bk}\,\cos\frac{\phi_\mathrm{m}}{2} = W_\mathrm{bk}\,\sin\frac{\hat{\phi}}{2}\,, \tag{67}$$

with $\hat{\phi} = \pi - \phi_\mathrm{m}$ being the maximum phase amplitude for an oscillation around the synchronous phase $\phi_\mathrm{s} = \pi$.

The corresponding maximum energy difference of a particle on this phase-space trajectory is

$$\left(\frac{\Delta E}{E_\mathrm{s}}\right)_\mathrm{b} = \left(\frac{\Delta E}{E_\mathrm{s}}\right)_\mathrm{RF}\cos\frac{\phi_\mathrm{m}}{2} = \left(\frac{\Delta E}{E_\mathrm{s}}\right)_\mathrm{RF}\sin\frac{\hat{\phi}}{2}\,. \tag{68}$$

When a particle bunch is injected into a synchrotron, the bunch has a given bunch length and energy spread. The different particles will move along phase-space trajectories that correspond to their initial phase and energy. If the shape of the injected bunch in phase space matches the shape of a phase-space trajectory for the given RF parameters, the shape of the bunch in phase space will be maintained.

If the shape does not match, it will vary during the synchrotron period. This is illustrated in Fig. 13 for a bunch that has a shorter bunch length and a larger energy spread compared with the phase-space trajectory. As the particles move along their individual trajectories, the bunch will be longer with a smaller energy spread after one-quarter of a synchrotron period, and will regain the initial shape after one-half of a period. This effect can be used to manipulate the shape of the bunch in phase space and trade off bunch length against energy spread (so-called *bunch rotation*). When the RF voltage in matched conditions is suddenly increased, the bunch will be shorter after a quarter of a synchrotron period. In this way, it can be shortened for a transfer to a higher-frequency RF system.

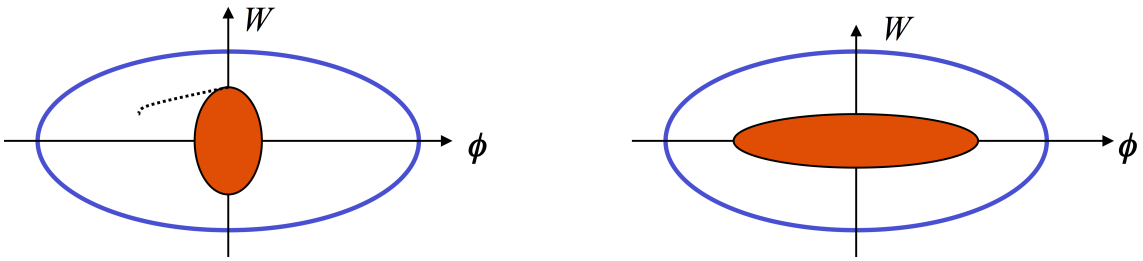


Fig. 13: Phase-space plots for a mismatched bunch one-quarter of a synchrotron period apart

Owing to the non-linear restoring force, the synchrotron period depends on the oscillation amplitude, and particles with larger amplitudes have a longer synchrotron period, as shown in Fig. 14. This will eventually lead to a *filamentation* of the bunch and an increase of the longitudinal emittance.

The same phenomenon will happen when the bunch shape is matched to the bucket but there is an error in the phase or the energy. The different particles in the bunch will perform their individual oscillations around the synchronous particle and will filament, leading to an increase in the longitudinal emittance. To avoid an emittance increase, it is important to match phase, energy, and the shape of the bunch and the bucket during the transfer.

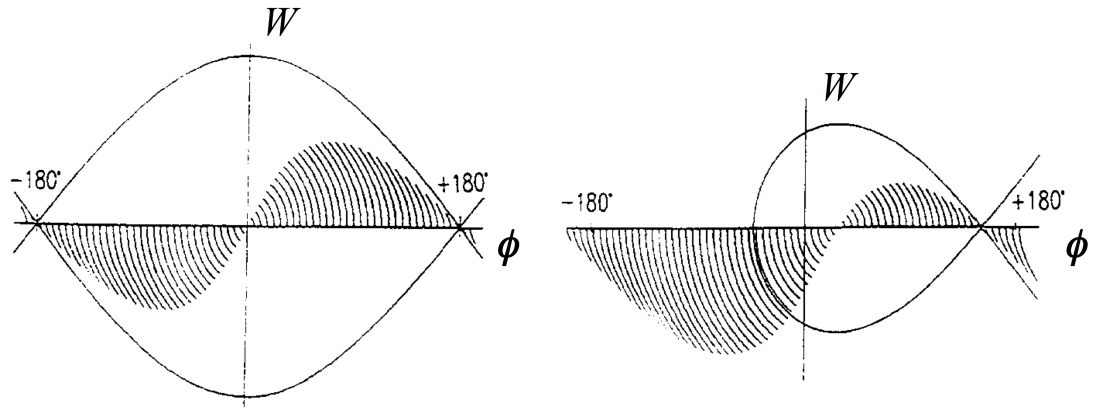


Fig. 14: Phase-space trajectories for one-eighth of the synchrotron period: left, stationary bucket; right, accelerating bucket [3].

3.5 Potential energy function and Hamiltonian

The longitudinal motion is produced by a force that can be derived from a scalar potential U:

$$\frac{\mathrm{d}^2\phi}{\mathrm{d}t^2} = F(\phi) = -\frac{\partial U}{\partial \phi}\,, \tag{69}$$

$$U(\phi) = -\int_0^{\phi} F(\phi)\,\mathrm{d}\phi = -\frac{\Omega_\mathrm{s}^2}{\cos\phi_\mathrm{s}}\left(\cos\phi + \phi\sin\phi_\mathrm{s}\right) - F_0\,. \tag{70}$$

The sum of the potential energy and the kinetic energy is constant and, by analogy, represents the total energy of a non-dissipative system:

$$\frac{\dot{\phi}^2}{2} + U(\phi) = F_0\,. \tag{71}$$

Since the total energy is conserved, we can describe the system as a Hamiltonian system. Different choices of the canonical variables are possible. With the variable

$$W = \frac{\Delta E}{\omega_\mathrm{rs}}\,, \tag{72}$$

the two first-order equations of the longitudinal motion become

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = -\frac{h\eta\omega_\mathrm{rs}}{pR}W\,, \tag{73}$$

$$\frac{\mathrm{d}W}{\mathrm{d}t} = \frac{e\hat{V}}{2\pi}\left(\sin\phi - \sin\phi_\mathrm{s}\right)\,. \tag{74}$$

The two variables ϕ and W are canonical, since these equations of motion can be derived from a Hamiltonian $H(\phi, W, t)$:

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = \frac{\partial H}{\partial W}, \qquad \frac{\mathrm{d}W}{\mathrm{d}t} = -\frac{\partial H}{\partial \phi}\,, \tag{75}$$

$$H(\phi, W, t) = \frac{e\hat{V}}{2\pi}\left[\cos\phi - \cos\phi_\mathrm{s} + (\phi - \phi_\mathrm{s})\sin\phi_\mathrm{s}\right] - \frac{1}{2}\frac{h\eta\omega_\mathrm{rs}}{pR}W^2\,. \tag{76}$$

The basic Hamiltonian shown here reproduces the equations of motions that we found before. In more complex cases, the general approach of the Hamiltonian formalism helps to treat and understand some fairly complicated dynamics (multiple harmonics, bunch splitting, etc.).

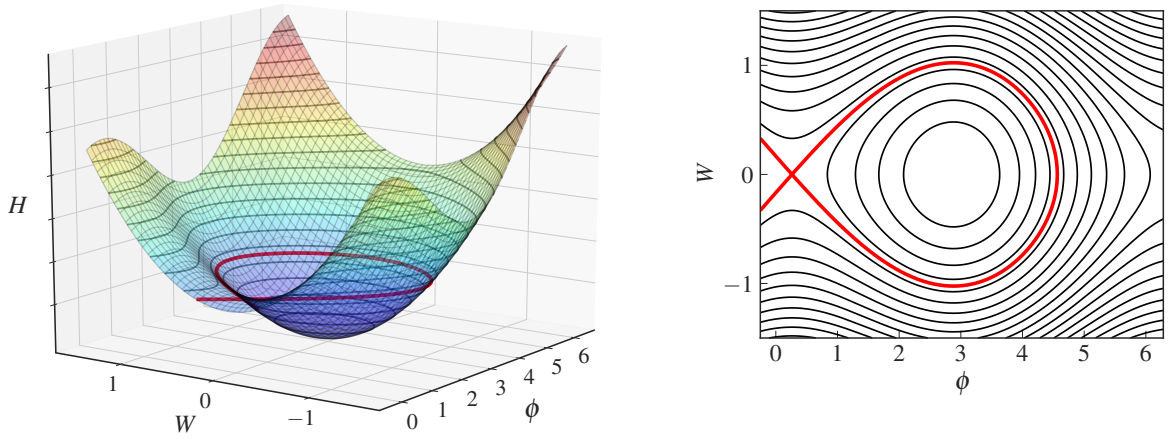


Fig. 15: Hamiltonian function H (left) and its projection (right) above transition, for a synchronous phase of $\phi_s = 170°$. Equipotential lines are possible phase-space trajectories. The separatrix is shown as a solid red line.

The Hamiltonian H represents the total energy of the system. In fact, if the total energy is conserved, the contours of constant H are particle trajectories in phase space, as illustrated in an example in Fig. 15.

Acknowledgement

I thank, in particular, Joël Le Duff, who gave this lecture series in the past, and allowed me to use the material of his lectures, from which I profited greatly. Many of the figures and derivations of this paper originate from him.

References

[1] E.J.N. Wilson, *An Introduction to Particle Accelerators* (Oxford University Press, Oxford, 2001).

[2] F. Tecker, in Proceedings of the CAS-CERN Accelerator School: Advanced Accelerator Physics, Trondheim, Norway, 19-29 August 2013, edited by W. Herr, CERN-2014-009 (CERN, Geneva, 2014), pp.1-21, https://doi.org/10.5170/CERN-2014-009.1.

[3] W. Pirkl, in Proceedings of the CAS-CERN Accelerator School: CAS Fifth Advanced Accelerator Physics Course, Rhodes, Greece, 20 Septempber-1 October 1993, edited by S. Turner, CERN 95-06, (CERN, Geneva, 1995), pp. 233-257, https://doi.org/10.5170/CERN-1995-006.233.

Bibliography

J. Le Duff, in Proceedings of the CAS-CERN Accelerator School: CAS Fifth General Accelerator Physics Course, Jyväskylä, Finland, 7-18 September 1992, edited by S. Turner, CERN 94-01, (CERN, Geneva, 1994), pp. 253-288, https://doi.org/10.5170/CERN-1994-001.253.

J. Le Duff, in Proceedings of the CAS-CERN Accelerator School: CAS Fifth General Accelerator Physics Course, Jyväskylä, Finland, 7-18 September 1992, edited by S. Turner, CERN 94-01, (CERN, Geneva, 1994), pp. 289-311, https://doi.org/10.5170/CERN-1994-001.289.

H. Wiedemann, *Particle Accelerator Physics* (Springer, Berlin, 2007).

K. Wille, *The Physics of Particle Accelerators: An Introduction* (Oxford University Press, Oxford, 2000).

T. Wangler, *RF Linear Accelerators* (Wiley-VCH, Weinheim, 2008).

Proceedings of the CAS–CERN Accelerator School: Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018)

Particle Interactions with Matter

A. Lechner
CERN, Geneva, Switzerland

Abstract
With steady advances in accelerator technology and new records in beam energy, intensity, brightness, power, and luminosity, it is essential to limit the adverse effects of beam losses for accelerator operation, equipment, and personnel. Particle losses can quench superconducting magnets, induce single-event effects in electronics, activate equipment, and alter material properties. In the worst case, high-energy particle beams can cause severe damage to equipment, if they are lost in an uncontrolled way. Modern accelerators typically rely on a system of protection absorbers, dumps, and beam manipulation devices to safely accelerate, store, extract, and transfer beams. Knowledge of particle–matter interactions is important to understand and predict the consequences of beam losses and to design beam-intercepting devices, shielding, and other equipment. This paper summarizes the principal physical processes of hadrons, photons, and leptons in matter and discusses related phenomena, such as particle showers.

Keywords
Particle–matter interactions; particle showers; energy deposition.

1 Introduction

All particles of a beam or their secondary interaction or decay products will eventually interact with surrounding matter. Depending on the particle type and energy, swift particles penetrating in a material will be subject to different atomic and nuclear processes. The complexity of particle–matter interactions can be illustrated for the example of high-energy protons: if high-energy protons ($\geq$GeV) impact on a material, they are subject to inelastic nuclear collisions, producing, on average, a number of fast secondary hadrons; some of these hadrons (protons, neutrons, charged pions) can have further nuclear collisions, resulting in a hadronic cascade; others (neutral pions) will decay almost instantly and their decay products (γ-pairs) can initiate electromagnetic showers, which are sustained by electron–positron pair production by photons and Bremsstrahlung losses of electrons and positrons. As the particle energy decreases, other processes become dominant (Compton scattering of photons, photoelectric effect, ionization losses of charged particles). Some of the charged pions and other hadrons will also decay and produce muons. Besides the production of fast hadrons, nuclear collisions also give rise to lower-energy (MeV) neutrons, protons, light ions (alpha particles) and gamma rays, which are emitted during the de-excitation of target nuclei. The protons and light ions will mainly range out because of ionizing losses. The neutrons can, however, be very penetrating and will mainly undergo elastic nuclear collisions until they thermalize and are captured. Eventually, the residual nuclei of elastic and inelastic collisions can also recoil and impart their energy on neighbouring nuclei through Coulomb interactions. This description provides only a glimpse of the copious particle production mechanisms in high-energy particle interactions in matter.

The interactions of particles in accelerator equipment can have many different negative consequences, from instantaneous damage or long-term damage of equipment because of shock heating or atomic displacements to the activation of equipment because of radionuclei production in nuclear collisions. Knowledge of particle–matter interactions is therefore essential for designing and operating new accelerators in order to mitigate consequences for equipment and personnel. This paper provides a very basic introduction to beam–matter interactions for a non-expert audience working in the field of

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-005.47

particle accelerators who seeks to get a first understanding of the topic. The paper aims to convey a basic overview of the relevant physical processes and related phenomena, such as particle showers, without an in-depth discussion of the underlying theories. It introduces commonly used definitions and quantities and describes the dependencies of processes and energy deposition profiles on particle energy and target material. The paper is partially based on previous papers by Ferrari [1,2]. Particle–matter interactions in accelerators have also been discussed in another paper by Mokhov and Cerutti [3]. The reader is referred to these papers for additional and complementary information. In addition, a wealth of literature is available, where the physical processes and in particular the underlying theories and models are discussed in much greater detail (see, for example, Refs. [4–7]). The paper is organized as follows. Section 2 gives a brief introduction of beam-intercepting devices, sources of secondary particles, and consequences of particle–matter interactions. Section 3 summarizes the properties of relevant particles and discusses the relation between cross-section, mean free path, and interaction probability. Sections 4 and 5 provide a basic account of atomic and nuclear processes and Section 6 concludes the paper, with a discussion of energy deposition profiles and particle showers.

2 Beam–machine interaction

2.1 Particle impact on beam-intercepting devices

Particle accelerators employ a variety of beam-intercepting devices, such as collimators, scrapers, dumps, stripping foils, crystals, or targets. These devices serve different purposes, such as machine protection, reduction of activation in certain areas, beam extraction, or particle production for experiments. A few examples are listed in the following.

- *Beam disposal on dumps.* In case of a beam abort, which can be deliberate or due to an emergency situation, the beam is steered onto a beam dump or a beam stopper, which should absorb a large fraction of the energy carried by the beam particles. Beam dumps must be robust enough to sustain the impact of full bunch trains stored in a machine or transferred between machines. Design features of a dump, such as the choice of absorber materials, dimensions, cooling circuits, etc., depend on the particle type and energy, as well as on operational parameters, such as the beam intensity, the beam brightness, and the dumping frequency. In high-energy hadron colliders, such as the LHC, beam dumps can be as long as 10 m and need to be surrounded by heavy shielding to reduce the activation of the environment [8]. In addition, the LHC beams need to be swept transversally across the dump front face by dilution kickers to avoid excessive temperatures and stresses in the graphitic dump core [8].
- *Machine protection and beam manipulation.* Beam particles can be intercepted by protection or beam manipulation devices, such as collimators, absorbers, or scrapers, which are usually placed in close proximity to the beam. Such devices are employed for different reasons, for example, to remove unwanted halo particles before they are lost in sensitive locations, such as superconducting magnets; to protect the machine in case of equipment malfunction (e.g., mis-steering of the beam during injection or asynchronous beam dumps); to reduce the background to experiments; or to concentrate losses in a certain location for radioprotection reasons. The latter can be crucial to limit the exposure of personnel during technical interventions.
- *Beam extraction.* The physical interaction of beam particles with different devices is exploited in certain extraction techniques. Beam particles can, for example, be extracted by changing their charge using a stripping foil. In other cases, bent crystals are used to selectively steer beam particles. Charged particles impacting on a crystal can be subject to different coherent phenomena, such as channelling or volume reflection, which have been successfully exploited for beam manipulation and extraction. Particles that do not have the right impact angle with respect to the crystal planes or that are de-channelled, interact with the crystal, as they would with an amorphous absorber.

– *Particle production in targets.* In many cases, beam particles are directed onto a fixed target to produce secondary particles for physics experiments or diagnostic purposes. The target design requirements depend on the specific use case, in particular on the type and spectrum of secondary particles to be produced. A typical example is the production of neutrons in spallation targets. Spallation targets can be solid (e.g., lead) or liquid (e.g., mercury) and are usually combined with a moderator (e.g., heavy water) to manipulate the neutron spectrum.

2.2 Sources of secondary particles

Not only beam particles but also secondary particles might interact with surrounding equipment. The production of secondary particles can be localized or spread across a machine. A few examples are given here.

– *Luminosity production.* Secondary particles are produced in the interaction points of a collider, where two beams are brought into collision. A fraction of the collision products may escape the experimental cavern and can interact with nearby accelerator equipment [9, 10].
– *Beam–gas collisions.* Other sources of secondary particles are interactions of beam particles with residual gas molecules in the vacuum chamber. Hadron beams can have inelastic nuclear encounters with gas nuclei, resulting in the production of secondary hadrons, muons, or photons, which can, for example, be a source of background to experiments [11]. Electron or positron beams can be subject to Bremsstrahlung losses.
– *Dust or macroparticle-induced losses.* Beam losses can also be induced by dust particles liberated into the beams. First observed in electron storage rings, where ionized dust particles were found to be trapped in the beam [12], they have also been observed in the LHC [13, 14], where they triggered many beam aborts and even magnet quenches. In electron machines, dust trapping can lead to Bremsstrahlung losses in the field of a charged macroparticle; this is generally referred to as duststrahlung [12].
– *Synchrotron radiation.* Synchrotron radiation is emitted if charged particles are bent by a magnet or if they are accelerated. Synchrotron radiation can be a significant source of beam power loss in circular electron and positron machines.

2.3 Consequences of particle–matter interactions for operation, equipment, and personnel

Particle interactions with matter can have negative consequences for an accelerator, leading to disruptions of machine operation or causing long-term effects. A few particle-induced effects are summarized in the following.

– *Activation of equipment.* Nuclear interactions lead to the activation of accelerator equipment [15]. Radioisotopes are not only produced in hadron machines, but also in electron accelerators because of photonuclear and electronuclear interactions. The types of radionuclides produced in a material depend on the elemental composition and decay chains. Activation of equipment may require access restrictions or dedicated cooling times before technical intervention. Radioactive isotopes can also be produced in the surrounding environment (soil, air, ground water), which may require special monitoring.
– *Long-term radiation damage of equipment.* Long-time exposure to radiation can lead to the embrittlement and swelling of materials and may degrade physical material properties, such as electrical or thermal conductivity. Radiation damage in non-organic materials is generally related to gas production (H, He) in nuclear reactions and to the displacement of atoms from the lattice due to the non-ionizing energy loss of particles. Atomic displacements depend on the damage threshold of a material. Organic materials can change their properties if they receive a large ionizing dose.

Organic insulators are, for example, used in some magnet coils and adequate shielding might be necessary in high-radiation areas to avoid degradation of the material properties.

– *Instantaneous damage of equipment because of thermal shock.* In the case of an uncontrolled fast beam loss on the aperture, the stresses generated by beam-induced thermal shock can lead to instantaneous material damage. Damage can occur when the energy densities reach the order of kJ/cm^3 (depending on the material properties and the spatial energy density distribution). If the beam-induced energy density is high enough, materials can even be subject to a phase transition (melting). Equipment damage can be very disruptive for operation and may lead to weeks of downtime for repairs. Particle beams in a high-energy collider, like the LHC, with a stored energy of hundreds of MJ, have an enormous damage potential. The LHC employs a sophisticated machine protection system, which extracts the beams on the dump block if equipment is found to be malfunctioning or if beam losses measured by beam loss monitors exceed a certain threshold [8]. The LHC employs, in addition, a multistage collimation system with robust collimators made of carbon-reinforced carbon, which provide another layer of protection [8].
– *Quenches of superconducting magnets.* If the beam-induced energy deposition in the coils of a superconducting magnet exceeds a certain level, the magnet can quench, i.e., it can become normally conducting. The minimum energy density to induce a quench depends on the beam loss duration, the coil geometry, the superconductor type, the magnet current, and the operation temperature. Quench levels are orders of magnitude less ($\sim$mJ/cm^3) than the damage limit of coils. A magnet quench can lead to a machine downtime of several hours. In high-energy machines, such as the LHC, a small fraction of beam particles ($\sim 10^{-7}$) can induce a quench if lost on the magnet aperture [16]. The risk of beam-induced magnet quenches must be assessed in the design phase of a superconducting machine and appropriate measures must be implemented (e.g., collimators, shielding inserts in magnets, beam loss monitors).
– *Radiation effects in electronics.* Electronics exposed to radiation fields can suffer single-event effects, which can be non-destructive or destructive. Even if non-destructive, single-event effects can limit the performance of an accelerator by triggering beam aborts [17]. Over longer periods, electronics can also be compromised by cumulative displacement damage or the total ionizing dose deposited by charged particles.

3 Particles and interaction probability

3.1 Particle types and properties

If a beam particle interacts with matter, secondary particles can be produced in atomic and nuclear processes, as illustrated in the example given in Section 1. Table 1 summarizes the mass, mean life τ, and main decay mode of the most relevant particle species found in hadron and lepton accelerator environments. Photons are present in high abundance in most machines. They are produced in Bremsstrahlung processes of charged particles, in electron–positron annihilation, in gamma de-excitation after nuclear reactions, in radiative neutron capture, or in the decay of particles (e.g., π^0 decay). The most common hadrons are protons, neutrons, pions, and also other species, such as kaons (the latter are not included in the table). Hadrons can be produced in inelastic hadron–nucleus collisions, but also in other processes, such as photonuclear reactions. Neutrons can be very penetrating, since they are neutral and have a relatively long mean lifetime of almost 15 min. Pions are the lightest hadrons and hence are produced with high multiplicities in nuclear collisions. Although charged pions are unstable, their mean range is sufficiently long that they typically reinteract in a material before decaying. Neutral pions have a much shorter lifetime and range and they decay predominantly into a γ-pair. They are the main source of electromagnetic showers in high-energy hadron machines. The most copiously produced leptons are electrons, positrons, and muons. Electrons are released in Coulomb interactions of charged particles and in various photon processes (photoelectric effect, Compton scattering, and electron–positron pair

Table 1: Particle properties (from Ref. [4])

	Particle type	Rest mass [MeV/c]	Mean life τ [s]	Main decay mode
Photons	(γ)	0	Stable	–
Leptons	Electron (e^-), positron (e^+)	0.511	Stable	–
	Muon (μ^+, μ^-)	105.66	2.2×10^{-6}	$\mu^- \to e^-\overline{\nu}_e\nu_\mu$
				$\mu^+ \to e^+\nu_e\overline{\nu}_\mu$
Hadrons	Proton (p)	938.27	Stable	–
	Neutron (n)	939.57	880	$n \to pe^-\overline{\nu}_e$
	Charged pion (π^+, π^-)	139.57	2.6×10^{-8}	$\pi^- \to \mu^-\overline{\nu}_\mu$
				$\pi^+ \to \mu^+\nu_\mu$
	Neutral pion (π^0)	134.98	8.5×10^{-17}	$\pi^0 \to \gamma\gamma$

production). Muons are mainly produced in the decay of pions and kaons.

3.2 Cross-section and mean free path

The collision probability of an incident particle with an atom (or nucleus) with atomic number Z and mass number A is generally quantified by the microscopic cross-section,

$$\sigma = \sigma(E, Z, A). \tag{1}$$

The microscopic cross-section depends on the projectile type and the projectile energy E and is often interpreted as the effective area for a collision. Cross-sections are generally expressed in terms of barns (1 b $= 10^{-24}$ cm^2). Considering a particle travelling inside a material, the average distance between two successive collisions, i.e., the mean free path λ, is then given by

$$\lambda = \frac{1}{N\sigma} = \frac{M}{\rho N_\mathrm{A}\sigma}\,, \tag{2}$$

where N is the atom density, M is the molar mass, ρ is the material density, and N_A is the Avogadro constant (6.022045×10^{23} mol^{-1}). In the case of a material mixture, the mean free path becomes

$$\lambda = \frac{1}{\sum_i N_i\sigma_i} = \frac{1}{\rho N_\mathrm{A}\sum_i(\omega_i/M_i)\sigma_i}, \tag{3}$$

where $\sum_i$ is the sum over all elemental constituents, N_i is the atom density of element i, M_i is the molar mass of element i, and ω_i is the relative weight fraction of element i in the mixture. The inverse of the mean free path is generally referred to as the macroscopic cross-section Σ,

$$\Sigma = \frac{1}{\lambda}. \tag{4}$$

In contrast with the microscopic cross-section in Eq. (1), the unit of the macroscopic cross-section is the unit of an inverse length, usually expressed in cm^{-1}. When quantifying particle penetration in matter, the amount of material traversed is often expressed as the product of length (cm) and material density (g/cm^3), which is also referred to as the surface density (g/cm^2). Surface density units are often used for the mean free path and macroscopic cross-section:

$$\lambda' = \lambda\rho, \quad \Sigma' = \frac{1}{\lambda\rho}. \tag{5}$$

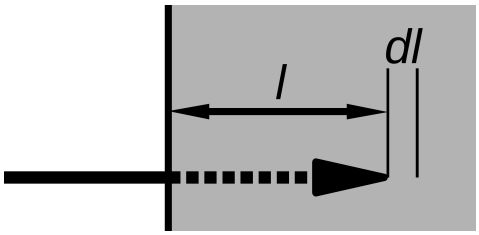


Fig. 1: Path length l of a particle inside a target.

3.3 Interaction probability

Assume that particles are normally incident on a homogeneous material and that they are subject to a process with a mean free path λ between collisions. Let us denote by l the distance travelled by a particle inside the material, as illustrated in Fig. 1. Then the probability that a particle suffers a collision between l and $l + \mathrm{d}l$ is given by

$$p(l)\mathrm{d}l = \frac{1}{\lambda}\exp\left(-\frac{l}{\lambda}\right)\mathrm{d}l. \tag{6}$$

By definition, the integral of $p(l)\mathrm{d}l$ over a semi-infinite target yields unity,

$$\int_0^\infty p(l')\mathrm{d}l' = 1, \tag{7}$$

and the first moment of the probability density function $p(l)$ is λ,

$$\int_0^\infty p(l')l'\mathrm{d}l' = \lambda. \tag{8}$$

The probability that a particle has an interaction up to a path length l is then given by the cumulative distribution function

$$P(l) = \int_0^l p(l')\mathrm{d}l' = 1 - \exp\left(-\frac{l}{\lambda}\right). \tag{9}$$

It follows from this expression that 63.2% of all particles will have a collision up to a path length of λ, 86.5% will have a collision within 2λ, 95% within 3λ, and 98.2% within 4λ. For a thin target, i.e., for a target thickness $d \ll \lambda$, the exponential term in Eq. (9) can be expanded into a Taylor series and $P(l)$ can be approximated as

$$P(d)\Big|_{d\ll\lambda} = 1 - \exp\left(-\frac{d}{\lambda}\right) = 1 - \left(1 - \frac{d}{\lambda} + \cdots\right) \approx \frac{d}{\lambda}, \tag{10}$$

i.e., the probability that the particle interacts in the target is simply given by the ratio of target thickness and mean free path. Analogous to the interaction probability, one can define the survival probability $P_\mathrm{s}(l)$, i.e., the probability that a particle does not suffer a collision until a distance l,

$$P_\mathrm{s}(l) = 1 - P(l) = \exp\left(-\frac{l}{\lambda}\right). \tag{11}$$

It follows from this expression that the chance that a particle did not have an interaction up to a distance l falls off exponentially.

As a practical example of the particle survival in an absorber, let us consider the accidental loss of a 450 GeV proton bunch train during transfer from the CERN SPS to the LHC. A nominal LHC injection train comprises 288 bunches with a bunch intensity of 1.15×10^{11} protons, which gives a total intensity of $I = 3.31 \times 10^{13}$ protons [8]. In the event of an extraction kicker erratic, a train can be

mis-steered and is then intercepted by collimators in the transfer line, which protect the transfer line and the LHC magnets from damage. Each of the SPS-to-LHC transfer line collimators is 1.2 m long and consists of graphite with a density of $1.84\ \mathrm{g/cm^3}$. Let us assume that an entire LHC injection train impacts on a single collimator: how many beam protons would then survive the impact, i.e., how many beam protons would escape from the downstream face of the collimator? A beam proton ceases to exist if it is subject to an inelastic nuclear collision inside the collimator. Elastic scattering processes (nuclear elastic or Coulomb scattering) and energy loss processes (ionization) do not affect proton survival and are therefore not considered here. The inelastic nuclear cross-section σ_i for a 450 GeV proton colliding with a carbon nucleus is about 245 mb, as predicted by the FLUKA code [18, 19]. Using the molar mass of carbon (12.0107 g/mol) and the specified graphite density, one can calculate the mean free path of protons in the collimator (in this case also referred to as the inelastic scattering length):

$$\lambda = \frac{M}{\rho N_{\mathrm{A}} \sigma_i} \tag{12}$$

$$= \frac{12.0107\ \mathrm{g/mol}}{1.84\ \mathrm{g/cm^3} \cdot 6.022 \times 10^{23}\ \mathrm{mol^{-1}} \cdot 245 \times 10^{-27} \mathrm{cm^2}} \tag{13}$$

$$= 44\ \mathrm{cm}\,. \tag{14}$$

Using Eq. (11), one can then derive the number of surviving beam protons:

$$I_s = I \cdot \exp\left(-\frac{l}{\lambda}\right) \tag{15}$$

$$= 3.31 \times 10^{13} \cdot \exp\left(-\frac{120\ \mathrm{cm}}{44\ \mathrm{cm}}\right) \tag{16}$$

$$= 2.2 \times 10^{12}\,. \tag{17}$$

This shows that 93.5% of the beam protons have an inelastic interaction in the collimator or, in other words, that the surviving proton intensity is equivalent to the intensity of 19 nominal bunches. It is to be noted that a large fraction of the initial energy of protons interacting in the collimator still escapes the absorber material in the form of particle showers. But contrary to the surviving beam protons, shower particles have different magnetic rigidities and they dissipate their energy in a more diluted way than focused proton bunches.

4 Atomic interactions

4.1 Photon interactions

Knowledge of photon interactions and absorption is important for most accelerator applications since photons are produced in a variety of processes. Several processes contribute to the scattering and to the absorption of photons in matter. Figure 2 illustrates the cross-sections of the different processes as a function of energy for carbon and lead. The different photon processes are as follows.

- *Coherent scattering or Rayleigh scattering.* The photon is subject to scattering from bound electrons without ionizing or exciting the atom. The recoil energy of the atom is generally negligible, i.e., the photon essentially retains its energy. The cross-section scales approximately with Z^2. Rayleigh scattering is only of marginal importance for many accelerator-related problems since it does not contribute to the absorption of photons.
- *Photoelectric effect.* The photon is absorbed and an electron is released from its atomic binding state. The photoelectric effect can only occur if the photon energy exceeds the binding energy U_i of an atomic shell. The energy of the emitted electron corresponds to the photon energy minus U_i. If the photoelectric effect involves an inner shell, the emission of an electron is typically followed by atomic relaxation. This means that the vacancy is filled by an electron from an outer shell

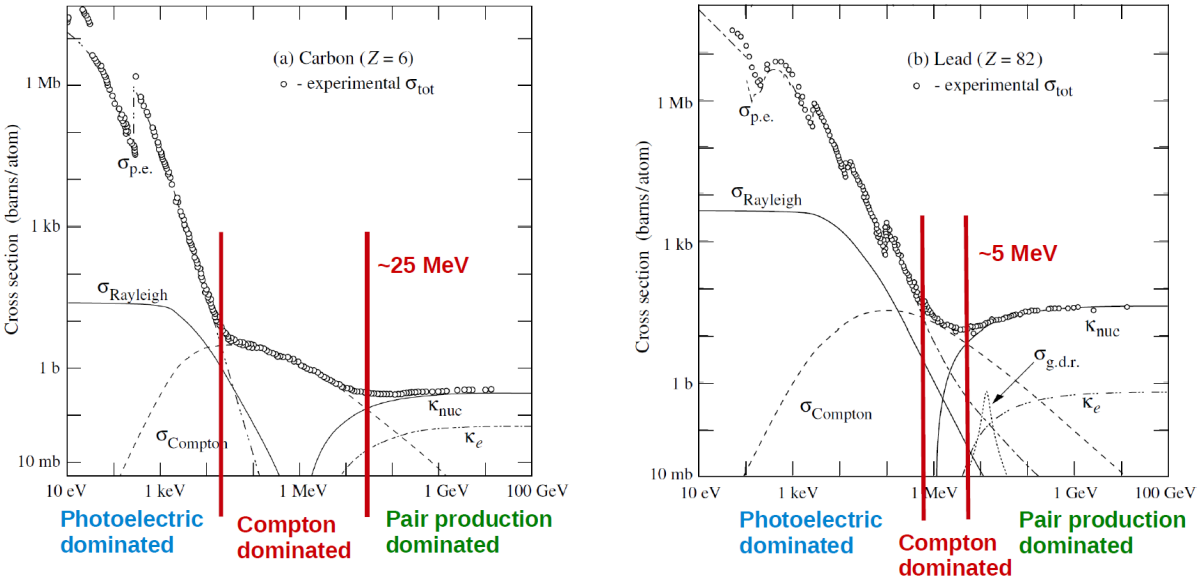


Fig. 2: Total cross-sections (circles) and partial cross-sections (lines) of photons in (left) carbon and (right) lead. The partial cross-sections correspond to the following processes: photoelectric effect ($\sigma_{\mathrm{p.e.}}$), coherent scattering ($\sigma_{\mathrm{Rayleigh}}$), Compton scattering ($\sigma_{\mathrm{Compton}}$), pair production in the field of the nucleus (κ_{nuc}), pair production in the field of the electrons (κ_{e}), and the giant dipole resonance ($\sigma_{\mathrm{g.d.r}}$), which is a photonuclear process. The figure originates from Ref. [4].

and characteristic X-rays or Auger electrons are emitted. The photoelectric effect is the dominant photon process at lower energies (see Fig. 2). The cross-section scales approximately with Z^{4-5} and exhibits distinct absorption edges which reflect the prevalence for certain atomic shells once the photon energy exceeds the corresponding binding energy.

– *Incoherent scattering or Compton scattering.* A photon collides with an electron, imparts some of its energy to the electron and is scattered into a different direction. The energy imparted on the electron can amount from zero to a large fraction of the incident gamma energy. Compton scattering can also be followed by atomic relaxation. The Compton effect is the dominant photon process at MeV energies (see Fig. 2). The cross-section scales approximately with Z, i.e., it has a weaker dependency on the atomic number than the photoelectric effect and electron–positron pair production. The Compton-dominated energy range is therefore smaller for higher-Z materials.

– *Electron–positron pair production.* At higher photon energies, pair production in the field of nuclei is the dominating absorption mechanism of photons. The threshold energy for electron–positron pair production is 1.022 MeV, i.e., the combined rest energy of an electron and a positron. The cross-section scales approximately with Z^2. The energy above which pair production dominates is less for higher-Z materials (around 25 MeV for carbon and around 5 MeV for lead). Pair production is the relevant photon process in electromagnetic cascades. Since high-energy electrons and positrons can lose energy in Bremsstrahlung processes, new photons can be produced. Electromagnetic cascades are described in more detail in Section 6.

In addition to atomic processes, photons can also be subject to photonuclear interactions, which are not discussed here. In photon-related problems, the mean free path λ multiplied by the material density ρ is generally referred to as the mass attenuation length or absorption length, and the inverse is generally referred to as the mass attenuation coefficient,

$$\mu_{\mathrm{m}} = \frac{1}{\lambda\rho} = \frac{N_{\mathrm{A}} \sum_j \sigma_j}{M\rho}, \tag{18}$$

where $\sum_j \sigma_j$ is the sum of cross-sections of all the processes discussed. The definition of the mass attenuation coefficient is identical to the macroscopic cross-section in surface density units (see Eq. (5)).

4.2 General remarks about charged particle interactions

Charged particles travelling inside a material are subject to Coulomb interactions with electrons and nuclei. As a consequence, the particles slow down and are deflected with respect to their initial direction.

- *Coulomb interactions with electrons lead to the excitation or ionization of atoms.* Energetic secondary electrons ejected in ionization processes are generally referred to as δ-rays. Collisions with electrons dominate the energy loss of charged particles up to energies above which radiative loss becomes important, i.e., typically up to tens of MeV for electrons and positrons, up to a few hundred GeV for muons, and up to even higher energies for charged hadrons. High-energy hadrons are also subject to inelastic nuclear collisions, which play a major role in the energy dissipation in a material if the hadron energy is of the order of GeV or higher. However, nuclear collisions are less frequent than Coulomb interactions and high-energy hadrons still lose some energy in Coulomb interactions before being subject to a nuclear collision.
  The energy loss due to Coulomb interactions with electrons is generally called electronic energy loss and leads to the heating of a material. In most Coulomb collisions, the energy transfer T is small (of the order of tens of eV), but cross-sections are large; hence, a charged particle will be subject to many collisions. In particle transport problems, the energy loss of charged particles is generally treated in a statistical way. Assuming that $\mathrm{d}\sigma(T)/\mathrm{d}T$ is the differential cross-section for an energy transfer T to an electron, then the mean energy loss per unit path length, which is generally referred to as the electronic stopping power, is given by

$$\left.\frac{\mathrm{d}E}{\mathrm{d}x}\right|_{\mathrm{elec}} = N \int_0^{T_{\max}} T \frac{\mathrm{d}\sigma}{\mathrm{d}T}(T)\mathrm{d}T, \tag{19}$$

  where $T_{\max}$ is the maximum energy transfer, which can be derived from particle kinematics. Stopping powers of hadrons, muons, and electrons are shown in the following subsections. Energy loss fluctuations are equally important when studying the ionizing energy loss of charged particles (often described by the Landau–Vavilov distribution [20–22]), but are not further discussed in this paper.
- *Coulomb interactions with nuclei dominate the angular deflection of charged particles in a material.* Particles are subject to smaller deflections if their energy is greater. Lighter particles scatter more if they have the same βc as heavier particles. Like the energy loss in electronic collisions, deflections in collisions with nuclei are often treated in theories where the net effect of multiple events is studied (multiple scattering theories). Multiple Coulomb scattering is described well by the theory of Molière [23] (see also Ref. [24]).
  The energy loss in collisions with nuclei is generally much smaller than the electronic energy loss, except for low-energy heavy projectiles (typically ions in the keV/u energy range). The energy loss is called non-ionizing energy loss and can lead to the displacement of atoms from the lattice if the energy transfer is higher than a certain material-dependent damage threshold, which can also depend on the particle direction in crystalline materials. In analogy to the electronic stopping power, the mean energy loss per unit length due to Coulomb collisions with nuclei is called the nuclear stopping power.

At higher energies, charged particles are also subject to radiative losses. Electrons with energies above a few tens of MeV are increasingly subject to Bremsstrahlung processes in the Coulomb field of nuclei, i.e., the electrons lose some of their energy, which is emitted in the form of a Bremsstrahlung photon. For other charged particles, radiative losses are only important at much higher energies because of their

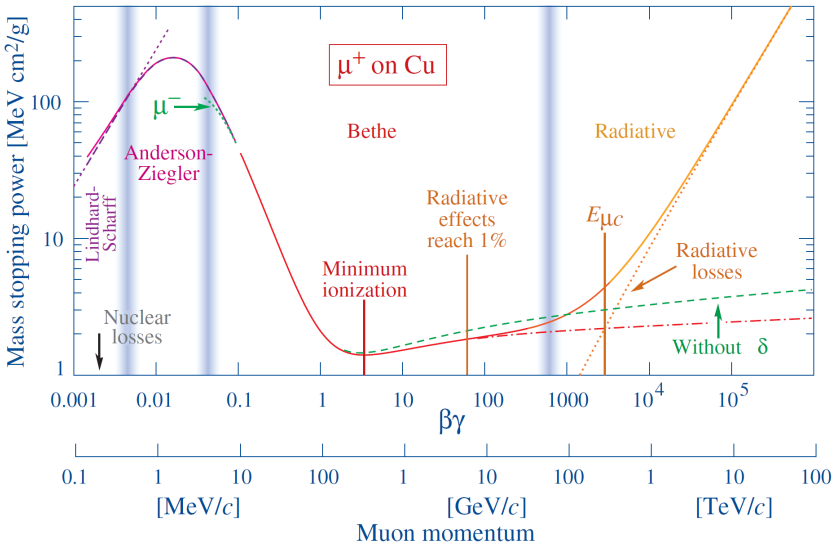


Fig. 3: Mass stopping power of positively charged muons in copper (solid line: total stopping power, dotted line: radiative stopping power, dash-dotted line: electronic stopping power according to Bethe). The figure originates from Ref. [4].

larger mass (e.g., hundreds of GeV for muons). High-energy muons can lose energy in Bremsstrahlung and electron–positron pair production processes, and they can also induce photonuclear interactions. The mean energy loss per unit length due to radiative losses is generally quantified by means of radiative stopping powers.

4.3 Energy loss of hadrons and muons

Figure 3 illustrates the mass stopping power of 100 keV/c–100 TeV/c muons in copper. Mass stopping power, which is generally expressed in MeV·cm^2/g, is defined as the stopping power multiplied by the material density. The solid line shows the sum of electronic and radiative stopping powers. The latter are dominant above a few hundred GeV. Over a large energy range (above a few MeV), electronic stopping powers of muons and hadrons can be described by the well-known Bethe equation [25], with the addition of some higher-order corrections (density effect, Bloch and Barkas–Andersen correction terms, see Refs. [26–28]). To a first order, electronic mass stopping powers of hadrons and muons exhibit only a weak material dependency in the Bethe regime (with the exception of hydrogen), since they are proportional to Z/A. In addition, the stopping power depends on the velocity and charge of a projectile, but only very weakly on the projectile mass (via T_{max}). The Bethe equation exhibits a distinct minimum around $\beta\gamma \approx 3$–3.5; charged projectiles with such velocities are generally referred to as minimum ionizing particles. The mass stopping power of minimum ionizing particles is 1–2 MeV·cm^2/g in most materials. At lower energies, where the Bethe equation cannot be applied, electronic stopping powers are often described by empirical formulae (see, for example, ICRU Report 49 [29] for stopping powers of protons and alpha particles).

4.4 Energy loss of electrons or positrons and the radiation length, X_0

For electrons and positrons, which are much lighter than muons, radiative losses are already important at much lower energies. This is illustrated in Fig. 4, which shows electronic and radiative mass stopping powers of electrons in carbon and lead. Radiative losses start to dominate above a few tens of MeV in low-Z materials, and above a few MeV in high-Z materials. The energy, at which electronic and radiative

Table 2: Critical energy of electrons E_c and radiation length X_0 for different materials. The values originate from Ref. [32]. The critical energy corresponds to Rossi's definition, which is slightly different from that given in Eq. (20). Rossi defines the critical energy as the energy where $-\mathrm{d}E/\mathrm{d}x|_\mathrm{elec} = E/X_0$, see Ref. [4].

Material	Z	Density [g/cm^3]	E_c [MeV]	X_0 [cm]
C	6	2.2	81.7	19.3
Al	13	2.7	42.7	8.90
Fe	26	7.8	21.7	1.76
Cu	29	8.96	19.4	1.44
W	74	19.3	8.0	0.35
Pb	82	11.4	7.4	0.56

stopping powers are equal is generally referred to as the critical energy E_c [4]:

$$\frac{\mathrm{d}E_\mathrm{ion}}{\mathrm{d}x}(E_\mathrm{c}) = \frac{\mathrm{d}E_\mathrm{rad}}{\mathrm{d}x}(E_\mathrm{c}). \tag{20}$$

Some numerical values of E_c for different materials are given in Table 2. As can be seen in Fig. 4, at higher energies, radiative stopping powers are linearly proportional to the electron energy, i.e., the mean energy loss can be written as

$$-\frac{\mathrm{d}E}{\mathrm{d}x} = \frac{E}{X_0}, \tag{21}$$

where the constant of proportionality, X_0^{-1}, is the inverse radiation length. It follows from this simple differential equation that the average energy of high-energy electrons decreases exponentially with travelled length:

$$\langle E(x) \rangle = E_0 \cdot \exp\left(-\frac{x}{X_0}\right), \tag{22}$$

where E_0 is the energy at $x = 0$. The radiation length is therefore the average distance needed to reduce the energy of a high-energy electron by a factor of $1/e$. At the same time, the radiation length is also a characteristic length for the penetration of high-energy photons in matter. The pair production cross-section (see Fig. 2) and, therefore, the mean free path are constant at high energies and can be written as [4]

$$\sigma_\mathrm{pair} \approx \frac{7}{9}\frac{M}{\rho N_\mathrm{A} X_0}, \qquad \lambda_\mathrm{pair} \approx \frac{9}{7}X_0\,, \tag{23}$$

where M and ρ are the molar mass and the material density, respectively. This means that the radiation length is about 7/9 of the mean free path of high-energy photons and therefore determines the photon survival probability, i.e.,

$$I(x) = I_0 \cdot \exp\left(-\frac{7}{9}\frac{x}{X_0}\right), \tag{24}$$

where $I(x)$ is the photon intensity at depth x. As will be discussed in Section 6.2, the radiation length is an important quantity for describing the longitudinal features of electromagnetic showers. The radiation length for different materials can be derived from the well-known formula from Tsai, as described in Ref. [4], but is commonly approximated by the following simplified expression [4]:

$$X_0' = \frac{716.4\ \mathrm{g/cm^2}\, A}{Z(Z+1)\ln(287/\sqrt{Z})}, \tag{25}$$

where X_0' is $X_0\rho$ in g/cm^2, and Z and A are the atomic and mass numbers, respectively. Some numerical values of the radiation length in different materials are given in Table 2.

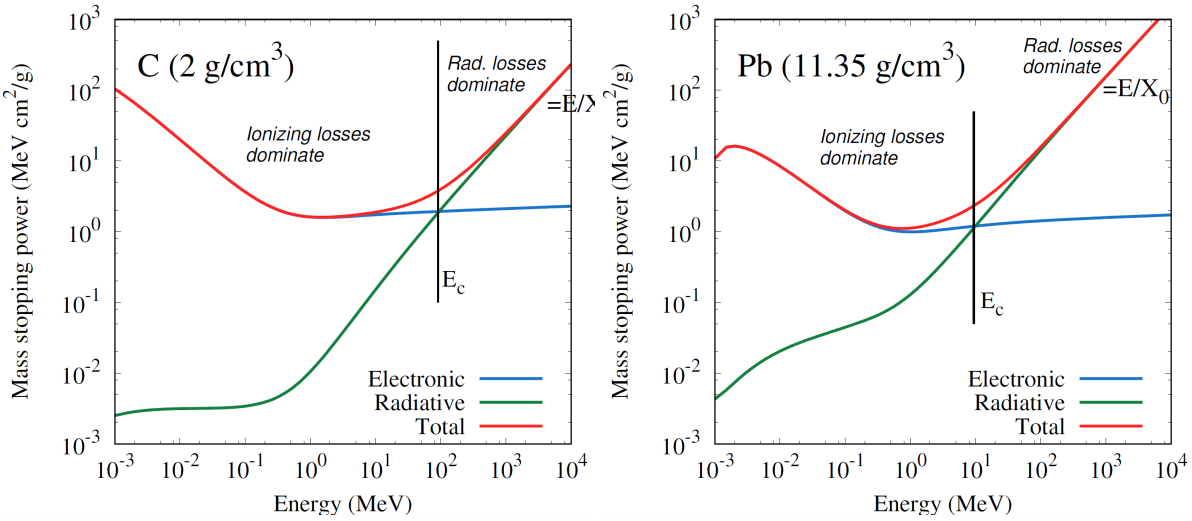


Fig. 4: Mass stopping powers of electrons in (left) carbon and (right) lead. The curves originate from the NIST ESTAR web database [30], which is based on tabulations from ICRU Report 37 [31]. The vertical lines indicate the critical energy defined in Eq. (20).

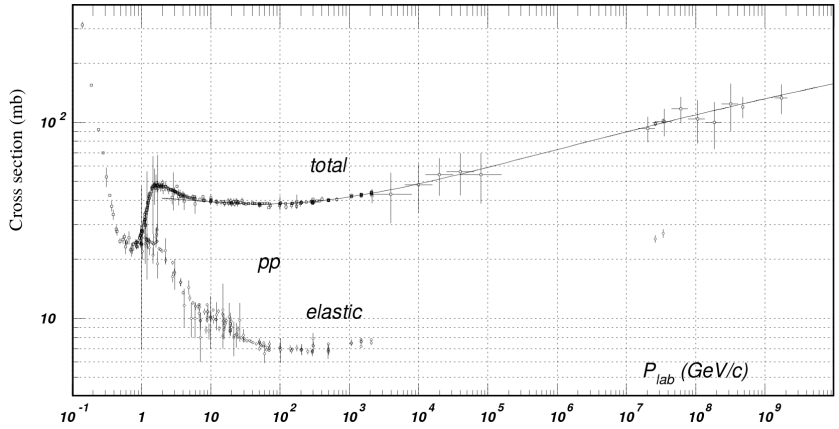


Fig. 5: Total and elastic proton–proton cross-sections. The figure originates from Ref. [4].

5 Nuclear interactions

5.1 Hadron–nucleus collisions

Hadrons travelling in matter can be subject to elastic and inelastic collisions with nuclei because of the strong force. Nuclear collisions are considered inelastic if new particles are produced or if the internal structure of the target nucleus or the projectile is changed. Some general features of nuclear collisions can be inferred from nucleon–nucleon cross-sections. Figure 5 shows total and elastic proton–proton cross-sections. The total cross-sections are the sum of the elastic and inelastic ones. As can be seen in the figure, above a few GeV, total cross-sections are relatively energy-independent up to the highest energies reached in today's accelerators. Moreover, over a wide energy range, inelastic cross-sections are much larger than elastic ones.

The microscopic cross-section for inelastic hadron–nucleus cross-section scales approximately with $A^{2/3}$, which means that the mean free path λ shows the following material dependency:

$$\lambda \propto \frac{A^{1/3}}{\rho}. \tag{26}$$

Table 3: Inelastic scattering length of 7 TeV protons in different materials. The values derive from the FLUKA Monte Carlo code [18, 19].

Material	Z	Density [g/cm^3]	λ [cm]
C	6	2.2	37.3
Al	13	2.7	35.4
Fe	26	7.8	15.1
Cu	29	8.96	13.9
W	74	19.3	8.9
Pb	82	11.4	15.7

The mean free path for inelastic collisions is generally known as the inelastic scattering length (as mentioned in Section 3.3). Table 3 lists the inelastic scattering lengths of 7 TeV protons in different elemental materials. The inelastic scattering length is a characteristic length for describing hadronic showers and is generally longer than the radiation length. Hadronic showers are therefore longer than electronic ones (see Section 6 for details).

Hadron–nucleus collisions have been described in more detail in the papers of Mokhov and Cerutti [3] and Ferrari [2]. A detailed account of nuclear reactions can also be found in Ref. [6]. The following description provides only a brief and very simplified picture of inelastic hadron–nucleus collisions at higher energies. In the initial fast stage of a hadron–nucleus interaction (10^{-22} s), the incident hadron interacts with the nucleons and new particles can be produced (in particular, pions) if the projectile energy is high enough (in particular for energies $\geq$GeV). If a nucleon collides with a free nucleon, pion production can occur at energies greater than 290 MeV [6]. The energy threshold is somewhat smaller for nucleons bound in a nucleus [6]. The hadrons (protons, neutrons, pions) can form intranuclear cascades and the more energetic particles can leave the nucleus. These fast particles are mainly forwards-directed and their multiplicity scales approximately with the logarithm of the incident particle energy. The other particles can deposit their energy in the nucleus and leave it in an excited state. After going through a pre-compound phase, the nucleus is subject to a slow de-excitation process (10^{-16} s). In this evaporation phase, lower-energy (MeV) neutrons, protons, or light fragments (alpha particles) are emitted isotropically. In addition, the residual nucleus can be subject to γ de-excitation. In heavy elements, evaporation competes with fission, where the nucleus breaks up in larger residuals. Fission products can also be subject to particle evaporation. The residual nuclei remaining from evaporation or fission can be radioactive.

5.2 Low-energy neutron interactions

Neutrons have a relatively long mean lifetime, of almost 15 min; hence, they are the only 'stable' neutral hadrons. Neutrons with energies of a few MeV will mainly slow down, owing to elastic scattering, until they are thermalized and captured by a nucleus. For illustration, Fig. 6 shows the cross-section of neutrons in ^{56}Fe as a function of the neutron energy (between 10^{-4} eV and 20 MeV). The cross-sections have characteristic resonances because of the specific nuclear structure. Over a wide energy range, the elastic scattering cross-section dominates. Inelastic interactions, where one or more less-energetic neutrons are emitted from the target nucleus or the nucleus gets excited, occur above a threshold energy and contribute to the deceleration of neutrons in the MeV region. Other reaction channels exist as well, where protons, deuterons, or alpha particles are emitted. The cross-section for capture is inversely proportional to the neutron velocity and thermal neutrons are hence more likely to be absorbed than fast neutrons. Capture is often followed by gamma emission (radiative capture). In heavier materials, neutrons can also induce fission. Neutron absorption depends strongly on the nuclear isotopes contained in a material; different evaluated data libraries exist, in which neutron cross-sections are tabulated (e.g., ENDF/B-VII, JENDL); the tables can be accessed via a web-based interface [34].

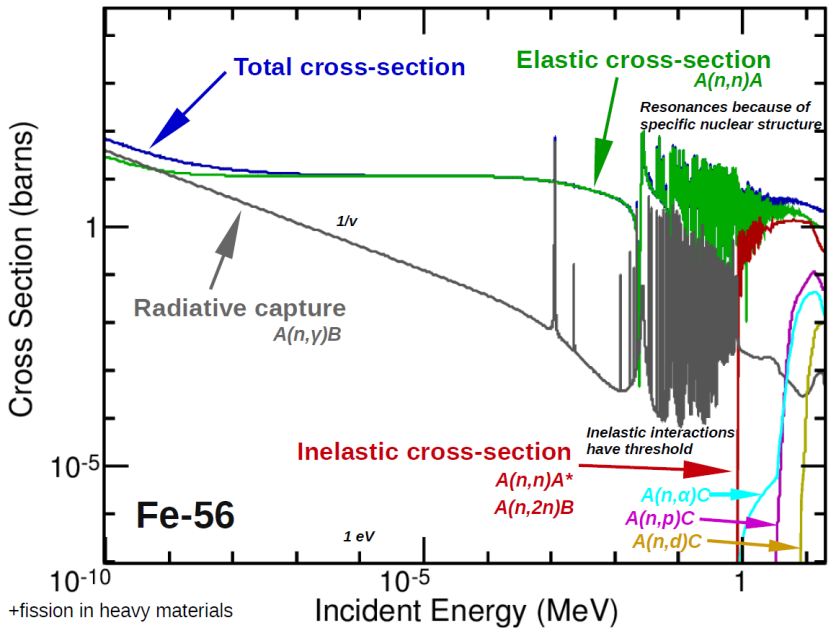


Fig. 6: Cross-sections of neutrons in ^{56}Fe. The figure originates from the web-based interface of the ENDF/B-VII.1 database [33].

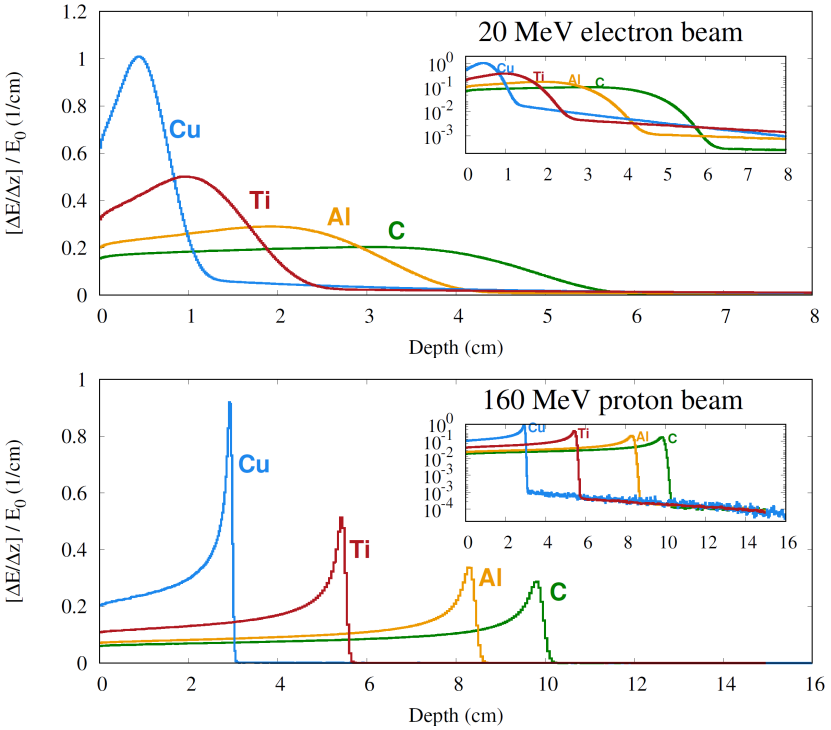


Fig. 7: Longitudinal energy deposition profiles of 20 MeV electron beams (top) and 160 MeV proton beams (bottom) in different materials. The profiles were obtained with the FLUKA Monte Carlo code and are expressed in the reduced form defined in Eq. (27). The inset figures are in logarithmic scale for a better visualization of the energy deposition tails.

6 Energy deposition and particle showers

The energy deposition in a material is mediated by Coulomb interactions of charged particles, which are set in motion by atomic and nuclear processes. The ionizing energy loss of a charged particle in a certain volume V can still be different from the energy deposition in V, since the energy can be transported away by δ-rays. This section illustrates, by means of a few examples, the dependency of energy deposition profiles on particle type, energy, and target material. Section 6.1 addresses lower particle energies, where particle showers are irrelevant (MeV regime), while Sections 6.2 and 6.3 are dedicated to electromagnetic and hadronic showers, which determine the energy deposition at higher energies (GeV regime and above). To highlight characteristic features of energy deposition profiles, it is assumed that a pencil beam impacts a laterally infinite material block. The longitudinal energy deposition profiles shown throughout this section are expressed as

$$\varepsilon(z_i) = \frac{1}{E_0}\frac{\Delta E(z_i)}{\Delta z}, \tag{27}$$

where E_0 is the incident particle energy and $\Delta E(z_i)$ is the energy deposited in the layer between z_i and $z_i + \Delta z$. In this form, energy deposition profiles are more easily comparable, since the area under the curves is identical (except for cases involving nuclear interactions, where a fraction of energy is needed to overcome the binding energy). Depending on the particle type and energy, the longitudinal co-ordinate is substituted by a characteristic length, such as the radiation length (electromagnetic showers) or the inelastic scattering length (hadronic showers).

6.1 Energy deposition by lower-energy particles (MeV regime)

Figure 7 illustrates energy deposition profiles of 20 MeV electron beams in different materials. The profiles are characterized by an initial build-up region, which is followed by a broad peak. The shape of the peak is much governed by the Coulomb scattering of electrons as they slow down. The profiles have a long tail induced by Bremsstrahlung photons, which can travel much further than the electrons. The tail is more pronounced in heavier materials, since the critical energy is lower and radiative losses are hence more important (E_c is roughly 80 MeV for carbon and about 20 MeV for copper, see Table 2).

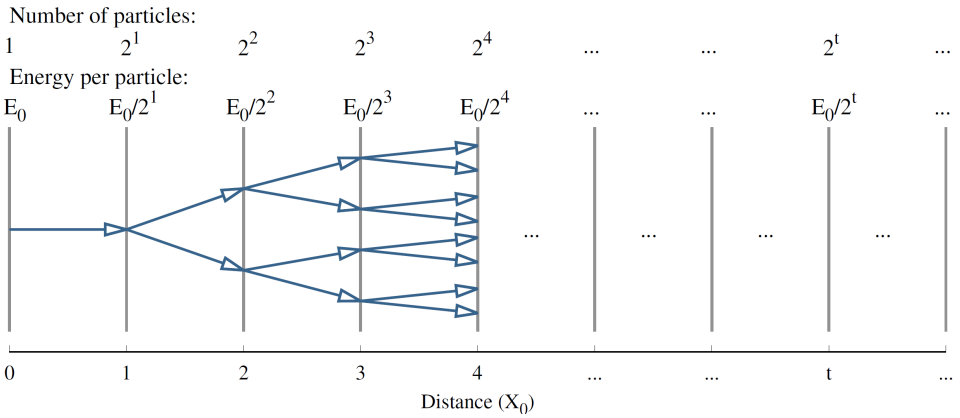


Fig. 8: Multiplication of electromagnetic cascade particles in the toy model of Heitler. The incident particle can be either an electron or positron or a photon. The longitudinal co-ordinate is expressed in terms of radiation lengths.

The energy deposition profiles of MeV protons (up to some hundreds of MeV) are characterized by a distinct peak, which is referred to as the Bragg peak. Compared with electrons, the angular deflection of protons as they slow down is much smaller and many protons stop within a certain depth range. Since the average energy loss per unit length is higher towards the end of their range (see Bethe formula), the energy deposition rises sharply before the protons stop. The exact shape of the Bragg peak

is determined by the beam energy spread, the energy loss straggling, and Coulomb and nuclear elastic scattering processes, but also by inelastic nuclear interactions, which become increasingly important at higher energies. Figure 7 illustrates the Bragg peak of 160 MeV protons in different materials. Beyond the peak, the energy deposition profiles feature a long tail, which is due to neutrons. Ions exhibit a similar Bragg peak as protons, but the tail beyond the peak is more pronounced because of secondary particles produced in the fragmentation of ions.

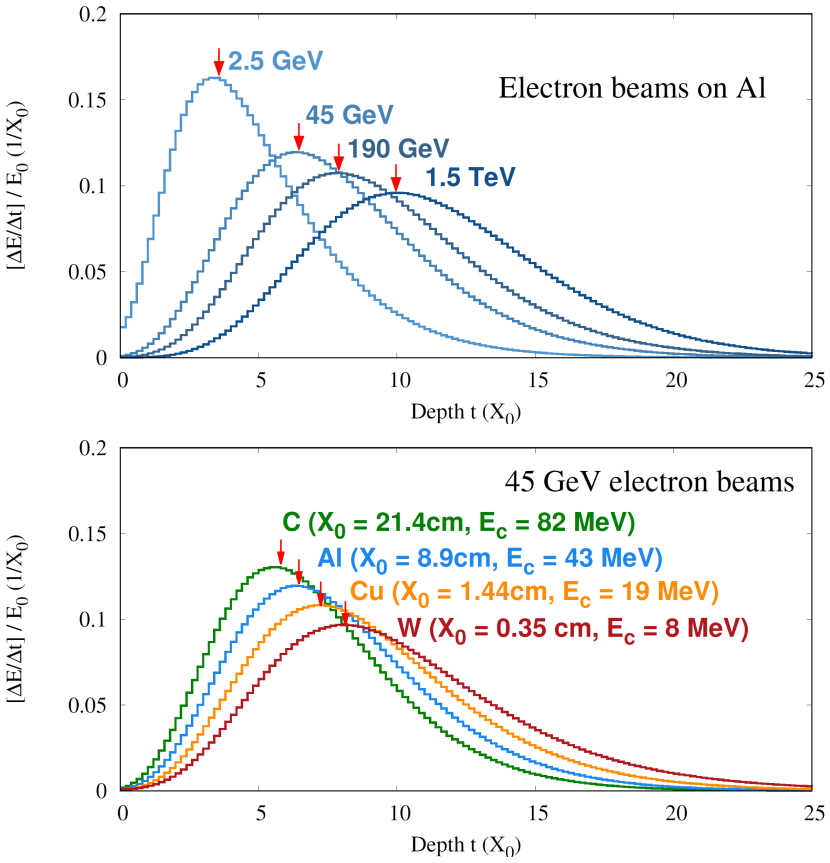


Fig. 9: Longitudinal energy deposition profiles of (top) 2.5 GeV–1.5 TeV electron beams in aluminium and (bottom) 45 GeV electron beams in carbon, aluminium, copper, and tungsten. The beam energies correspond to the operational energies of past and future machines: KEK/PF linac (2.5 GeV), LEP at start-up (45 GeV), CLIC stage 1 (190 GeV), and CLIC stage 3 (1.5 TeV). The profiles were obtained with the FLUKA Monte Carlo code and are expressed in the reduced form defined in Eq. (27). The longitudinal co-ordinate is given in terms of radiation lengths. The red arrows indicate the position of the shower maximum, as predicted by Eq. (31).

6.2 Electromagnetic showers

As discussed in Section 4, high-energy electrons and positrons mainly lose energy through Bremsstrahlung processes, whereas high-energy photon interactions are dominated by electron–positron pair production. If the energy of an incident electron, positron or photon is high enough ($\geq$GeV), these processes give rise to a particle multiplication, which is referred to as an electromagnetic shower. The average particle energy decreases from shower generation to shower generation and the cascade development approximately ceases when the energy of electrons and positrons falls below the critical energy, E_c, where ionization and excitation processes become dominant.

The location where the number of cascade particles is a maximum is referred to as the shower maximum. Since the radiation length X_0 is a characteristic length for Bremsstrahlung and pair production

processes, it is also a central quantity for describing the longitudinal development of electromagnetic showers. Some qualitative features of the longitudinal shower profile can be derived from Heitler's model [35], which assumes that interactions (Bremsstrahlung and pair production) always take place after a distance X_0 and that, at each interaction, the energy is equally split between the two outgoing particles. In this simple model, the number of particles after t radiation lengths is given by

$$N(t) = 2^t \tag{28}$$

and the energy per particle is

$$E(t) = E_0/2^t, \tag{29}$$

where E_0 is the energy of the incident particle (see Fig. 8). If one assumes that the cascade stops, i.e., reaches its maximum, when the energy per particle is equal to E_c, then it follows from Eq. (29) that the depth of the maximum, t_max, depends logarithmically on the initial energy:

$$t_\mathrm{max} \propto \log\left(\frac{E_0}{E_\mathrm{c}}\right). \tag{30}$$

Since t_max expresses the depth of the shower maximum in terms of radiation lengths, and since X_0 is approximately proportional to $A/Z^2\rho$ (see Section 4), it follows that electromagnetic showers are shorter the higher the atomic number and the higher the material density. If one inserts Eq. (30) into Eq. (28), one can also see that the shower size, i.e., the number of particles at the maximum, increases with the initial particle energy.

Although the Heitler model provides a qualitative description of the longitudinal shower development, it fails to predict absolute particle numbers or relative features, such as the ratio of electrons or positrons and photons at the shower maximum. In reality, there is an excess of photons, since Bremsstrahlung photons are much more frequently produced than is assumed in the Heitler model. In addition, the model neglects other processes, such as Compton scattering and the photoelectric effect, which lead to the production of low-energy electrons. These electrons are important for dissipating the energy carried by the shower.

A more accurate description of the electromagnetic shower development can be obtained from Monte Carlo simulations. Figure 9 shows longitudinal energy deposition profiles of electron beams in different materials. The profiles were obtained using the FLUKA shower simulation code [18, 19]. The longitudinal co-ordinate is expressed in terms of radiation length. The two plots demonstrate the dependency on the beam energy E_0 and the target material (E_c), respectively. In higher-Z materials, where the critical energy E_c is lower, the multiplication process continues down to lower energies and the shower maximum therefore occurs at larger depths (in terms of X_0). The red arrows in the plots indicate the position of the shower maximum predicted by the following empirical formula [4]:

$$t_\mathrm{max} = log\left(\frac{E_0}{E_\mathrm{c}}\right) - 0.5. \tag{31}$$

The formula applies to electron- and positron-induced showers, while the latter term has to be replaced by '$+0.5$' for photon-induced showers [4]. This small difference results from the different nature of the initial interactions and is, by construction, not predicted by the Heitler model. Beyond the shower maximum, the energy deposition profiles do not scale with X_0, since other processes than Bremsstrahlung and pair production become dominant. The longitudinal shower tails decay more rapidly in lower-Z materials.

The transverse profile of the shower core around the shower maximum is roughly independent of the beam energy. The lateral shower opening is mainly governed by the multiple scattering of lower-energy electrons. Lateral features of the shower core can be well described by the Molière radius R_M,

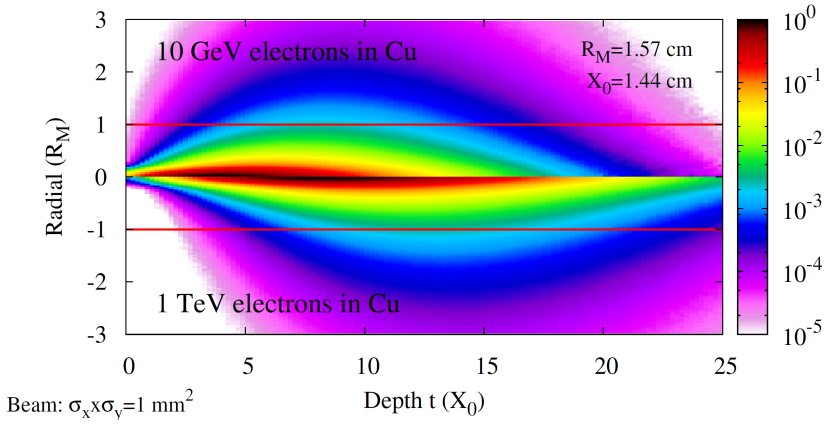


Fig. 10: Two-dimensional energy density maps for (top half) 10 GeV and (bottom half) 1 TeV electron beams incident on copper. The profiles were obtained with FLUKA Monte Carlo simulations and are normalized to the maximum energy density. The longitudinal co-ordinate is given in terms of radiation lengths and the radial co-ordinate in terms of Molière radii. The assumed beam size (σ) is 1 mm.

which represents the average lateral deflection of electrons with $E = E_{\rm c}$ after traversing one radiation length X_0 [4]:

$$R_{\rm M} = \frac{\sqrt{4\pi\alpha} m_{\rm e} c^2}{E_{\rm c}} X_0 \approx \frac{21\ {\rm MeV}}{E_{\rm c}} X_0, \tag{32}$$

where α is the fine structure constant, and $m_{\rm e}c^2$ is the electron mass. Typically, a large fraction of the beam energy (>90%) is deposited within a cylinder with a radius of $R_{\rm M}$ around the beam axis (if the beam size is much smaller than $R_{\rm M}$). For illustration, Fig. 10 shows two-dimensional energy density maps for 10 GeV and 1 TeV electron beams incident on copper. The radial energy density profile at the shower maximum is comparable for the two energies.

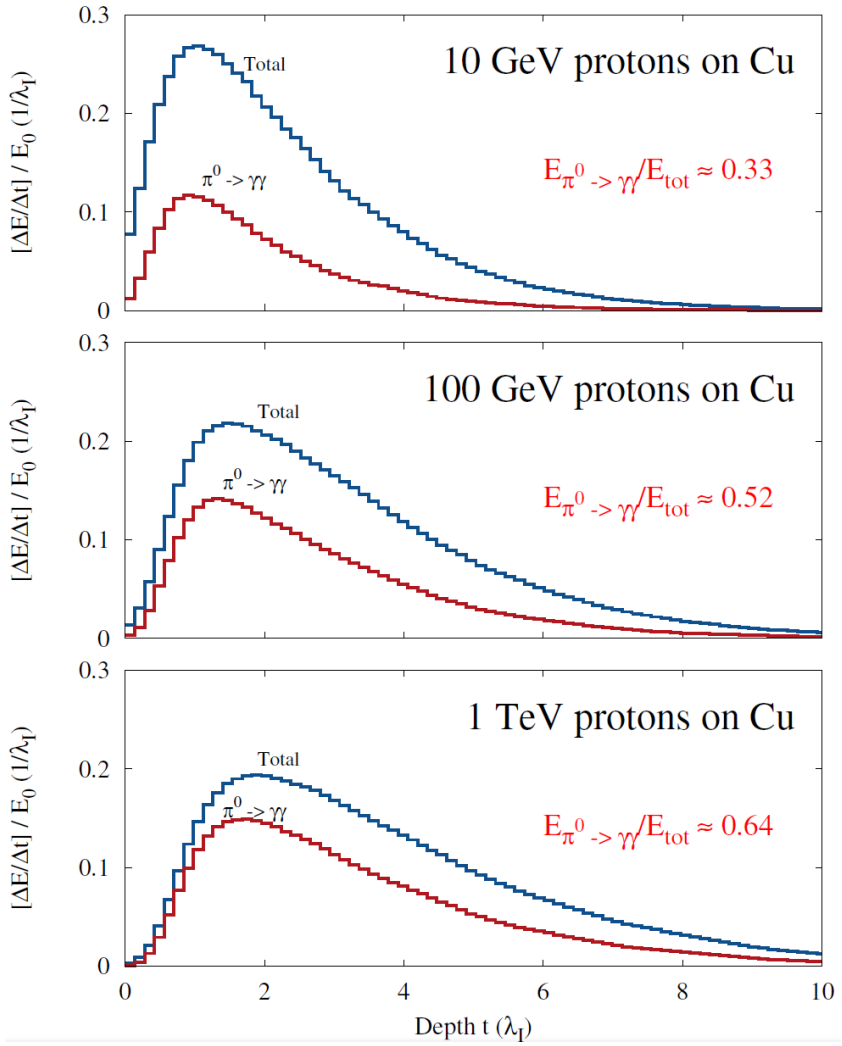


Fig. 11: Longitudinal energy deposition profiles of (top) 10 GeV, (centre) 100 GeV, and (bottom) 1 TeV proton beams in copper. Profiles were calculated with the FLUKA Monte Carlo code and are expressed in the reduced form defined in Eq. (27). Blue curves indicate the total energy deposition, while red curves show the energy deposited by electromagnetic showers initiated by $\pi^0 \to \gamma\gamma$.

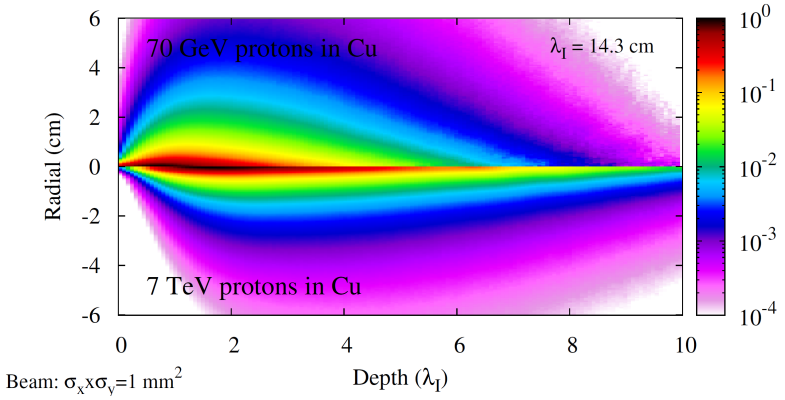


Fig. 12: Two-dimensional energy density maps for 70 GeV (top half) and 7 TeV (bottom half) proton beams incident on copper. The profiles were obtained with FLUKA Monte Carlo simulations and are normalized to the maximum energy density. The longitudinal co-ordinate is given in terms of inelastic scattering lengths. The assumed beam size (σ) is 1 mm.

6.3 Hadronic showers

If a high-energy hadron undergoes an inelastic nuclear collision in a material, the energetic, mainly forwards-directed collision products (n, p, $\pi^{+/-}$, ...) will be subject to further collisions. Since, on average, several hadrons are produced in each collision, the re-interaction of secondary products gives rise to a particle multiplication, which is called a hadronic shower. Since the multiplication process is governed by inelastic collisions, the longitudinal shower development can be characterized by the inelastic scattering length λ, introduced in Section 5. A non-negligible fraction of the initial energy can go into binding energy or can be transferred to recoils. Lower-energy particles are produced in evaporation, gamma de-excitation, and fission, and some energy is also carried away by neutrinos. The multiplication process roughly continues until the energy falls below the pion production threshold. As for electromagnetic showers, the shower length, i.e., the depth of the shower maximum, increases logarithmically with energy.

Hadronic showers also have an electromagnetic component, owing to the production of neutral pions, which decay into energetic γ pairs (neutral pions also have other decay channels, but with smaller branching ratios). Other secondary particles can also induce electromagnetic showers (e.g., η mesons), but they are produced with much smaller multiplicities. The electromagnetic shower component is mainly concentrated around the shower core, since electromagnetic showers are narrower than hadronic ones. In addition, they are shorter since the radiation length is generally much smaller than the inelastic scattering length (compare Tables 2 and 3). While hadronic showers can transfer a significant fraction of energy to electromagnetic showers, the coupling in the other direction is much weaker. Hadrons can be produced in photonuclear or electronuclear interactions, but the cross-sections of these processes are small and the fraction of energy transferred back to hadrons is often negligible. Photonuclear or electronuclear processes can still be of importance for some applications (e.g., neutron production in electron accelerators).

The fraction of energy dissipated by the electromagnetic shower component increases with particle energy. At each nuclear interaction, about one-third of the energy is carried away by π^0s. Conversely, the higher the initial energy, the more interactions are needed, such that hadrons in the cascade reach GeV energies and the shower eventually ceases. As a consequence, the contribution of electromagnetic showers to the energy deposition increases and becomes dominant at higher energies. This is illustrated in Fig. 11, which shows longitudinal energy deposition profiles of 10 GeV, 100 GeV, and 1 TeV proton beams in copper. The longitudinal co-ordinate is expressed in terms of inelastic scattering lengths. At 100 GeV, about one-half of the initial energy is dissipated by γ-induced showers.

The transverse momentum of hadrons produced in nuclear collisions is more or less invariant with energy (the average being 300–400 MeV/c) [6]. As a consequence, the shower opening angle becomes narrower with increasing energy. This in turn also leads to a greater overlap of electromagnetic showers around the core. Figure 12 shows two-dimensional energy density maps for 70 GeV and 7 TeV beams incident on copper. The radial profile drops more sharply in the latter case.

Acknowledgements

The author thanks F. Cerutti for proofreading the manuscript.

References

[1] A. Ferrari, A primer on particle interactions with matter, (CERN, Geneva, 2011).

[2] A. Ferrari, Interactions of particles with matter, CERN Accelerator School, Accelerators for Medical Applications, Vösendorf, 2015.

[3] N.V. Mokhov and F. Cerutti, Beam–material interactions, Proc. 2014 Joint International Accelerator School: Beam Loss and Accelerator Protection, Newport Beach, CA, USA, 5-14 November 2014, CERN-2016-002 (CERN, Geneva, 2016) p. 83. http://dx.doi.org/10.5170/CERN-2016-002.

[4] C. Patrignani *et al.* (Particle Data Group), *Chin. Phys. C* **40** (2016) 100001, updated 2017, https://doi.org/10.1088/1674-1137/40/10/100001.

[5] P. Sigmund, *Particle Penetration and Radiation Effects*, (Springer-Verlag, Berlin, 2006).

[6] A. Ferrari and P.R. Sala, Proc. Workshop on Nuclear Reaction Data and Nuclear Reactors, Physics, design and safety, Trieste, Italy, 15 April-17 May 1996, Eds. A. Gandini and G. Reffo (World Scientific, Singapore, 1998), p. 424, https://doi.org/10.1142/9789814530385.

[7] B.B. Rossi, *High-Energy Particles*, (Prentice-Hall, New York, 1952).

[8] LHC Design Report, edited by O.S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole and P. Proudlock, CERN-2004-003-V-1, (CERN, Geneva, 2004), http://dx.doi.org/10.5170/CERN-2004-003-V-1.

[9] N.V. Mokhov *et al.*, Protecting LHC IP1/IP5 components against radiation resulting from colliding beam interactions, LHC-Project-Report-633, 2003.

[10] R. Bruce *et al.*, *Phys. Rev. ST Accel. Beams* **12** (2009) 071002, https://doi.org/10.1103/PhysRevSTAB.12.071002.

[11] R. Bruce *et al.*, *Nucl. Instr. Meth. Phys. Res. A* **729** (2013) 825, https://doi.org/10.1016/j.nima.2013.08.058.

[12] F. Zimmermann *et al.*, Trapped macroparticles in electron storage rings, Proc. 1995 Particle Accelerator Conf., Dallas, TX, USA, 1995 (IEEE, Piscataway, 1996), p. 517, https://doi.org/10.1109/PAC.1995.504705.

[13] T. Baer *et al.*, UFOs in the LHC, Proc. 2nd International Particle Accelerator Conf., TUPC137, San Sebastian, Spain, 2011, Ed. C. Petit-Jean-Genaz (EPS-AG, Geneva, 2011), p. 1347, http://accelconf.web.cern.ch/AccelConf/IPAC2011/index.htm.

[14] G. Papotti *et al.*, Macroparticle-induced losses during 6.5 TeV LHC operation, Proc. 7th International Particle Accelerator Conf., TUPMW023, Busan, Korea, 2016, Eds. K.S. Kim, C. Petit-Jean-Genaz, I.S. Ko, K.R. Kim and V.R.W Schaa (JACoW, Geneva, 2016), p. 1481, http://dx.doi.org/10.18429/JACoW-IPAC2016-TUPMW023.

[15] H. Vincke *et al.*, *Radiat. Prot. Dosimetry* **146** (2011) 434, https://doi.org/10.1093/rpd/ncr245.

[16] B. Auchmann *et al.*, *Phys. Rev. ST Accel. Beams* **18** (2015) 061002, https://doi.org/10.1103/PhysRevSTAB.18.061002.

[17] M. Brugger, Radiation damage to electronics at the LHC, Proc. 3rd International Particle Accelerator Conf., MOPPC004, New Orleans, LA, USA, 2012, Eds. F. Zimmermann and C. Eyberger (IEEE, Piscataway, 2012), p. 3734.

[18] T.T. Böhlen *et al.*, *Nucl. Data Sheets* **120** (2014) 211, https://doi.org/10.1016/j.nds.2014.07.049.

[19] A. Ferrari *et al.*, FLUKA: a multi-particle transport code, CERN-2005-10, INFN/TC_05/11, SLAC-R-773 (CERN, Geneva, 2005), http://dx.doi.org/10.5170/CERN-2005-010.

[20] L. Landau, *J. Phys. USSR* **8** (1944) 201.

[21] P.V. Vavilov, *Sov. Phys. JETP* **5** (1957) 749.

[22] P.V. Vavilov, *Zh. Eksp. Teor. Fiz.* **32** (1957) 920.

[23] G. Molière, *Z. Naturforsch. A* **3** (1948) 78.

[24] H.A. Bethe, *Phys. Rev.* **89** (1953) 1256, https://doi.org/10.1103/PhysRev.89.1256.

[25] H.A. Bethe, *Ann. Phys.* **397** (1930) 325, https://doi.org/10.1002/andp.19303970303.

[26] F. Bloch, *Ann. Phys.* **408** (1933) 285, https://doi.org/10.1002/andp.19334080303.

[27] S. P. Ahlen, *Phys. Rev. A* **17** (1978) 1236, https://doi.org/10.1103/PhysRevA.17.1236.

[28] J. C. Ashley *et al.*, *Phys. Rev. B* **5** (1972) 2393, https://doi.org/10.1103/PhysRevB.5.2393.

[29] M. J. Berger *et al.*, *J. Int. Commission Radiat. Units Measurements* **25** (1993), https://doi.org/10.1093/jicru/os25.2.Report49.

[30] https://physics.nist.gov/PhysRefData/Star/Text/ESTAR.html, last accessed January 30th, 2018.

[31] M. J. Berger *et al.*, *J. Int. Commission Radiat. Units Measurements* **19** (1984), https://doi.org/10.1093/jicru/os19.2.Report37.

[32] http://pdg.lbl.gov/2017/AtomicNuclearProperties/, last accessed January 30th, 2018.

[33] http://www.nndc.bnl.gov/exfor/endf00.jsp, last accessed January 30th, 2018.

[34] https://www-nds.iaea.org/exfor/endf.htm, last accessed January 30th, 2018.

[35] W. Heitler, *The Quantum Theory of Radiation*, (Clarendon Press, Oxford, 1954).

Proceedings of the CAS–CERN Accelerator School: Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018)

Recapitulation of Electromagnetism

T. Flisgen[*]*, J. Heller, and U. van Rienen*
University of Rostock, Germany

Abstract
Electromagnetism plays a vital role in many areas of our modern life, such as telecommunication, navigation, electrical drives, and electrical grids. Moreover, electromagnetism is of special relevance for particle accelerators. Magnetic fields are used to bend and to focus bunches of charged particles, whereas electric fields are used to increase the energy of charged particles. This contribution recapitulates selected fundamentals of the theory of electromagnetic fields and sketches thereby important facts related to vector analysis, electric fields, magnetic fields, and their mutual interaction.

Keywords
Maxwell's equations; vector analysis; conservation principles; continuity constraints of fields; wave equations; field attenuation in conductors.

1 Introduction

Today's daily life would not be as comfortable as it is without a profound understanding of electromagnetic phenomena. This report summarizes the most relevant foundations of electromagnetic theory. However, this article does not discuss electromagnetics in a complete manner, owing to lack of space. The report can be considered as a reminder and as an overview. It is meant to be an appetizer and an invitation for more profound studies with textbooks [1–8]. These textbooks describe electromagnetism in a very detailed way and often comprise hundreds of pages.

Subsection 1.1 presents the quantities describing electromagnetic effects and introduces the famous four equations, which relate these quantities to each other. In this context, important vector algebraic operators, such as the divergence and the curl operator, are introduced and defined. Subsequently, forces acting on charged particles as a result of electromagnetic fields are discussed in Subsection 1.2. Subsections 1.3 and 1.4 discuss conservation principles, such as the conservation of charge and the conservation of energy. Section 2 describes the interaction of electromagnetic fields with matter. A special focus is on continuity constraints of fields at interfaces between different materials. Subsequently, the governing equations for static electric and for static magnetic fields are derived in Sections 3 and 4, respectively. In a next step, wave equations for electric and magnetic waves are deduced from Maxwell's equations in Section 5. A simple example for a solution of the wave equations is provided as an example. Section 6 discusses the attenuation of fields in a conductor and Section 7 concludes the report.

Note that a consistent nomenclature is employed throughout the entire report. Vectors are denoted with boldfaced letters. Where quantities depend on time or space, they are highlighted with t or $\mathbf{r}$, respectively. For instance, a transient spatially dependent vector field is denoted $\mathbf{D}(\mathbf{r},t)$. In the case of harmonic fields, complex-valued quantities arise, which are highlighted by underscores, e.g., $\mathbf{D}(\mathbf{r},t) = \mathrm{Re}\,[\underline{\mathbf{D}}(\mathbf{r},t)]$. The complex-valued quantity $\underline{\mathbf{D}}(\mathbf{r},t)$ can be split up into a phasor $\underline{\mathbf{D}}(\mathbf{r})$ and an exponential function $\exp(\mathrm{j}\omega t)$ accounting for the time dependency, i.e., $\underline{\mathbf{D}}(\mathbf{r},t) = \underline{\mathbf{D}}(\mathbf{r})\exp(\mathrm{j}\omega t)$, where $\mathrm{j}^2 = -1$.

1.1 Maxwell's equations

The following text closely follows the introductory section in Ref. [9]. In the nineteenth century, James Clerk Maxwell summarized and improved important laws on electric and magnetic phenomena [10]

[*]Currently at Helmholtz Zentrum Berlin.

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-005.69

originating from experimental observations. In the following decades, Oliver Heaviside and Josiah Willard Gibbs independently reformulated the equations into the now commonly known form. These so-called Maxwell's equations are the foundation of classical electrodynamics. The integral form of Maxwell's equations is given by

$$\oiint_{\partial\mathbf{\Omega}} \mathbf{D}(\mathbf{r},t)\cdot\mathrm{d}\mathbf{A} = \iiint_{\mathbf{\Omega}} \rho(\mathbf{r},t)\,\mathrm{d}V\,, \tag{1}$$

$$\oiint_{\partial\mathbf{\Omega}} \mathbf{B}(\mathbf{r},t)\cdot\mathrm{d}\mathbf{A} = 0\,, \tag{2}$$

$$\oint_{\partial\mathbf{\Gamma}} \mathbf{E}(\mathbf{r},t)\cdot\mathrm{d}\mathbf{s} = -\iint_{\mathbf{\Gamma}} \frac{\partial}{\partial t}\mathbf{B}(\mathbf{r},t)\cdot\mathrm{d}\mathbf{A}\,, \tag{3}$$

$$\oint_{\partial\mathbf{\Gamma}} \mathbf{H}(\mathbf{r},t)\cdot\mathrm{d}\mathbf{s} = \iint_{\mathbf{\Gamma}} \left(\frac{\partial}{\partial t}\mathbf{D}(\mathbf{r},t)+\mathbf{J}(\mathbf{r},t)\right)\cdot\mathrm{d}\mathbf{A}\,. \tag{4}$$

These four equations specify the relationship between the *electric flux density* $\mathbf{D}(\mathbf{r},t)$, the *electric charge density* $\rho(\mathbf{r},t)$, the *magnetic flux density* $\mathbf{B}(\mathbf{r},t)$, the *electric field strength* $\mathbf{E}(\mathbf{r},t)$, the *magnetic field strength* $\mathbf{H}(\mathbf{r},t)$, and the *electric current density* $\mathbf{J}(\mathbf{r},t)$. All six vector-valued quantities depend on the spatial co-ordinate $\mathbf{r}$ and the time t. *Gauss's law* (1) states that the total electric flux through the closed boundary $\partial\mathbf{\Omega}$ of a domain $\mathbf{\Omega}$ is equal to the total charge contained in the domain. Figure 1(a) depicts a volume containing the electric charge Q and indicates the electric displacement field as well as the infinitely small area element. *Gauss's law for magnetism* (2) requires the total magnetic flux through the closed boundary $\partial\mathbf{\Omega}$ of a domain $\mathbf{\Omega}$ to be equal to zero. Figure 1(b) shows an example volume, as well as the magnetic flux density and the infinitely small area element. In other words: the field lines of the magnetic flux density have neither a source nor a sink but are closed, owing to the absence of magnetic charges. *Faraday's law of induction* (3) states that the negative time derivative of the total magnetic flux through a surface $\mathbf{\Gamma}$ equals the integration of the electric field along the closed boundary $\partial\mathbf{\Gamma}$ of the surface, as highlighted in Fig. 1(c). Similarly, *Ampère's law with Maxwell's extension* (4) states that the total current through a surface $\mathbf{\Gamma}$ is equal to the integration of the magnetic field along the closed boundary $\partial\mathbf{\Gamma}$ of the surface, as sketched in Fig. 1(d). It is worth noting that Maxwell's equations in their integral form hold for arbitrary surfaces $\mathbf{\Gamma}$ or volumes $\mathbf{\Omega}$. For instance, $\mathbf{\Gamma}$ and $\mathbf{\Omega}$ can be independently chosen from material boundaries.

Maxwell's equations, in their integral representation (Eqs. (1)–(4)) can be transferred to their differential form,

$$\nabla\cdot\mathbf{D}(\mathbf{r},t) = \rho(\mathbf{r},t)\,, \tag{5}$$

$$\nabla\cdot\mathbf{B}(\mathbf{r},t) = 0\,, \tag{6}$$

$$\nabla\times\mathbf{E}(\mathbf{r},t) = -\frac{\partial}{\partial t}\mathbf{B}(\mathbf{r},t)\,, \tag{7}$$

$$\nabla\times\mathbf{H}(\mathbf{r},t) = \frac{\partial}{\partial t}\mathbf{D}(\mathbf{r},t)+\mathbf{J}(\mathbf{r},t)\,, \tag{8}$$

by considering infinitely small integration surfaces $\mathbf{\Gamma}$ and integration volumes $\mathbf{\Omega}$. Here, $\nabla\cdot$ is the divergence operator, which contains spatial derivatives. The divergence operator acts on a vector field and delivers scalar fields. These scalar fields describe the source strength of the field per unit volume. Note that in some textbooks the divergence is denoted div. In a Cartesian co-ordinate system, the divergence of the vector field $\mathbf{D}(x,y,z)$ is given by

$$\nabla\cdot\mathbf{D}(x,y,z) = \mathrm{div}\,\mathbf{D}(x,y,z) = \frac{\partial}{\partial x}D_x(x,y,z)+\frac{\partial}{\partial y}D_y(x,y,z)+\frac{\partial}{\partial z}D_z(x,y,z)\,, \tag{9}$$

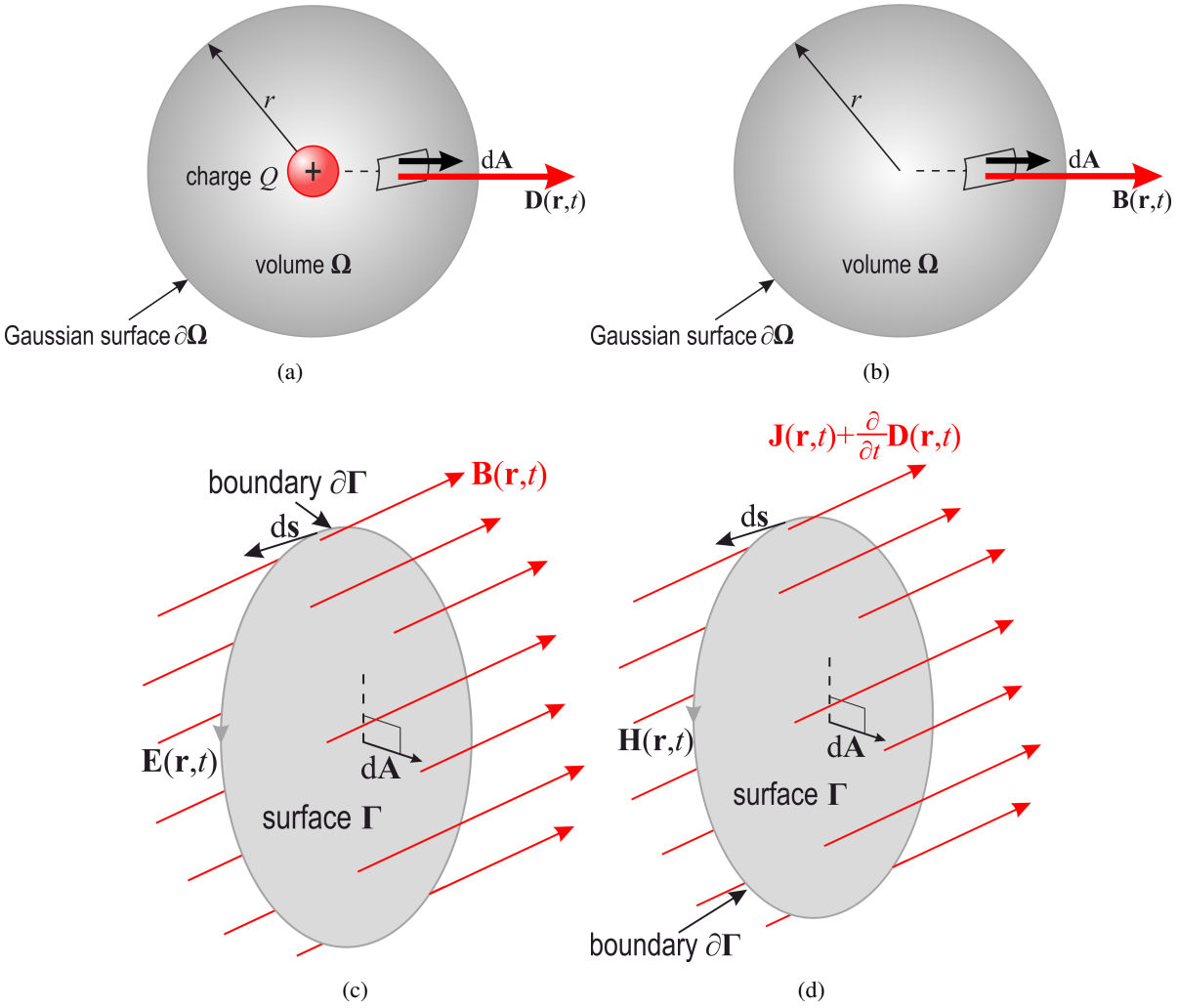
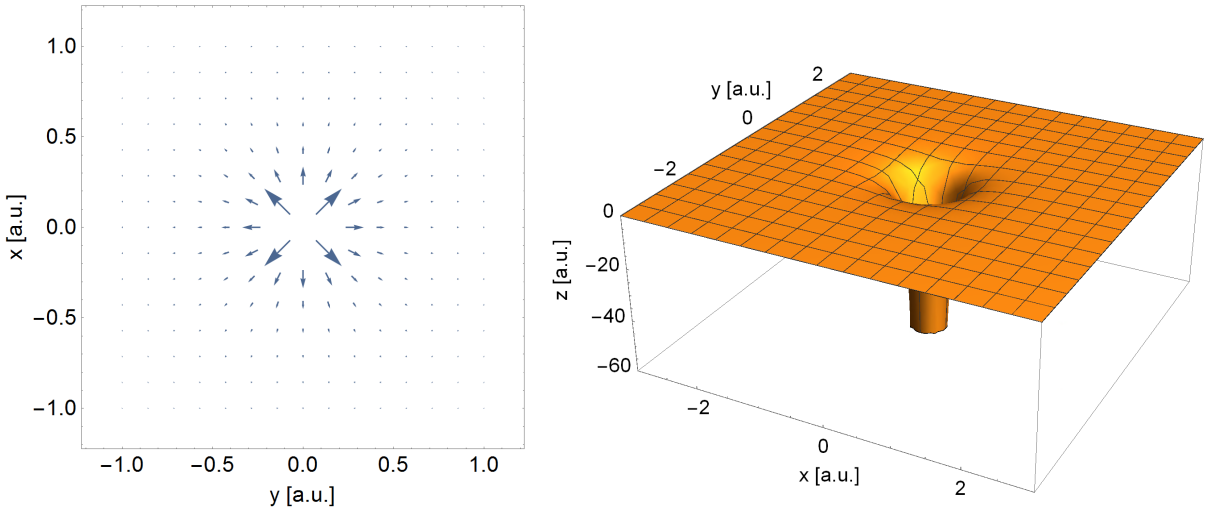


Fig. 1: Relation between electromagnetic quantities expressed by Maxwell's equations: (a) Gauss's law; (b) Gauss's law for magnetism; (c) Faraday's law of induction; (d) Ampère's law with Maxwell's extension. The figures closely are adapted from Ref. [6].

where $D_x(x, y, z)$, $D_y(x, y, z)$, and $D_z(x, y, z)$ are spatially dependent components of the vector field $\mathbf{D}(x, y, z)$. Figure 2(a) depicts an arbitrarily chosen vector field as an example of a divergence field. The divergence of this field, which is a scalar field, as the divergence gives the source strength, is shown in Fig. 2(b).

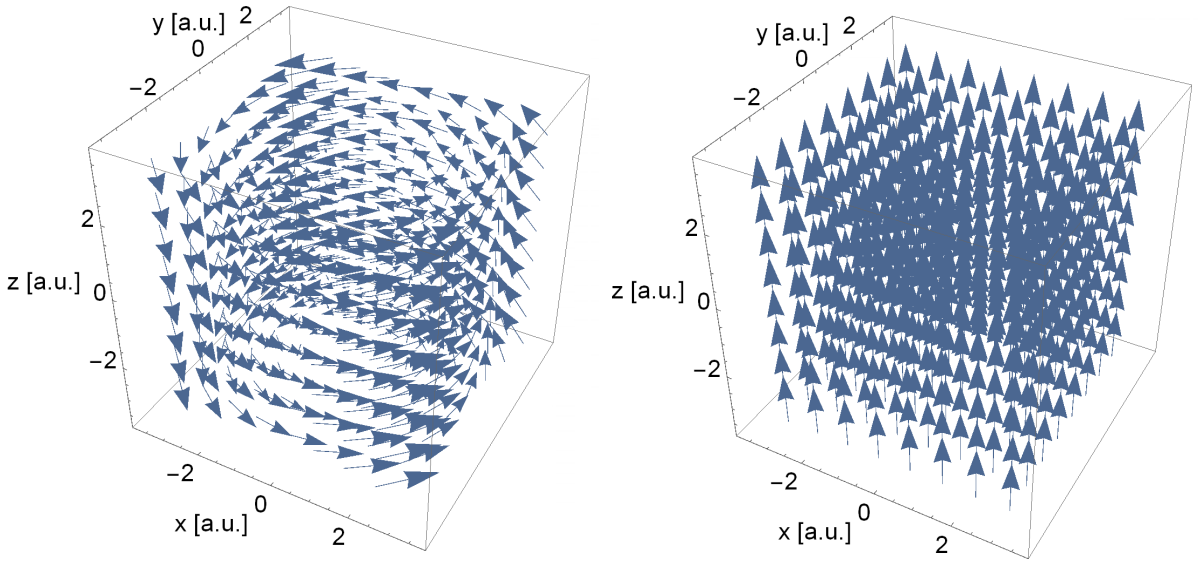
The operator $\nabla\times$ in Eqs. (7)–(8) represents the curl operator, which computes the rotation of a vector field about a point. Like the divergence operator, the curl operator contains spatial derivatives. In contrast with the divergence operator, the curl operator acts on vector fields and delivers vector fields. The vector fields resulting from the curl operator describe the curl strength of the field per unit area. In some textbooks, the curl is abbreviated by curl or rot. In a Cartesian co-ordinate system, the curl of the vector field $\mathbf{E}(x, y, z)$ is determined by

$$\nabla \times \mathbf{E}(x, y, z) = \operatorname{curl} \mathbf{E}(x, y, z) = \begin{vmatrix} \mathbf{e}_x & \mathbf{e}_y & \mathbf{e}_z \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ E_x(x, y, z) & E_y(x, y, z) & E_z(x, y, z) \end{vmatrix}$$



(a) (b)

Fig. 2: (a) Vector field $\mathbf{D}(x, y, z) = [x, y, 0]^{\mathrm{T}}/\sqrt{x^2 + y^2}^3$ as an example of a divergence field. (b) Divergence of the vector field, $\nabla \cdot \mathbf{D}(x, y, z) = -1/\sqrt{x^2 + y^2}^3$, shown in (a).



(a) (b)

Fig. 3: (a) Vector field $\mathbf{E}(x, y, z) = [-y, x, 0]^{\mathrm{T}}$ as an example of a curl field. (b) $\nabla \times \mathbf{E}(x, y, z) = [0, 0, 2]^{\mathrm{T}}$, i.e., the curl of the vector field depicted in (a).

$$= \begin{pmatrix} \frac{\partial}{\partial y} E_z(x, y, z) - \frac{\partial}{\partial z} E_y(x, y, z) \\ \frac{\partial}{\partial z} E_x(x, y, z) - \frac{\partial}{\partial x} E_z(x, y, z) \\ \frac{\partial}{\partial x} E_y(x, y, z) - \frac{\partial}{\partial y} E_x(x, y, z) \end{pmatrix}, \tag{10}$$

with the unit vectors $\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z$.

Figure 3(a) sketches an arbitrarily chosen vector field which is curled. Figure 3(b) plots the corresponding curl of this field, which is again a vector field.

1.2 Lorentz force

The interaction of electromagnetic fields with a charged particle is described by

$$\mathbf{F}(\mathbf{r},t) = q\left(\mathbf{E}(\mathbf{r},t) + \mathbf{v}\times\mathbf{B}(\mathbf{r},t)\right), \tag{11}$$

with the Lorentz force $\mathbf{F}(\mathbf{r},t)$ acting on a particle moving with velocity $\mathbf{v}$ and carrying charge q. Note that in some textbooks the term Lorentz force exclusively refers to $q\,\mathbf{v}\times\mathbf{B}(\mathbf{r},t)$, whereas the remaining term $q\,\mathbf{E}(\mathbf{r},t)$ is called the Coulomb force.

Independent of naming conventions, Eq. (11) is of crucial relevance for particle accelerators, as electric fields are used to transfer energy to the charged particles and magnetic fields to bend particles and to focus bunches of particles. The force on charged particles due to an electric field is always parallel to the electrical field lines (in fact that is one way to define electric fields). The force on charged particles due to a magnetic field is always orthogonal to the velocity $\mathbf{v}$ of the charged particle. The direction of the force acting on the particle due to magnetic fields can be found using the right-hand rule, where the thumb reflects the velocity $\mathbf{v}$ of the particle and the forefinger the direction of the magnetic flux density $\mathbf{B}(\mathbf{r},t)$. The direction of the forces is indicated by the direction of the middle finger. The Lorentz force is given component-wise in a Cartesian system by

$$F_x(\mathbf{r},t) = q\left(E_x(\mathbf{r},t) + v_y\,B_z(\mathbf{r},t) - v_z\,B_y(\mathbf{r},t)\right), \tag{12}$$
$$F_y(\mathbf{r},t) = q\left(E_y(\mathbf{r},t) + v_z\,B_x(\mathbf{r},t) - v_x\,B_z(\mathbf{r},t)\right), \tag{13}$$
$$F_z(\mathbf{r},t) = q\left(E_z(\mathbf{r},t) + v_x\,B_y(\mathbf{r},t) - v_y\,B_x(\mathbf{r},t)\right). \tag{14}$$

1.3 Conservation of charges

Maxwell's equations inherently comprise the conservation of electrical charges. To show this, the divergence of Ampère's law with Maxwell's extension (4) is determined as

$$\underbrace{\nabla\cdot\Big(\nabla\times\mathbf{H}(\mathbf{r},t)\Big)}_{0} = \nabla\cdot\frac{\partial}{\partial t}\mathbf{D}(\mathbf{r},t) + \nabla\cdot\mathbf{J}(\mathbf{r},t)\,. \tag{15}$$

The left-hand side is equal to zero as the divergence of any curl operations equals zero. On the right-hand side, the order of the derivation with respect to space and with respect to time can be exchanged by means of Schwarz's theorem, i.e., $\nabla\cdot\partial_t\mathbf{D}(\mathbf{r},t) = \partial_t\nabla\cdot\mathbf{D}(\mathbf{r},t)$, where ∂_t denotes the derivative with respect to time. Using Eq. (5) to replace the divergence of the displacement current delivers a differential form of the charge conservation formula,

$$\nabla\cdot\mathbf{J}(\mathbf{r},t) = -\frac{\partial}{\partial t}\rho(\mathbf{r},t)\,. \tag{16}$$

This formula can be transferred to an integral form through integration over the domain and the application of Gauss' law:

$$\oiint_{\partial\mathbf{\Omega}}\mathbf{J}(\mathbf{r},t)\cdot\mathrm{d}\mathbf{A} = -\frac{\partial}{\partial t}\underbrace{\iiint_{\mathbf{\Omega}}\rho(\mathbf{r},t)\,\mathrm{d}V}_{Q_{\mathrm{tot}}}\,. \tag{17}$$

The equation states that a total net electric flux through the closed surface $\partial\mathbf{\Omega}$ of the volume $\mathbf{\Omega}$ equals the reduction of the total electric charge in the volume $\mathbf{\Omega}$. Equation (16) states the same for an infinitely small volume, $\mathbf{\Omega}$.

1.4 Conservation of energy or Poynting's theorem

In addition to the conservation of charges, Maxwell's equations comprise the conservation of energy. This results from the multiplication of Eq. (7) with the magnetic field $\mathbf{H}(\mathbf{r},t)$, the multiplication of Eq. (8) with the electric field $\mathbf{E}(\mathbf{r},t)$, and the subtraction of both equations:

$$\begin{aligned}\mathbf{H}(\mathbf{r},t)\cdot\big(\nabla\times\mathbf{E}(\mathbf{r},t)\big)-\mathbf{E}(\mathbf{r},t)\cdot\big(\nabla\times\mathbf{H}(\mathbf{r},t)\big)\\ =-\underbrace{\mathbf{H}(\mathbf{r},t)\cdot\frac{\partial}{\partial t}\mathbf{B}(\mathbf{r},t)}_{\partial_t w_\mathrm{m}(\mathbf{r},t)}-\underbrace{\mathbf{E}(\mathbf{r},t)\cdot\frac{\partial}{\partial t}\mathbf{D}(\mathbf{r},t)}_{\partial_t w_\mathrm{e}(\mathbf{r},t)}-\underbrace{\mathbf{E}(\mathbf{r},t)\cdot\mathbf{J}(\mathbf{r},t)}_{p_\mathrm{int}(\mathbf{r},t)}\,,\end{aligned}\tag{18}$$

where $\partial_t w_\mathrm{e}(\mathbf{r},t)$ and $\partial_t w_\mathrm{m}(\mathbf{r},t)$ describe the change of energy stored in an infinitely small volume element in the form of electric and magnetic fields, respectively. The loss of energy per volume element is denoted $p_\mathrm{int}(\mathbf{r},t)$. Simplifying the left-hand side of Eq. (18) by means of the vector calculus identity $\nabla\cdot\big(\mathbf{E}(\mathbf{r},t)\times\mathbf{H}(\mathbf{r},t)\big)=\mathbf{H}(\mathbf{r},t)\cdot\big(\nabla\times\mathbf{E}(\mathbf{r},t)\big)-\mathbf{E}(\mathbf{r},t)\cdot\big(\nabla\times\mathbf{H}(\mathbf{r},t)\big)$ (see e.g., Ref. [3, p. 879, (I. 24)]) and shifting $p_\mathrm{int}(\mathbf{r},t)$ from the right-hand side to the left-hand side of the equation gives

$$p_\mathrm{int}(\mathbf{r},t)+\nabla\cdot\big(\underbrace{\mathbf{E}(\mathbf{r},t)\times\mathbf{H}(\mathbf{r},t)}_{\mathbf{S}(\mathbf{r},t)}\big)=-\frac{\partial}{\partial t}w_\mathrm{e}(\mathbf{r},t)-\frac{\partial}{\partial t}w_\mathrm{m}(\mathbf{r},t)\,.\tag{19}$$

The cross product of the electric field with the magnetic field is often referred to as the Poynting vector, $\mathbf{S}(\mathbf{r},t)$. It specifies the flow of energy per unit area due to electromagnetic fields. Integrating Eq. (19) over the volume $\mathbf{\Omega}$ and applying Gauss's divergence theorem yields

$$\underbrace{\iiint_{\mathbf{\Omega}}p_\mathrm{int}(\mathbf{r},t)\,\mathrm{d}V}_{P_\mathrm{int}(t)}+\underbrace{\oiint_{\partial\mathbf{\Omega}}\mathbf{S}(\mathbf{r},t)\cdot\mathrm{d}\mathbf{A}}_{P_\mathrm{ext}(t)}=-\frac{\partial}{\partial t}\underbrace{\iiint_{\mathbf{\Omega}}w_\mathrm{e}(\mathbf{r},t)\,\mathrm{d}V}_{W_\mathrm{e}(t)}-\frac{\partial}{\partial t}\underbrace{\iiint_{\mathbf{\Omega}}w_\mathrm{m}(\mathbf{r},t)\,\mathrm{d}V}_{W_\mathrm{m}(t)}\,.\tag{20}$$

The equation balances the internal energy losses, $P_\mathrm{int}(t)$, in $\mathbf{\Omega}$ and the energy flowing through the boundary $\partial\mathbf{\Omega}$, often referred to as external energy losses, $P_\mathrm{ext}(t)$, with the change of total energy stored in electric, $W_\mathrm{e}(t)$, and magnetic, $W_\mathrm{m}(t)$, fields. The internal losses, $P_\mathrm{int}(t)$, can result from ohmic dissipation effects as a consequence of finite resistivity and dielectric or magnetic losses. Furthermore, the transfer of energy to charged particles residing in $\mathbf{\Omega}$ is covered in $P_\mathrm{int}(t)$ as well. The external losses, $P_\mathrm{ext}(t)$, account, e.g., for the propagation of energy through waveguide ports. Note that $P_\mathrm{int}(t)$ and $P_\mathrm{ext}(t)$ can become negative, for instance if charged particles are decelerated in $\mathbf{\Omega}$ or energy propagates into the structure through $\partial\mathbf{\Omega}$, owing to an external excitation of the waveguide ports. Consequently, $P_\mathrm{int}(t)$ and $P_\mathrm{ext}(t)$ can also model energy gains.

2 Interaction of electromagnetic fields with matter

The interaction of electromagnetic fields with matter is described by the material equations. They relate the electric flux density $\mathbf{D}(\mathbf{r},t)$ with the electric field strength $\mathbf{E}(\mathbf{r},t)$ (see Subsection 2.1), the magnetic flux density $\mathbf{B}(\mathbf{r},t)$ with the magnetic field strength $\mathbf{H}(\mathbf{r},t)$ (see Subsection 2.2), and the electric current density $\mathbf{J}(\mathbf{r},t)$ with the electric field strength $\mathbf{E}(\mathbf{r},t)$ (see Subsection 2.3). All subsequent material equations describe the interaction of electromagnetic fields with matter, based on a macroscopic level, accounting for a sufficiently large number of molecules and atoms. Typically, these interactions are transient effects, as polarization of matter does not take place instantaneously. These transient effects are not considered hereinafter. In other words, it is assumed that the polarization takes place much faster than the change of the respective field strengths. If this does not hold, the dynamic properties of polarizations are often accounted for by frequency-dependent and complex-valued material properties, as discussed for instance in Refs. [1, 2]. Furthermore, the properties of some materials depend on the

direction. For these cases of anisotropy, it is not sufficient to describe the material properties by means of scalars but tensors are required. However, these cases are not discussed within this report, nor is the class of non-linear materials, such as iron.

2.1 Electric fields

The interaction of matter with electric fields is described by

$$\mathbf{D}(\mathbf{r},t) = \varepsilon_0 \mathbf{E}(\mathbf{r},t) + \mathbf{P}(\mathbf{r},t)\,, \tag{21}$$

where $\varepsilon_0 = 8.8542 \times 10^{-12}\,\mathrm{As/Vm}$ is the permittivity of free space and $\mathbf{P}(\mathbf{r},t)$ is the electric polarization vector. The polarization vector $\mathbf{P}(\mathbf{r},t)$ models the electric polarization of the material as a result of an external electric field. In fact, the following three polarization effects [2] can be distinguished.

- Dipole polarization results from the fact that polar materials, such as water, have a permanent dipole moment. In the presence of an electric field, the dipoles align themselves parallel to the electric field lines.
- Molecular polarization describes the displacement of positively or negatively charged atoms in a molecule.
- Electronic polarization is referred to as the displacement of an electron cloud centre in the vicinity of an atomic nucleus.

Figure 4 depicts these mechanisms. The electric polarization depends on the electric field via

$$\mathbf{P}(\mathbf{r},t) = \varepsilon_0 \chi_\mathrm{e} \mathbf{E}(\mathbf{r},t)\,, \tag{22}$$

where the material property χ_e is referred to as the electric susceptibility. Combining Eq. (21) with Eq. (22) gives

$$\mathbf{D}(\mathbf{r},t) = \varepsilon_0 \mathbf{E}(\mathbf{r},t) + \varepsilon_0 \chi_\mathrm{e} \mathbf{E}(\mathbf{r},t) = \varepsilon_0 \underbrace{(1+\chi_\mathrm{e})}_{\varepsilon_\mathrm{r}} \mathbf{E}(\mathbf{r},t)\,, \tag{23}$$

with the relative dielectric constant ε_r. Typical values for ε_r are 1.0006 for air, 3.8 for quartz, and 80 for water.

2.2 Magnetic fields

The interaction between matter and magnetic fields is characterized by

$$\mathbf{B}(\mathbf{r},t) = \mu_0 \mathbf{H}(\mathbf{r},t) + \mathbf{M}(\mathbf{r},t)\,, \tag{24}$$

with the permeability of free space $\mu_0 = 4\pi \times 10^{-7}\,\mathrm{Vs/Am}$ and the magnetic polarization vector $\mathbf{M}(\mathbf{r},t)$. The polarization vector $\mathbf{M}(\mathbf{r},t)$ describes the magnetic polarization of the material in the presence of an external magnetic field strength. Employing a very simple atomic model, magnetic dipoles result from negatively charged electrons orbiting around positively charged nuclei or from electrons spinning around their own axes. As shown in Fig. 5, these orbiting electrons generate fields similar to those of a magnetic bar, i.e., fields of a magnetic dipole. The magnetic polarization is connected with magnetic field strength by means of

$$\mathbf{M}(\mathbf{r},t) = \mu_0 \chi_\mathrm{m} \mathbf{H}(\mathbf{r},t)\,, \tag{25}$$

where the material property χ_m is referred to as the magnetic susceptibility. Replacing Eq. (25) in Eq. (24) yields

$$\mathbf{B}(\mathbf{r},t) = \mu_0 \mathbf{H}(\mathbf{r},t) + \mu_0 \chi_\mathrm{m} \mathbf{H}(\mathbf{r},t) = \mu_0 \underbrace{(1+\chi_\mathrm{m})}_{\mu_\mathrm{r}} \mathbf{H}(\mathbf{r},t)\,, \tag{26}$$

where μ_r is the relative permeability. The relative permeability ranges from 0.999834 for bismuth to $1{,}000{,}000$ for supermalloy [2].

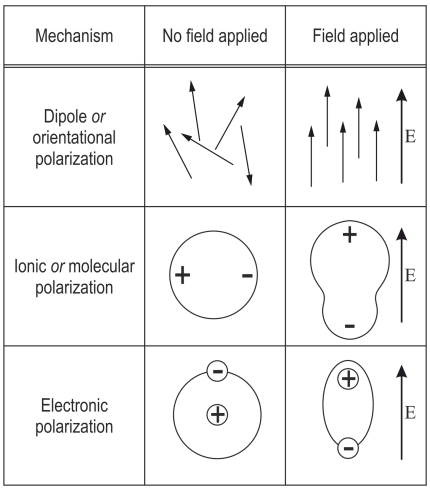


Fig. 4: Most relevant electrical polarization effects in dielectric materials. The behaviour of the dipoles is shown in the absence (middle column) and the presence (right column) of an external electric field $\mathbf{E}(\mathbf{r}, t)$. Different time constants can be attributed to each effect, so that the respective polarization effects occur at different frequencies. Adapted from Ref. [2].

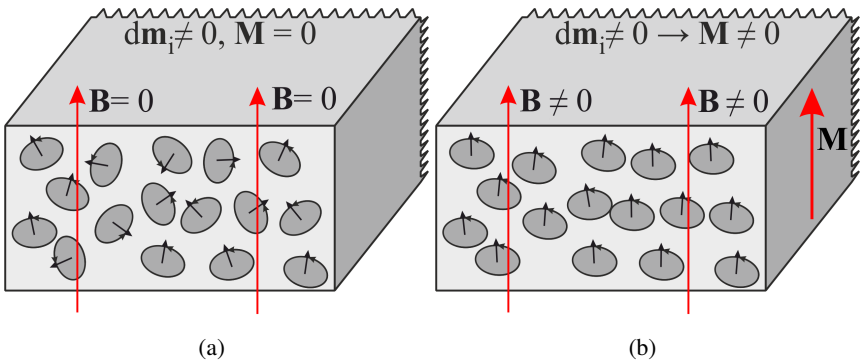


Fig. 5: Orientation of magnetic dipoles (grey circles). The arrows at the boundaries of the grey circles describe electric currents arising from negatively charged electrons orbiting around positively charged nuclei or from electrons spinning around their own axes. (a) Random orientation of magnetic dipoles in the absence of a magnetic flux density $\mathbf{B}(\mathbf{r}, t)$. (b) Alignment of magnetic dipoles in the presence of a magnetic flux density $\mathbf{B}(\mathbf{r}, t)$. Adapted from Ref. [2].

2.3 Ohm's law

If the electrons of a material can move from one atom to the other, they can drift as a result of an applied electric field, in consequence of Eq. (11). Owing to the drifting electron, an electric current flows. In good electric conductors, such as metals, the number of these free electrons is very large. The relationship between the electric current density due to ohmic conduction and the electric field strength is determined by

$$\mathbf{J}(\mathbf{r}, t) = \sigma\, \mathbf{E}(\mathbf{r}, t)\,, \tag{27}$$

with the electric conductivity σ. Depending on the electric conductivity, matter is classified as conducting, semiconducting, or dielectric (insulators). The conductivity σ varies from $1.33 \times 10^{-18}\,\mathrm{S/m}$ for fused quartz (an insulator) via $2.17\,\mathrm{S/m}$ for germanium (a semiconductor) to $6.29 \times 10^{7}\,\mathrm{S/m}$ for silver (a conductor) [6].

2.4 Boundary conditions of electromagnetic fields at material interfaces

With the help of Maxwell's equations in integral form (Eqs. (1)–(4)), boundary conditions of electromagnetic fields at the interface between different materials can be derived. They read

$$\mathbf{n} \times \Big[\mathbf{E}_2(\mathbf{r}_{\mathrm{if}}, t) - \mathbf{E}_1(\mathbf{r}_{\mathrm{if}}, t)\Big] = \mathbf{0}\,, \tag{28}$$

$$\mathbf{n} \times \Big[\mathbf{H}_2(\mathbf{r}_{\mathrm{if}}, t) - \mathbf{H}_1(\mathbf{r}_{\mathrm{if}}, t)\Big] = \mathbf{J}_{\mathrm{s}}(\mathbf{r}_{\mathrm{if}}, t)\,, \tag{29}$$

$$\mathbf{n} \cdot \Big[\mathbf{D}_2(\mathbf{r}_{\mathrm{if}}, t) - \mathbf{D}_1(\mathbf{r}_{\mathrm{if}}, t)\Big] = \rho_{\mathrm{s}}(\mathbf{r}_{\mathrm{if}}, t)\,, \tag{30}$$

$$\mathbf{n} \cdot \Big[\mathbf{B}_2(\mathbf{r}_{\mathrm{if}}, t) - \mathbf{B}_1(\mathbf{r}_{\mathrm{if}}, t)\Big] = 0\,, \tag{31}$$

where $\mathbf{n}$ denotes a unit vector normal to the interface, $\mathbf{r}_{\mathrm{if}}$ a location on the interface, $\mathbf{J}_{\mathrm{s}}(\mathbf{r}_{\mathrm{if}}, t)$ an electric surface current density, and $\rho_{\mathrm{s}}(\mathbf{r}_{\mathrm{if}}, t)$ an electric surface charge density. The latter two quantities refer to charges residing or flowing at the interface between the different materials. The subscripts of the fields indicate whether they refer to material 1 or to material 2.

Despite the fact that the statements (Eqs. (28)–(31)) look very formal, they have an intuitively accessible meaning. An explicit evaluation of the cross and the dot product reveals that Eq. (28) states that the tangential components of the electric fields are equal to each other across the interface, i.e., $E_{\mathrm{t},2}(\mathbf{r}_{\mathrm{if}}, t) = E_{\mathrm{t},1}(\mathbf{r}_{\mathrm{if}}, t)$. In the absence of electric surface currents ($\mathbf{J}_{\mathrm{s}}(\mathbf{r}_{\mathrm{if}}, t) = \mathbf{0}$), this also holds for tangential magnetic fields, i.e., $H_{\mathrm{t},2}(\mathbf{r}_{\mathrm{if}}, t) = H_{\mathrm{t},1}(\mathbf{r}_{\mathrm{if}}, t)$. In the presence of electric surface currents ($\mathbf{J}_{\mathrm{s}}(\mathbf{r}_{\mathrm{if}}, t) \neq \mathbf{0}$), the tangential fields are allowed to be discontinuous. While Eqs. (28) and (29) specify tangential components of the fields, Eqs. (30) and (31) refer to the normal components of the respective fields. Equation (30) claims that the normal component of the electric displacement fields is continuous in the absence of surface charges ($\rho_{\mathrm{s}}(\mathbf{r}_{\mathrm{if}}, t) = 0$), i.e., $D_{\mathrm{n},2}(\mathbf{r}_{\mathrm{if}}, t) = D_{\mathrm{n},1}(\mathbf{r}_{\mathrm{if}}, t)$. In the presence of surface charges, the normal components are related by $D_{\mathrm{n},2}(\mathbf{r}_{\mathrm{if}}, t) = D_{\mathrm{n},1}(\mathbf{r}_{\mathrm{if}}, t) + \rho_{\mathrm{s}}(\mathbf{r}_{\mathrm{if}}, t)$. Finally, Eq. (31) claims that the normal components of the magnetic flux density are equal to each other across the interface, i.e., $B_{\mathrm{n},2}(\mathbf{r}_{\mathrm{if}}, t) = B_{\mathrm{n},1}(\mathbf{r}_{\mathrm{if}}, t)$. Please refer to Fig. 6 for a graphical representation of the boundary conditions of electromagnetic fields at material interfaces. Note that the laws of refraction for the field lines can be derived directly from the boundary conditions (Eqs. (28)–(31)).

3 Electrostatics

To describe electric fields that do not depend on time, it is sufficient to restrict ourselves to electric and electric displacement fields because coupling between electric and magnetic field quantities does not take place in the static case, as the derivatives with respect to time are zero:

$$\nabla \times \mathbf{E}(\mathbf{r}) = -\underbrace{\frac{\partial}{\partial t}\mathbf{B}(\mathbf{r})}_{\mathbf{0}}\,, \tag{32}$$

$$\nabla \cdot \mathbf{D}(\mathbf{r}) = \rho(\mathbf{r})\,. \tag{33}$$

In a next step, the curl-free static electric field strength $\mathbf{E}(\mathbf{r})$ may be expressed in terms of the gradient of a scalar electrostatic potential $\phi(\mathbf{r})$:

$$\mathbf{E}(\mathbf{r}) = -\nabla\phi(\mathbf{r})\,. \tag{34}$$

Here, $\nabla\phi(\mathbf{r})$ denotes the gradient of the electric potential. The gradient describes the change of a scalar field with respect to the spatial co-ordinates. The gradient operator acts on a scalar field and delivers a vector field. In a Cartesian system, the gradient is determined by

$$\nabla\phi(x, y, z) = \begin{pmatrix} \frac{\partial}{\partial x}\phi(x, y, z) \\ \frac{\partial}{\partial y}\phi(x, y, z) \\ \frac{\partial}{\partial z}\phi(x, y, z) \end{pmatrix}\,. \tag{35}$$

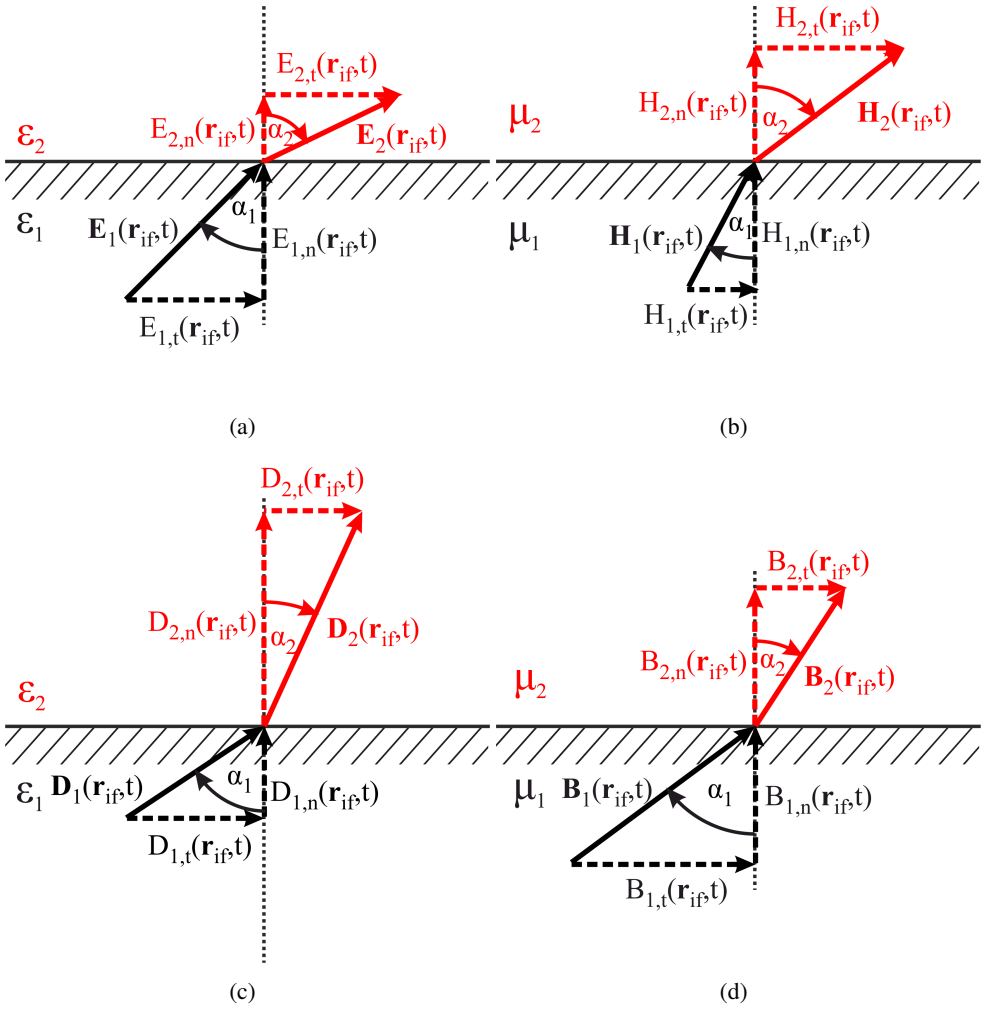


Fig. 6: Visualization of (a) electric fields, (b) magnetic fields, (c) electric displacement fields, and (d) magnetic flux densities across the interface of material 1 and material 2. At the interface, each vector field can be decomposed in terms of a normal component and a tangential component.

Expressing the static electric field $\mathbf{E}(\mathbf{r})$ in terms of the electrostatic potential $\phi(\mathbf{r})$, as in Eq. (34), has two advantages.

- The electric field $\mathbf{E}(\mathbf{r})$ is, *a priori*, free of curl, as $\nabla \times \nabla\phi(\mathbf{r}) = \mathbf{0}$, so that Eq. (32) holds in general. This can be seen by combining Eq. (35) with Eq. (10).
- Rather than searching for three unknown functions representing the components of the vector $\mathbf{E}(\mathbf{r})$, the gradient approach allows the number of sought functions to be reduced to one, i.e., the electric scalar potential $\phi(\mathbf{r})$.

Substituting the electric field strength in Eq. (33) by the ansatz (Eq. (34)) yields

$$\nabla \cdot \Big[\underbrace{-\varepsilon(\mathbf{r})\nabla\phi(\mathbf{r})}_{\mathbf{D}(\mathbf{r})} \Big] = \rho(\mathbf{r})\,. \tag{36}$$

This partial differential equation is often referred to as a potential equation. Mathematically, it is a Poisson equation. It describes electric potentials and consequently electric fields in static scenarios and must

hold in the domain $\mathbf{\Omega}$. In the case of a homogeneous material distribution ($\varepsilon(\mathbf{r}) = \varepsilon = \text{const.}$), the permittivity is a constant, so can be shifted in front of the divergence operator. Dividing the resulting equation on both sides by ε delivers

$$\nabla \cdot \Big[\nabla\phi(\mathbf{r})\Big] = -\frac{\rho(\mathbf{r})}{\varepsilon}\,. \tag{37}$$

The differential operator on the left-hand side corresponds to the Laplace operator:

$$\Delta\phi(\mathbf{r}) = \nabla \cdot \Big[\nabla\phi(\mathbf{r})\Big]\,. \tag{38}$$

In a Cartesian system, the Laplace operator reads as

$$\Delta\phi(x,y,z) = \frac{\partial^2}{\partial x^2}\phi(x,y,z) + \frac{\partial^2}{\partial y^2}\phi(x,y,z) + \frac{\partial^2}{\partial z^2}\phi(x,y,z)\,. \tag{39}$$

The solution of the Poisson equation (Eq. (37)) is given by

$$\phi(\mathbf{r}) = \frac{1}{4\pi\varepsilon}\iiint\limits_{\mathbf{\Omega}} \frac{1}{|\mathbf{r}-\mathbf{r}'|}\,\rho(\mathbf{r}')\mathrm{d}\mathbf{r}'\,, \tag{40}$$

where the potential at infinity was set to zero, to ensure the uniqueness of the potential. Note that the static electric field strength directly results from the electrostatic potential by means of Eq. (34).

To have a unique solution for a specific field problem, boundary conditions are required. Important boundary conditions are the Dirichlet and the Neumann boundary conditions, prescribing fixed boundary values and normal derivatives, respectively. In electrostatics, the Dirichlet boundary condition is used to prescribe potentials on the boundary of the solution domain:

$$\phi(\mathbf{r}) = g_\mathrm{D}(\mathbf{r}) \text{ on } \partial\mathbf{\Omega}_\mathrm{D}\,, \tag{41}$$

with the function $g_\mathrm{D}(\mathbf{r})$ specifying the values of the potentials on the Dirichlet boundary $\partial\mathbf{\Omega}_\mathrm{D}$. This boundary condition is also used to model perfectly conducting boundaries. For these cases, constant potentials along the boundary are prescribed, so that the tangential component of the electric field vanishes. The very special case where $g_\mathrm{D}(\mathbf{r}) = 0$ is called the homogeneous Dirichlet boundary condition.

The Neumann boundary condition is employed to prescribe the change of the potential normal to the boundary of the computational domain:

$$\mathbf{n} \cdot \big[\nabla\phi(\mathbf{r})\big] = g_\mathrm{N}(\mathbf{r}) \text{ on } \partial\mathbf{\Omega}_\mathrm{N}. \tag{42}$$

Here, $\mathbf{n}$ is a unit vector normal to the boundary $\partial\mathbf{\Omega}_\mathrm{N}$ and $g_\mathrm{N}(\mathbf{r})$ is a function specifying the normal derivative of the potential on the Neumann boundary. To acquire a more intuitive understanding of Eq. (42), the gradient of the scalar electric potential is replaced by the electric field strength using Eq. (34):

$$\mathbf{n} \cdot \mathbf{E}(\mathbf{r}) = E_\mathrm{n}(\mathbf{r}) = -g_\mathrm{N}(\mathbf{r}) \text{ on } \partial\mathbf{\Omega}_\mathrm{N}\,. \tag{43}$$

The dot product of the normal vector $\mathbf{n}$ and the electric field strength $\mathbf{E}(\mathbf{r})$ delivers the normal component $E_\mathrm{n}(\mathbf{r})$ of the electric field strength on the Neumann boundary $\partial\mathbf{\Omega}_\mathrm{N}$. Following Eq. (43), the Neumann boundary condition allows the normal component of the electric field to be prescribed on the respective boundary of the computational domain. The special case where $g_\mathrm{N}(\mathbf{r}) = 0$ is called the homogeneous Neumann boundary condition and is often used to model perfect electric insulators. Figure 7 depicts the electric potential of a spherical charge as an example. The potential is shown for open boundary conditions (left picture) and for homogeneous Dirichlet boundary conditions (right picture). Open boundary conditions are not further discussed in this report. By employing certain approximation strategies, open

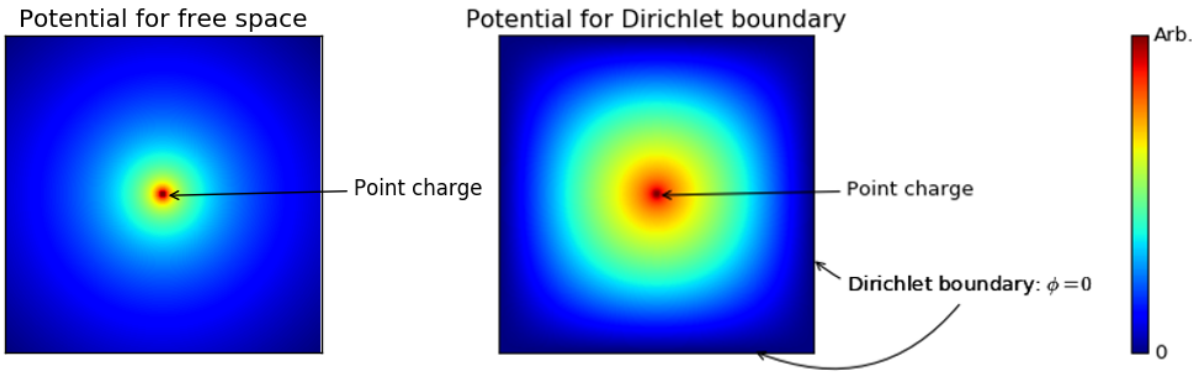


Fig. 7: Electric potential of a small spherical charge representing a single charge. The left picture shows the potential resulting from open boundaries; at the right picture, homogeneous Dirichlet conditions are enforced at the boundary of the computational domain. Both plots are created by means of numerical techniques using Ref. [11].

boundary conditions aim to mimic the computational domain to be extended to infinity. The influence of the boundary condition on the potential distribution inside the domain is clearly visible. In this example, for demonstration purposes, the boundary was set close to the source and obviously impacts the results.

In addition to the use of boundary conditions, Eqs. (41) and (42) are often used to specify symmetry conditions. The use of symmetries for symmetric field problems generally allows for a reduction of both memory requirements and computational burden.

4 Magnetostatics

As mentioned in Section 3, electric and magnetic fields are decoupled in the static case. Thus, it is sufficient to consider Eqs. (6) and (8) for magnetostatics:

$$\nabla \cdot \mathbf{B}(\mathbf{r}) = 0\,, \tag{44}$$

$$\nabla \times \mathbf{H}(\mathbf{r}) = \underbrace{\frac{\partial}{\partial t}\mathbf{D}(\mathbf{r})}_{\mathbf{0}} + \mathbf{J}(\mathbf{r})\,. \tag{45}$$

Fig. 8 depicts the magnetic flux density in an iron yoke with two air gaps as an example for a field distribution obeying Eq. (44) and Eq. (45).

Rather than solving directly for the magnetic flux density, the flux density is often expressed in terms of the curl of a so-called magnetic vector potential:

$$\mathbf{B}(\mathbf{r}) = \nabla \times \mathbf{A}(\mathbf{r})\,. \tag{46}$$

The advantage of this approach results from the fact that Eq. (44) is then, *a priori*, fulfilled as the divergence of a field resulting from the curl operator always equalling zero. This follows, for instance, from Eqs. (9) and (10) for Cartesian co-ordinate systems. It is worth mentioning that the magnetic flux density $\mathbf{B}(\mathbf{r})$ does not uniquely determine the magnetic vector potential $\mathbf{A}(\mathbf{r})$. Adding a gradient of a scalar field, the magnetic vector potential does not influence the resulting magnetic flux density,

$$\mathbf{B}(\mathbf{r}) = \nabla \times \big[\mathbf{A}(\mathbf{r}) + \nabla\phi(\mathbf{r})\big] = \nabla \times \mathbf{A}(\mathbf{r}) + \underbrace{\nabla \times \nabla\phi(\mathbf{r})}_{\mathbf{0}} = \nabla \times \mathbf{A}(\mathbf{r})\,, \tag{47}$$

as the vector operators are linear and the curl of the gradient of scalar fields equals zero. For the static problem, Coulomb gauging is appropriate to achieve uniqueness:

$$\nabla \cdot \mathbf{A}(\mathbf{r}) = 0\,. \tag{48}$$

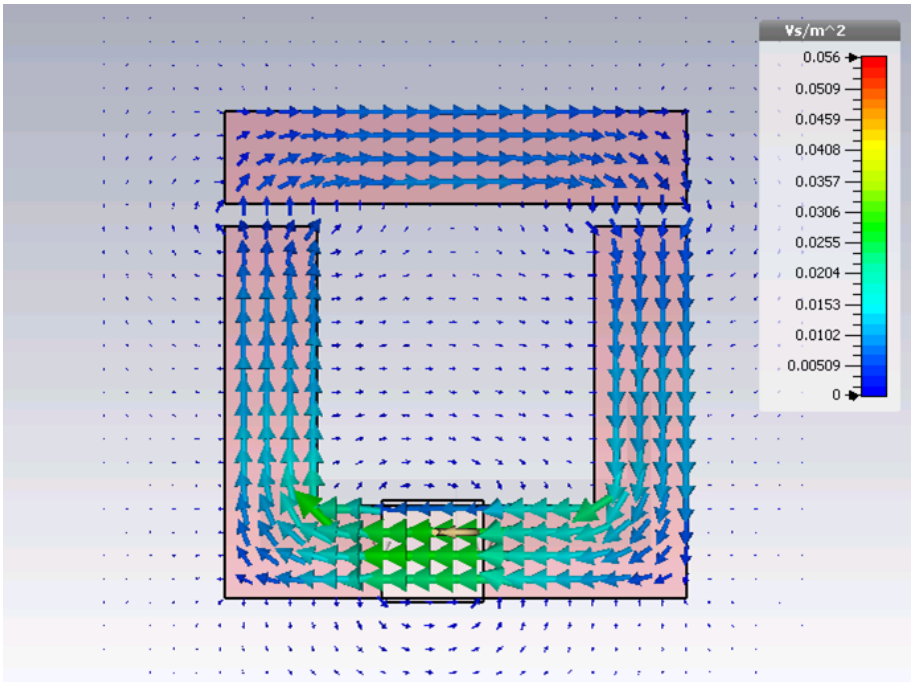


Fig. 8: Static field distribution of the magnetic flux density $\mathbf{B}(\mathbf{r})$ in an iron yoke with two air gaps. The magnetic field arises from electric currents flowing in the coil at the bottom of the yoke, where the magnetic field strength is largest. The field distribution is determined numerically using commercial software [12].

Substituting the magnetic field strength in Eq. (45) using the respective material relation and employing the ansatz (Eq. (46)) delivers

$$\nabla \times \left[\frac{1}{\mu(\mathbf{r})} \nabla \times \mathbf{A}(\mathbf{r}) \right] = \mathbf{J}(\mathbf{r}) \,. \tag{49}$$

Assuming the material to be homogeneous ($\mu(\mathbf{r}) = \mu = \text{const.}$), the permittivity is a constant, so it can be shifted in front of the divergence operator to obtain

$$\nabla \times \nabla \times \mathbf{A}(\mathbf{r}) = \mu \mathbf{J}(\mathbf{r}) \,. \tag{50}$$

Note that the material distribution of the example structure depicted in Fig. 8 is inhomogeneous as vacuum and iron are present in the domain. The statement Eq. (50) can be modified by means of the vector calculus identity,

$$\nabla \times \nabla \times \mathbf{A}(\mathbf{r}) = \nabla \left(\nabla \cdot \mathbf{A}(\mathbf{r}) \right) - \Delta \mathbf{A}(\mathbf{r}) \,, \tag{51}$$

to

$$\nabla \left(\nabla \cdot \mathbf{A}(\mathbf{r}) \right) - \Delta \mathbf{A}(\mathbf{r}) = \mu \mathbf{J}(\mathbf{r}) \,. \tag{52}$$

This can be simplified to the vector Poisson equation,

$$\Delta \mathbf{A}(\mathbf{r}) = -\mu \mathbf{J}(\mathbf{r}) \,, \tag{53}$$

by enforcing the Coulomb gauge (Eq. (48)) for the magnetostatic potential. In a Cartesian co-ordinate system, the vector Laplace operator reads as

$$\Delta \mathbf{A}(x,y,z) = \begin{pmatrix} \frac{\partial^2}{\partial x^2} A_x(x,y,z) + \frac{\partial^2}{\partial y^2} A_x(x,y,z) + \frac{\partial^2}{\partial z^2} A_x(x,y,z) \\ \frac{\partial^2}{\partial x^2} A_y(x,y,z) + \frac{\partial^2}{\partial y^2} A_y(x,y,z) + \frac{\partial^2}{\partial z^2} A_y(x,y,z) \\ \frac{\partial^2}{\partial x^2} A_z(x,y,z) + \frac{\partial^2}{\partial y^2} A_z(x,y,z) + \frac{\partial^2}{\partial z^2} A_z(x,y,z) \end{pmatrix} , \tag{54}$$

where $A_x(x,y,z)$, $A_y(x,y,z)$, and $A_z(x,y,z)$ are the components of the magnetic vector potential $\mathbf{A}(x,y,z)$. It follows from Eq. (54) that the vector Laplace operator, in fact, works as a Laplace operator on each individual component of the vector field. Note that the vector Laplace operator acts on a vector and delivers a vector.

In close analogy to Eq. (40), the solution of the vector Poisson equation (Eq. (53)) is given by

$$\mathbf{A}(\mathbf{r}) = \frac{\mu}{4\pi} \iiint\limits_{\Omega} \frac{1}{|\mathbf{r}-\mathbf{r}'|}\, \mathbf{J}(\mathbf{r}')\mathrm{d}\mathbf{r}' \,. \tag{55}$$

The magnetic field strength is uniquely determined by the magnetic vector potential by the employment of the Coulomb gauge (Eq. (48)). For a specific field problem, boundary conditions similar to Eqs. (41) and (42) can be formulated.

5 Electromagnetic waves

A closer inspection of Maxwell's equations reveals that these equations are mutually coupled. Before quantifying the coupling in a formal way, a qualitative description of the interplay between the field quantities is presented.

1. A time-dependent electric displacement field $\mathbf{D}(\mathbf{r},t)$ generates a time-dependent curled magnetic field $\mathbf{H}(\mathbf{r},t)$ according to Eq. (4).
2. This time-dependent magnetic field $\mathbf{H}(\mathbf{r},t)$ corresponds to a time-dependent magnetic flux density $\mathbf{B}(\mathbf{r},t)$, owing to the material relation (Eq. (26)).
3. A time-dependent magnetic flux density $\mathbf{B}(\mathbf{r},t)$ creates a curled electric field strength $\mathbf{E}(\mathbf{r},t)$ as stated in Eq. (3).
4. Owing to Eq. (23), this electric field strength $\mathbf{E}(\mathbf{r},t)$ is proportional to the electric displacement field $\mathbf{D}(\mathbf{r},t)$. In other words, a time-dependent electric displacement field $\mathbf{D}(\mathbf{r},t)$ is generated and the process can be started again from 1.

For the formal description of electromagnetic problems, Maxwell's equations are often not employed directly, since each equation has more than one unknown field quantity to be specified. Instead, Maxwell's equations are inserted into each other so that they describe one field quantity exclusively. The resulting equations are referred to as wave equations and can be deduced for electric and for magnetic fields. The described approach, however, increases the order of the time derivative from one to two. It is worth highlighting that, despite the fact that the wave equation describes only one field quantity, the remaining quantities are generally unequal to zero and can be computed by means of Maxwell's equations. Hereinafter, the wave equations for electric and magnetic fields are derived. The derived wave equations allow for a multitude of different solutions, as discussed in Refs. [1–8] (See Fig. 9). In Subsection 5.3, the most simple frequency-domain solution of these wave equations is presented in closer detail as an example.

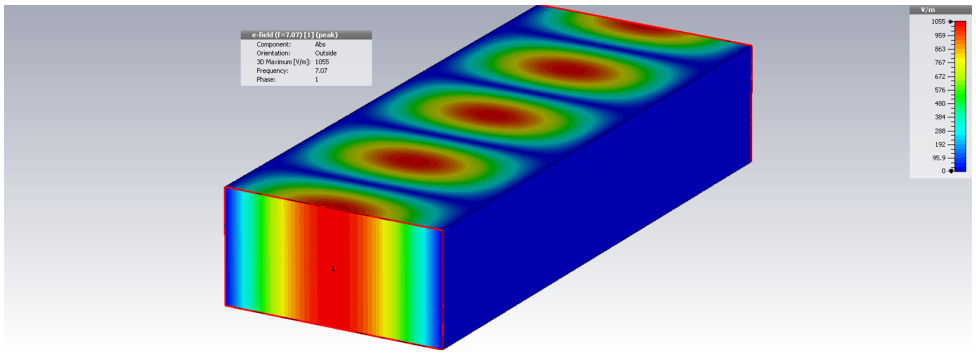
5.1 Wave equation for electric fields

In a first step, the curl of Faraday's law of induction (Eq. (7)) is evaluated:

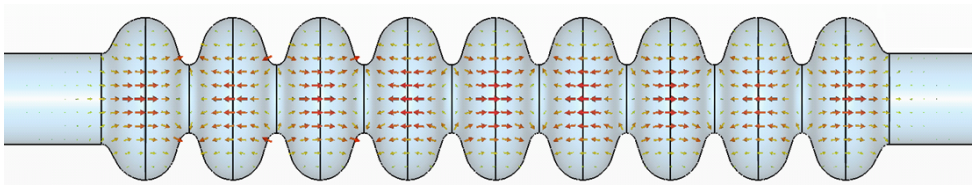
$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r},t) = \nabla \times \left[-\frac{\partial}{\partial t} \mathbf{B}(\mathbf{r},t) \right] . \tag{56}$$

Note that the minus sign can be shifted in front of the curl operator as it corresponds to the constant (-1) and that the order of derivatives can be changed as a consequence of Schwarz's theorem:

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r},t) = -\frac{\partial}{\partial t} \nabla \times \mathbf{B}(\mathbf{r},t) \,. \tag{57}$$



(a)



(b)

Fig. 9: Arbitrarily chosen examples for electromagnetic waves computed using commercial software [12]. (a) Absolute value of the electric field of a propagating TE_{10} mode in a rectangular waveguide. (b) Electric field of a standing wave in a third-harmonic TESLA cavity [13, 14] with a phase shift of π from cell to cell. This mode type is typically used for particle acceleration.

Next, the magnetic flux density is replaced by the magnetic field strength using Eq. (26):

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r},t) = -\frac{\partial}{\partial t}\nabla \times [\mu \mathbf{H}(\mathbf{r},t)] \; . \tag{58}$$

Provided that the material distribution is homogeneous, the permeability μ does not depend on spatial co-ordinates, so it can be shifted in front of the operators:

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r},t) = -\mu\frac{\partial}{\partial t}\nabla \times \mathbf{H}(\mathbf{r},t) \, . \tag{59}$$

The curl of the magnetic field on the right-hand side is now replaced by employing Ampère's law with Maxwell's extension (Eq. (8)) to obtain

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r},t) = -\mu\frac{\partial^2}{\partial t^2}\mathbf{D}(\mathbf{r},t) - \mu\frac{\partial}{\partial t}\mathbf{J}(\mathbf{r},t) \, . \tag{60}$$

Replacing the electric flux density by the electric field strength by using Eq. (23) yields the wave equation for electric fields in homogeneous media, often also denoted the curl-curl equation

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r},t) + \varepsilon\mu\frac{\partial^2}{\partial t^2}\mathbf{E}(\mathbf{r},t) = -\mu\frac{\partial}{\partial t}\mathbf{J}(\mathbf{r},t) \, . \tag{61}$$

By employing the vector calculus identity (Eq. (51)), the statement can be transferred to

$$\nabla \left(\nabla \cdot \mathbf{E}(\mathbf{r},t)\right) - \Delta\mathbf{E}(\mathbf{r},t) + \varepsilon\mu\frac{\partial^2}{\partial t^2}\mathbf{E}(\mathbf{r},t) = -\mu\frac{\partial}{\partial t}\mathbf{J}(\mathbf{r},t) \, . \tag{62}$$

As a result of Eqs. (5) and (23), the divergence of the electric field in homogeneous media is the charge density over the permittivity. Employing this property finally delivers the wave equation for electric fields,

$$\Delta\mathbf{E}(\mathbf{r},t) - \varepsilon\mu\frac{\partial^2}{\partial t^2}\mathbf{E}(\mathbf{r},t) = \mu\frac{\partial}{\partial t}\mathbf{J}(\mathbf{r},t) + \frac{1}{\varepsilon}\nabla\rho(\mathbf{r},t) \, . \tag{63}$$

5.2 Wave equation for magnetic fields

To derive the wave equation for magnetic fields, the curl of Ampère's law with Maxwell's extension (Eq. (8)) is initially evaluated to obtain

$$\nabla \times \nabla \times \mathbf{H}(\mathbf{r},t) = \nabla \times \left[\frac{\partial}{\partial t}\mathbf{D}(\mathbf{r},t) + \mathbf{J}(\mathbf{r},t)\right] = \nabla \times \left[\frac{\partial}{\partial t}\mathbf{D}(\mathbf{r},t)\right] + \nabla \times \mathbf{J}(\mathbf{r},t)\,. \tag{64}$$

The order of spatial and time derivatives is exchanged by the application of Schwarz's theorem:

$$\nabla \times \nabla \times \mathbf{H}(\mathbf{r},t) = \frac{\partial}{\partial t}\nabla \times \mathbf{D}(\mathbf{r},t) + \nabla \times \mathbf{J}(\mathbf{r},t)\,. \tag{65}$$

Subsequently, the electric flux density is substituted by the electric field strength using Eq. (23):

$$\nabla \times \nabla \times \mathbf{H}(\mathbf{r},t) = \frac{\partial}{\partial t}\nabla \times [\varepsilon\mathbf{E}(\mathbf{r},t)] + \nabla \times \mathbf{J}(\mathbf{r},t)\,. \tag{66}$$

Assuming the material distribution to be homogeneous, the permittivity ε does not depend on spatial co-ordinates, so this material property can be moved in front of the operators:

$$\nabla \times \nabla \times \mathbf{H}(\mathbf{r},t) = \varepsilon\frac{\partial}{\partial t}\nabla \times \mathbf{E}(\mathbf{r},t) + \nabla \times \mathbf{J}(\mathbf{r},t)\,. \tag{67}$$

The curl of the electric field on the right-hand side is now substituted by using Faraday's law of induction (Eq. (7)):

$$\nabla \times \nabla \times \mathbf{H}(\mathbf{r},t) = -\varepsilon\frac{\partial^2}{\partial t^2}\mathbf{B}(\mathbf{r},t) + \nabla \times \mathbf{J}(\mathbf{r},t)\,. \tag{68}$$

Expressing the magnetic flux density by means of the magnetic field strength according to Eq. (26) delivers the wave equation or curl-curl equation for magnetic fields in homogeneous media:

$$\nabla \times \nabla \times \mathbf{H}(\mathbf{r},t) = -\varepsilon\mu\frac{\partial^2}{\partial t^2}\mathbf{H}(\mathbf{r},t) + \nabla \times \mathbf{J}(\mathbf{r},t)\,. \tag{69}$$

The vector calculus identity (Eq. (51)) allows this statement to be transferred to:

$$\nabla\left(\nabla \cdot \mathbf{H}(\mathbf{r},t)\right) - \Delta\mathbf{H}(\mathbf{r},t) = -\varepsilon\mu\frac{\partial^2}{\partial t^2}\mathbf{H}(\mathbf{r},t) + \nabla \times \mathbf{J}(\mathbf{r},t)\,. \tag{70}$$

Combining Eqs. (6) and (26) shows that divergence of the magnetic field in homogeneous media is equal to zero. Accounting for this property finally yields the wave equation for magnetic fields:

$$\Delta\mathbf{H}(\mathbf{r},t) - \varepsilon\mu\frac{\partial^2}{\partial t^2}\mathbf{H}(\mathbf{r},t) = -\nabla \times \mathbf{J}(\mathbf{r},t)\,. \tag{71}$$

5.3 Plane waves

The simplest solution of the wave equations is a harmonic plane wave in free space. For this case, the excitation current density ($\mathbf{J}(\mathrm{r},t) = \mathbf{0}$) is assumed to be zero, as is the charge density ($\rho(\mathrm{r},t) = 0$). The electric field of a plane wave polarized in the x-direction and propagating parallel to the z-axis is given by

$$\underline{\mathbf{E}}(\mathbf{r},t) = \underbrace{\mathbf{n}_x\left(E_\mathrm{p}\mathrm{e}^{-\mathrm{j}kz} + E_\mathrm{m}\mathrm{e}^{\mathrm{j}kz}\right)}_{\underline{\mathbf{E}}(\mathbf{r})}\mathrm{e}^{\mathrm{j}\omega t}\,, \tag{72}$$

where $\mathbf{n}_x$ is the normal component in the x-direction, E_p the amplitude of a wave propagating in the positive z-direction, and E_m the amplitude of a wave propagating in the negative z-direction. It is a crucial property of the wave equation that it allows for propagation in both positive and negative directions. The

parameter k is referred to as the wavenumber. Substituting the equation for the plane wave in Eq. (63) and using the fact that the amplitudes E_p and E_m can be freely chosen delivers

$$k = \frac{\omega}{c}\,, \tag{73}$$

with the speed of light $c = 1/\sqrt{\varepsilon\mu}$.

Despite the fact that magnetic fields neither show up in Eq. (63) nor in Eq. (72), they are present, owing to the mutual coupling of electric and magnetic fields for frequencies greater than zero. To determine the magnetic field strength, which accompanies the electric field of a plane wave, Eq. (72) is inserted into Faraday's law of induction (Eq. (7)). Subsequently, the magnetic flux density is replaced by the magnetic field strength using the material relation (Eq. (26)). These two steps yield the magnetic field strength,

$$\underbrace{\underline{\mathbf{H}}(\mathbf{r},t) = \mathbf{n}_y \underbrace{\frac{\mathrm{j}k}{\mathrm{j}\omega\mu}}_{Z} \left(E_\mathrm{p}\mathrm{e}^{-\mathrm{j}kz} - E_\mathrm{m}\mathrm{e}^{\mathrm{j}kz)}\right)}_{\underline{\mathbf{H}}(\mathbf{r})} \mathrm{e}^{\mathrm{j}\omega t}\,, \tag{74}$$

of the plane wave, where $\mathbf{n}_y$ is the normal component in the y direction and $Z = \sqrt{\mu/\varepsilon}$ is the so-called wave impedance. The wave impedance of the plane wave is a quantity depending on the material in which the wave is propagating. This impedance results from the ratio of the amplitude of the electric field and the amplitude of the magnetic field of a propagating wave.

The power per unit area transported by the plane wave is determined by the frequency-domain representation of the Poynting vector (Eq. (19)) and reads as

$$\underline{\mathbf{S}}(\mathbf{r}) = \frac{1}{2}\underline{\mathbf{E}}(\mathbf{r}) \times \underline{\mathbf{H}}^*(\mathbf{r})\,, \tag{75}$$

where $\underline{\mathbf{H}}^*(\mathbf{r})$ denotes the complex conjugate of $\underline{\mathbf{H}}(\mathbf{r})$. According to Eqs. (72), (74), and (75), a wave travelling in the positive z-direction with the amplitude E_p transports the following power per unit area:

$$\underline{\mathbf{S}}(\mathbf{r}) = \mathbf{n}_z \frac{1}{2Z} E_\mathrm{p}^2\,. \tag{76}$$

6 Field attenuation in conductors

To discuss the attenuation of fields in conductors, a plane wave (Eq. (72)) with amplitude E_p is considered in lossy material. For this purpose, the wavenumber k is assumed to be complex-valued, i.e., $\underline{k} = k' - \mathrm{j}k''$. Here, k' and k'' denote the real and the negative imaginary part of $\underline{k}$, respectively. The plane wave (Eq. (72)) with the complex-valued wavenumber must satisfy Eq. (63). However, as the material under consideration is a conductor, ohmic currents can flow, so that Eq. (63) becomes

$$\Delta\mathbf{E}(\mathbf{r},t) - \varepsilon\mu\frac{\partial^2}{\partial t^2}\mathbf{E}(\mathbf{r},t) = \mu\sigma\frac{\partial}{\partial t}\mathbf{E}(\mathbf{r},t)\,, \tag{77}$$

by employment of Eq. (27). Note that the conductor under consideration is assumed to be free of free charges, so that $\rho(\mathbf{r},t) = 0$ holds. Plugging the complex-valued electric field strength of the plane wave into Eq. (77) determines the complex-valued wavenumber:

$$\underline{k}^2 = \varepsilon\mu\omega^2 - \mathrm{j}\omega\mu\sigma\,. \tag{78}$$

Taking the square root of this equation gives

$$\underline{k} = \underbrace{\frac{\mu\sigma\omega}{\sqrt{2}\sqrt{\sqrt{\mu^2\sigma^2\omega^2 + \varepsilon^2\mu^2\omega^4} - \varepsilon\mu\omega^2}}}_{k'} - \mathrm{j}\underbrace{\frac{1}{\sqrt{2}}\sqrt{\sqrt{\mu^2\sigma^2\omega^2 + \varepsilon^2\mu^2\omega^4} - \varepsilon\mu\omega^2}}_{k''}\,. \tag{79}$$

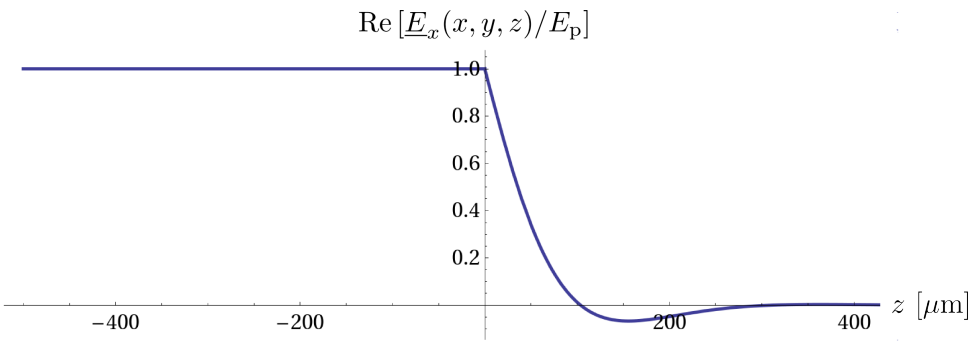


Fig. 10: Amplitude of a plane wave with frequency $f = 1\,\mathrm{MHz}$ polarized in the x-direction and propagating in the z-direction. For $z < 0$, vacuum with the material properties ε_0, μ_0, and $\sigma = 0\,\mathrm{S/m}$ is assumed. For $z > 0$, copper with the material properties ε_0, μ_0, and $\sigma = 58 \times 10^6\,\mathrm{S/m}$ is assumed. In other words, there is an interface between vacuum and copper at $z = 0$. The exponential decay of the field strength is clearly visible in copper, i.e., for $z > 0$. In this scenario, the penetration depth, $\delta = 66.085$ µm, is much smaller than the wavelength $\lambda = 1/(f\sqrt{\varepsilon_0\mu_0}) = 299.863\,\mathrm{m}$ in vacuum, so that the field strength appears to be constant for $z < 0$ on the given scale. Note that the electric field resulting from the reflection of the wave at the copper surface (i.e., a wave propagating in negative z-direction) is not depicted.

(Recall that a quadratic equation always has two solutions. Here, the solution referring to a wave travelling in the positive z-direction is selected.) Substituting the complex-valued wave number into the formula of the electric field of the plane wave (Eq. (72)) and exclusively assuming a wave propagating in the positive z-direction yields

$$\underline{\mathbf{E}}(\mathbf{r}) = \mathbf{n}_x E_\mathrm{p} \mathrm{e}^{-\mathrm{j}\underline{k}z} = \mathbf{n}_x E_\mathrm{p} \mathrm{e}^{-\mathrm{j}k'z} \mathrm{e}^{-k''z}\,. \tag{80}$$

The statement shows that the constant k'' describes the exponential decay of the field strength owing to the conductivity of the material. In this context, the inverse of k'',

$$\delta = \frac{1}{k''} = \frac{\sqrt{2}}{\sqrt{\sqrt{\mu^2\omega^2\left(\sigma^2 + \varepsilon^2\omega^2\right)} - \varepsilon\mu\omega^2}} \approx \sqrt{\frac{2}{\mu\omega\sigma}}\,, \tag{81}$$

is of special relevance. The so-called penetration depth δ is the length that is required for the wave amplitude to decay by a factor of $\mathrm{e}^{-1} \approx 0.3678$ in the direction of propagation. Typical values for the penetration depth are of the order of $10^{-5}\,\mathrm{m}$ for copper at $1\,\mathrm{MHz}$. Figure 10 shows the field strength of a plane wave with a frequency of $1\,\mathrm{MHz}$ at an interface from vacuum ($z < 0$) to copper ($z > 0$). Note that, for the described scenario, the wavelength $\lambda = 1/(f\sqrt{\varepsilon_0\mu_0})$ in free space is much larger than the penetration depth δ, so that the field strength appears to be constant in vacuum in the diagram.

7 Summary and conclusion

This report shows that profound knowledge of mathematical tools, such as vector analysis, is required for the formal description of electromagnetic phenomena. However, once one is able to apply and to work with these mathematical tools, one realizes the beauty of the theory of electromagnetic fields: only four differential equations plus three constitutive equations are required to describe electromagnetic phenomena completely in a classical and macroscopic way. Moreover, it is very satisfactory to see that these equations inherently conserve energy. Recall that conservation of energy can be considered as one of the 'holy grails' of physics.

After providing a short overview with this report, readers are invited to continue their studies in the exciting field of the theory of electromagnetic fields by reading the literature listed in the reference

section. The authors would like to point out that in addition to the analytical techniques for solving electromagnetic field problems, numerical methods for field problems are a very active research area. In contrast with analytical approaches, numerical methods can tackle field problems that are by far much more complex. Unfortunately, there exist a large variety of pitfalls when numerically solving field problems. To avoid these pitfalls, a profound knowledge of numerical mathematics and good command of analytical methods of the theory of electromagnetic fields is mandatory.

Acknowledgements

The authors thank the organizers of the CERN accelerator school for the opportunity to give a presentation and to prepare this report. Moreover, the authors express their appreciation to Franziska Reimann from the University of Rostock for kindly contributing Figs. 4 to 6. In addition, the authors are grateful to Dirk Hecht from the University of Rostock for carefully proofreading the manuscript and providing various proposals for improvements.

References

[1] K. Zhang and D. Li, *Electromagnetic Theory for Microwaves and Optoelectronics*, 2nd ed. (Springer, Berlin, 2008).

[2] C.A. Balanis, *Advanced Engineering Electromagnetics* (John Wiley & Sons, New York, 1989).

[3] R.E. Collin, *Field Theory of Guided Waves*, 2nd ed. (IEEE Press, New York, 1991).

[4] R.E. Collin, *Foundations for Microwave Engineering* (McGraw-Hill, New York,1992).

[5] D.J. Griffiths, *Introduction to Electrodynamics* (Prentice-Hall International, London, 1999).

[6] D. Halliday *et al.*, *Fundamentals of Physics* (John Wiley & Sons, New York, 2014).

[7] J.D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1998).

[8] U. van Rienen, *Numerical Methods in Computational Electrodynamics: Linear Systems in Practical Applications* (Springer, Berlin, 2001), https://doi.org/10.1007/978-3-642-56802-2.

[9] T. Flisgen, Ph.D. thesis, Universität Rostock, 2015, CERN-ACC-2015-0145.

[10] J.C. Maxwell, *A Treatise on Electricity and Magnetism*, vol. 1 (Cambridge University Press, Cambridge, 2010.

[11] A. Logg *et al. Automated Solution of Differential Equations by the Finite Element Method*. (Springer, Berlin, 2012).

[12] CST AG. *CST Studio Suite* (CST, Darmstadt, 2017).

[13] J. Sekutowicz *et al.* A design of a 3rd harmonic cavity for the TTF 2 photoinjector, TESLA-FEL 2002-05, 2002.

[14] T. Khabibouline *et al.* Higher order modes of a 3rd harmonic cavity with an increased end-cup iris, TESLA-FEL 2003-01, 2003.

Space Charge Mitigation[1]

M. Ferrario
Frascati National Laboratory, National Institute for Nuclear Physics, Rome, Italy

Abstract
In this paper we introduce, from basic principles, the main concepts of beam focusing and transport of space charge dominated beams in high brightness accelerators using the beam envelope equation as a convenient mathematical tool. Matching conditions suitable for preserving beam quality are derived from the model for significant beam dynamics regimes.

Keywords
Beam matching; rms emittance; laminar beam; space charge effects; rms envelope equations.

1 Introduction

Light sources based on high-gain free electron lasers or future high-energy linear colliders require the production, acceleration, and transport up to the interaction point of low divergence, high charge-density electron bunches [1]. Many effects contribute in general to degradation of the final beam quality, including chromatic effects, wake fields, emission of coherent radiation, and accelerator misalignments. Space charge effects and mismatch with focusing and accelerating devices typically contribute to emittance degradation of high charge-density beams [2]; hence, control of beam transport and acceleration is the leading edge for high-quality beam production.

Space charge effects represent a very critical issue and a fundamental challenge for high-quality beam production and its applications. Without proper matching, significant emittance growth may occur when the beam is propagating through different stages and components owing to the large differences of transverse focusing strength. This unwanted effect is even more serious in the presence of finite energy spread.

In this paper we introduce, from basic principles, the main concepts of beam focusing and transport in modern accelerators using the beam envelope equation as a convenient mathematical tool. Matching conditions suitable for preserving beam quality are derived from the model for significant beam dynamics regimes. A more detailed discussion of the previous topics can be found in the many classical textbooks on this subject, as listed in Refs. [3–6].

2 Laminar and non-laminar beams

An ideal high-charge particle beam has orbits that flow in layers that never intersect, as occurs in a laminar fluid. Such a beam is often called a laminar beam. More precisely, a laminar beam satisfies the following two conditions [6]:

i) all particles at a given position have identical transverse velocities. On the contrary, the orbits of two particles that start at the same position could separate and later cross each other;

[1] This paper has already been published in the Proceedings of the CAS-CERN Accelerator School on Free Electron Lasers and Energy Recovery Linacs, https://doi.org/10.23730/CYRSP-2018-001.177.

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-005.89

ii) assuming that the beam propagates along the z axis, the magnitudes of the slopes of the trajectories in the transverse directions x and y, given by $x'(z) = \mathrm{d}x/\mathrm{d}z$ and $y'(z) = \mathrm{d}y/\mathrm{d}z$, are linearly proportional to the displacement from the z axis of beam propagation.

Trajectories of interest in beam physics are always confined to the inside of small, near-axis regions, and the transverse momentum is much smaller than the longitudinal momentum, $p_{x,y} << p_z \approx p$. As a consequence, it is possible in most cases to use the small angle, or *paraxial*, approximation, which allows us to write the useful approximate expressions $x' = p_x/p_z \approx p_x/p$ and $y' = p_y/p_z \approx p_x/p$.

To help understand the features and advantages of a laminar beam propagation, the following figures compare the typical behaviour of a laminar and a non-laminar (or thermal) beam.

Figure 1 illustrates an example of orbit evolution of a laminar mono-energetic beam with half width x_0 along a simple beam line with an ideal focusing element (solenoid, magnetic quadrupoles, or electrostatic transverse fields are usually adopted to this end), represented by a thin lens located at the longitudinal coordinate $z = 0$. In an ideal lens, focusing (defocusing) forces are linearly proportional to the displacement from the symmetry axis z, so that the lens maintains the laminar flow of the beam.

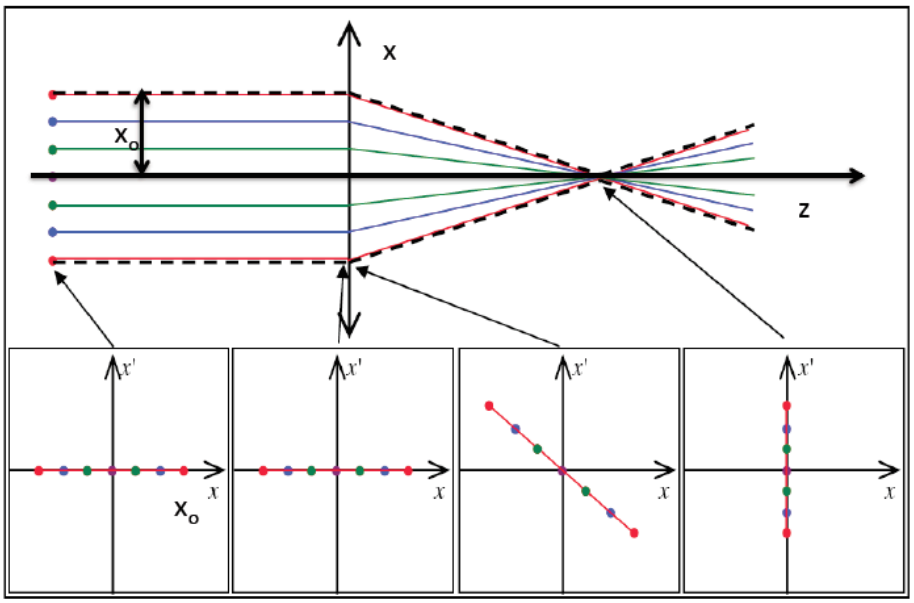


Fig. 1: Particle trajectories and phase space evolution of a laminar beam [7]

The beam shown in Fig. 1 starts propagating completely parallel to the symmetry axis z; in this particular case, the particles all have zero transverse velocity. There are no orbits that cross each other in such a beam. Ignoring collisions and inner forces, such as Coulomb forces, a parallel beam could propagate an infinite distance with no change in its transverse width. When the beam crosses the ideal lens, it is transformed into a converging laminar beam. Because the transverse velocities after the linear lens are proportional to the displacement off axis, particle orbits define similar triangles that converge to a single point. After passing through the singularity at the focal point, the particles follow diverging orbits. We can always transform a diverging (or converging) beam into a parallel beam by using a lens of the proper focal length, as can be seen by reversing the propagation axis of Fig. 1.

The small boxes in the lower part of the figure depict the particle distributions in the trace space (x, x'), equivalent to the canonical phase space $(x, p_x \approx x'p)$ when p is constant, i.e., without beam acceleration. The phase space area occupied by an ideal laminar beam is a straight segment of zero thickness. As can be easily verified, the condition that the particle distribution has zero thickness proceeds from condition 1; the segment straightness is a consequence of condition 2. The distribution

of a laminar beam propagating through a transport system with ideal linear focusing elements is thus a straight segment with variable slope.

Particles in a non-laminar beam have a random distribution of transverse velocities at the same location and a spread in directions, as shown in Fig. 2. Because of the disorder of a non-laminar beam, it is impossible to focus all particles from a location in the beam toward a common point. Lenses can influence only the average motion of particles. Focal spot limitations are a major concern for a wide variety of applications, from electron microscopy to free electron lasers and linear colliders. The phase space plot of a non-laminar beam is no longer a straight line: the beam, as shown in the lower boxes of Fig. 2, occupies a wider area of the phase space.

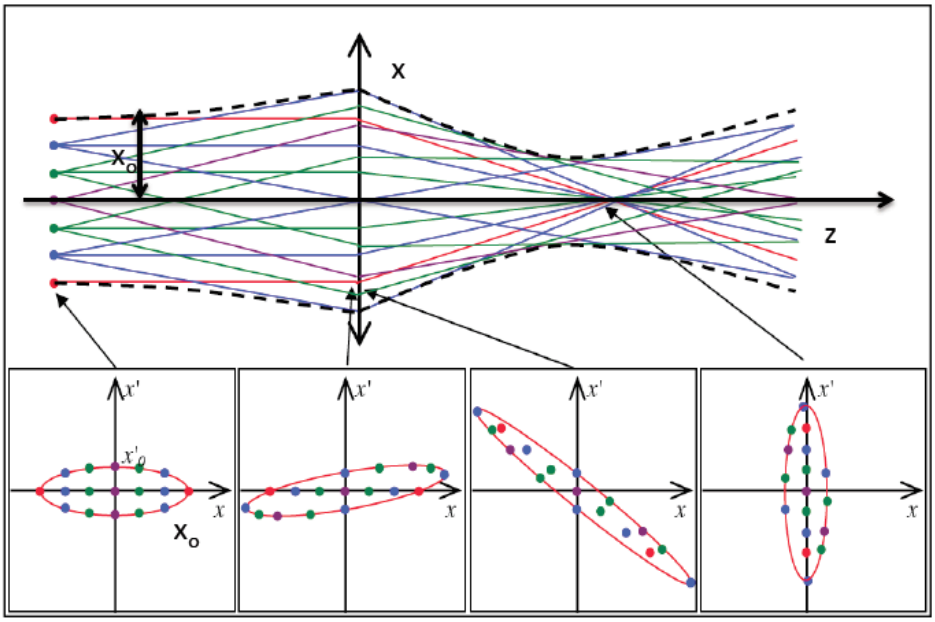


Fig. 2: Particle trajectories and phase space evolution of a non-laminar beam [7]

3 The emittance concept

The phase space surface A occupied by a beam is a convenient figure of merit for designating the quality of a beam. This quantity is the emittance ε_x and is usually represented by an ellipse that contains the whole particle distribution in the phase space (x, x'), such that $A = \pi\varepsilon_x$. An analogous definition holds for the (y, y') and (z, z') planes. The original choice of an elliptical shape comes from the fact that when linear focusing forces are applied to a beam, the trajectory of each particle in phase space lies on an ellipse, which may be called the trajectory ellipse. Being the area of the phase space, the emittance is measured in metres radians. More often is expressed in millimetres milliradians or, equivalently, in micrometres.

The ellipse equation is written as

$$\gamma_x x^2 + 2\alpha_x x x' + \beta_x x'^2 = \varepsilon_x \ , \tag{1}$$

where x and x' are the particle coordinates in the phase space and the coefficients $\alpha_x(z)$, $\beta_x(z)$, and $\gamma_x(z)$ are called Twiss parameters, which are related by the geometrical condition:

$$\beta_x \gamma_x - \alpha_x^2 = 1 \ . \tag{2}$$

As shown in Fig. 3, the beam envelope boundary X_{max}, its derivative $(X_{max})'$, and the maximum beam divergence X'_{max}, i.e., the projection on the axes x and x' of the ellipse edges, can be expressed as a function of the ellipse parameters:

$$\begin{cases} X_{\max} = \sqrt{\beta_x \varepsilon_x} \\ \left(X_{\max}\right)' = -\alpha\sqrt{\dfrac{\varepsilon}{\beta}} \\ X'_{\max} = \sqrt{\gamma_x \varepsilon_x} \end{cases} . \tag{3}$$

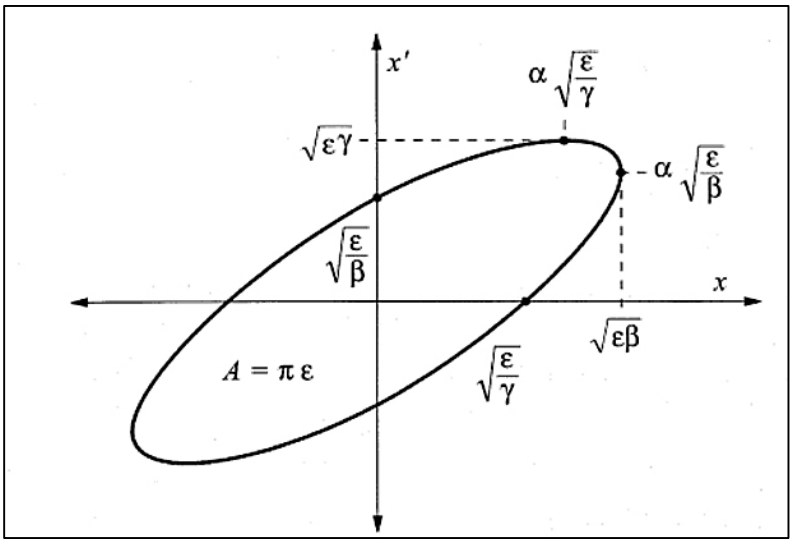


Fig. 3: Phase space distribution in a skewed elliptical boundary, showing the relationship of Twiss parameters to the ellipse geometry [6].

According to Liouville's theorem, the six-dimensional (x, p_x, y, p_y, z, p_z) phase space volume occupied by a beam is constant, provided that there are no dissipative forces, no particles lost or created, and no Coulomb scattering among particles. Moreover, if the forces in the three orthogonal directions are uncoupled, Liouville's theorem also holds for each reduced phase space surface, (x, p_x), (y, p_y), (z, p_z), and hence emittance also remains constant in each plane [3].

Although the net phase space surface occupied by a beam is constant, non-linear field components can stretch and distort the particle distribution in the phase space, and the beam will lose its laminar behaviour. A realistic phase space distribution is often very different from a regular ellipse, as shown in Fig. 4.

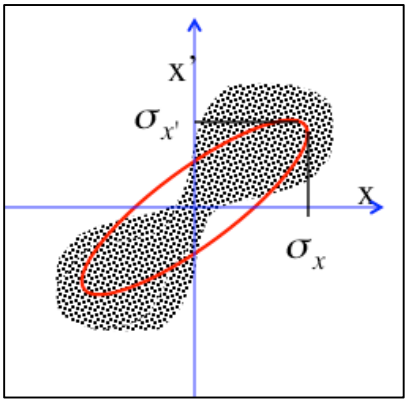


Fig. 4: Typical evolution of phase space distribution (black dots) under the effects of non-linear forces with the equivalent ellipse superimposed (red line).

We introduce, therefore, a definition of emittance that measures the beam quality rather than the phase space area. It is often more convenient to associate a statistical definition of emittance with a generic distribution function $f(x, x', z)$ in the phase space; this is the so-called *root mean square (rms) emittance*:

$$\gamma_x x^2 + 2\alpha_x xx' + \beta_x x'^2 = \varepsilon_{x,\mathrm{rms}} \ . \tag{4}$$

The rms emittance is defined such that the equivalent-ellipse projections on the x and x' axes are equal to the rms values of the distribution, implying the following conditions:

$$\begin{cases} \sigma_x = \sqrt{\beta_x \varepsilon_{x,\mathrm{rms}}} \\ \sigma_{x'} = \sqrt{\gamma_x \varepsilon_{x,\mathrm{rms}}} \end{cases}, \tag{5}$$

where

$$\begin{cases} \sigma_x^2(z) = \langle x^2 \rangle = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^2 f(x, x', z)\,\mathrm{d}x\mathrm{d}x' \\ \sigma_{x'}^2(z) = \langle x'^2 \rangle = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x'^2 f(x, x', z)\,\mathrm{d}x\mathrm{d}x' \end{cases} \tag{6}$$

are the second moments of the distribution function $f(x, x', z)$. Another important quantity that accounts for the degree of (x, x') correlations is defined as

$$\sigma_{xx'}(z) = \langle xx' \rangle = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xx' f(x, x', z)\,\mathrm{d}x\mathrm{d}x' \ . \tag{7}$$

From Eq. (3) it also holds that

$$\sigma'_x = \frac{\sigma_{xx'}}{\sigma_x} = -\alpha_x \sqrt{\frac{\varepsilon_{x,\mathrm{rms}}}{\beta_x}}$$

see also Eq. (16), which allows us to link the correlation moment, Eq. (7), to the Twiss parameter as

$$\sigma_{xx'} = -\alpha_x \varepsilon_{x,\mathrm{rms}} \ . \tag{8}$$

One can easily see from Eqs. (3) and (5) that

$$\alpha_x = -\frac{1}{2}\frac{\mathrm{d}\beta_x}{\mathrm{d}z}$$

also holds.

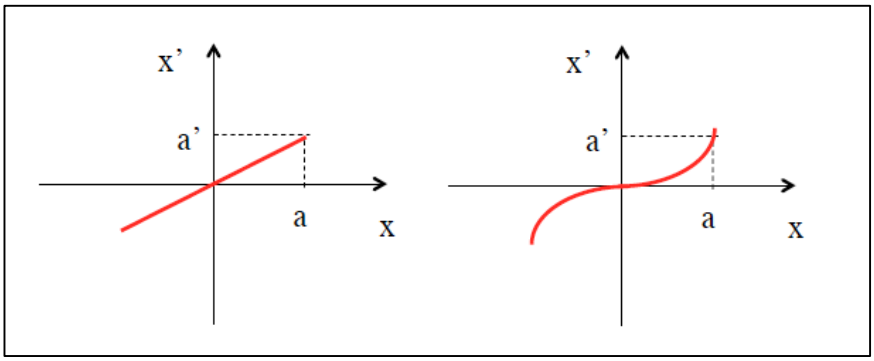
By substituting the Twiss parameter defined by Eqs. (5) and (8) into condition 2 we obtain [5]

$$\frac{\sigma_{x'}^2}{\varepsilon_{x,\mathrm{rms}}}\frac{\sigma_x^2}{\varepsilon_{x,\mathrm{rms}}} - \left(\frac{\sigma_{xx'}}{\varepsilon_{x,\mathrm{rms}}}\right) = 1 \quad . \tag{9}$$

Reordering the terms in Eq. (8) we obtain the definition of *rms emittance* in terms of the second moments of the distribution:

$$\varepsilon_{\mathrm{rms}} = \sqrt{\sigma_x^2\sigma_{x'}^2 - \sigma_{xx'}^2} = \sqrt{\left(\langle x^2\rangle\langle x'^2\rangle - \langle xx'\rangle^2\right)} \ , \tag{10}$$

where we omit, from now on, the subscript x in the emittance notation: $\varepsilon_{\mathrm{rms}} = \varepsilon_{x,\mathrm{rms}}$. The rms emittance tells us some important information about phase space distributions under the effect of linear or non-linear forces acting on the beam. Consider, for example, an idealized particle distribution in phase space that lies on some line that passes through the origin, as illustrated in Fig. 5.



(a) (b)

Fig. 5: Phase space distributions under the effect of (a) linear or (b) non-linear forces acting on the beam

Assuming a generic correlation of the type $x' = Cx^n$ and computing the rms emittance according to Eq. (10) we have

$$\varepsilon_{\mathrm{rms}}^2 = C\sqrt{\langle x^2\rangle\langle x^{2n}\rangle - \langle x^{n+1}\rangle^2} \quad \begin{cases} n = 1 \Rightarrow \ \varepsilon_{\mathrm{rms}} = 0 \\ n > 1 \Rightarrow \ \varepsilon_{\mathrm{rms}} \neq 0 \end{cases} . \tag{11}$$

When $n = 1$, the line is straight and the rms emittance is $\varepsilon_{\mathrm{rms}} = 0$. When $n > 1$ the relationship is non-linear, the line in phase space is curved, and the rms emittance is, in general, not zero. Both distributions have zero area. Therefore, we conclude that even when the phase space area is zero, if the distribution is lying on a curved line, its rms emittance is not zero. The rms emittance depends not only on the area occupied by the beam in phase space, but also on distortions produced by non-linear forces.

If the beam is subject to acceleration, it is more convenient to use the rms normalized emittance, for which the transverse momentum $p_x = p_z x' = m_0 c\beta\gamma x'$ is used instead of the divergence:

$$\varepsilon_{n,\mathrm{rms}} = \frac{1}{m_0 c}\sqrt{\sigma_x^2\sigma_{p_x}^2 - \sigma_{xp_x}^{\ 2}} = \frac{1}{m_0 c}\sqrt{\left(\langle x^2\rangle\langle p_x^2\rangle - \langle xp_x\rangle^2\right)} = \sqrt{\left(\langle x^2\rangle\langle(\beta\gamma x')^2\rangle - \langle x\beta\gamma x'\rangle^2\right)} \ . \tag{12}$$

The reason for introducing a normalized emittance is that the divergences of the particles $x' = p_x/p$ are reduced during acceleration as p increases. Thus, acceleration reduces the un-normalized emittance, but does not affect the normalized emittance.

It is interesting to estimate the fundamental limit of the beam emittance that is set by quantum mechanics on the knowledge of the two conjugate variables (x, p_x). The state of a particle is actually not exactly represented by a point, but by a small uncertainty volume of the order of $\hbar^3$ in the 6D phase space. According to the Heisenberg uncertainty relation $\sigma_x \sigma_{p_x} \geq \frac{\hbar}{2}$ one gets from Eq. (12) $\varepsilon_{n,rms}^{QM} \geq \frac{1}{2}\frac{\hbar}{m_o c} = \frac{\lambda\!\!\!^{-}_c}{2}$, where $\lambda\!\!\!^{-}_c$ is the reduced Compton wavelength. For electrons it gives: $\varepsilon_{n,rms}^{QM} \geq 1.9\times10^{-13} m$.

In the classical limit we see also from Eq. (12) that the single particle emittance is zero.

Assuming a small energy spread within the beam, the normalized and un-normalized emittances can be related by the approximated relation $\langle\beta\gamma\rangle\varepsilon_{rms}$. This approximation, which is often used in conventional accelerators, may be strongly misleading when adopted for describing beams with significant energy spread, like those currently produced by plasma accelerators. A more careful analysis is reported next [8].

When the correlations between the energy and transverse positions are negligible (as in a drift without collective effects), Eq. (12) can be written as

$$\varepsilon_{n,\mathrm{rms}}^2 = \left\langle \beta^2\gamma^2 \right\rangle \left\langle x^2 \right\rangle \left\langle x'^2 \right\rangle - \left\langle \beta\gamma \right\rangle^2 \left\langle xx' \right\rangle^2 . \tag{13}$$

Consider now the definition of relative energy spread

$$\sigma_\gamma^2 = \frac{\left\langle \beta^2\gamma^2 \right\rangle - \left\langle \beta\gamma \right\rangle^2}{\left\langle \beta\gamma \right\rangle^2}$$

which can be inserted into Eq. (13) to give

$$\varepsilon_{n,\mathrm{rms}}^2 = \left\langle \beta^2\gamma^2 \right\rangle \sigma_\gamma^2 \left\langle x^2 \right\rangle \left\langle x'^2 \right\rangle + \left\langle \beta\gamma \right\rangle^2 \left(\left\langle x^2 \right\rangle \left\langle x'^2 \right\rangle - \left\langle xx' \right\rangle^2 \right) . \tag{14}$$

Assuming relativistic particles ($\beta = 1$), we get

$$\varepsilon_{n,\mathrm{rms}}^2 = \left\langle \gamma^2 \right\rangle \left(\sigma_\gamma^2 \sigma_x^2 \sigma_{x'}^2 + \varepsilon_{\mathrm{rms}}^2 \right) . \tag{15}$$

If the first term in the parentheses is negligible, we find the conventional approximation of the normalized emittance as $\left\langle \gamma \right\rangle \varepsilon_{\mathrm{rms}}$. For a conventional accelerator, this might generally be the case. Considering, for example, beam parameters for the SPARC_LAB photoinjector [9]: at 5 MeV the ratio between the first and the second term is ~10^{-3}; while at 150 MeV it is ~10^{-5}. Conversely, using typical beam parameters at the plasma–vacuum interface, the first term is of the same order of magnitude as for conventional accelerators at low energies; however, owing to the rapid increase of the bunch size outside the plasma ($\sigma_{x'}$ ~ mrad) and the large energy spread (σ_γ > 1%), it becomes predominant compared with the second term after a drift of a few millimetres. *Therefore, the use of approximated formulas when measuring the normalized emittance of plasma accelerated particle beams is inappropriate* [10].

4 The root mean square envelope equation

We are now interested in following the evolution of the particle distribution during beam transport and acceleration. One can use the collective variable defined in Eq. (6), the second moment of the distribution termed the rms beam envelope, to derive a differential equation suitable for describing the rms beam envelope dynamics [11]. To this end, let us compute the first and second derivative of σ_x [4]:

$$\begin{aligned}\frac{\mathrm{d}\sigma_x}{\mathrm{d}z} &= \frac{\mathrm{d}}{\mathrm{d}z}\sqrt{\langle x^2\rangle} = \frac{1}{2\sigma_x}\frac{\mathrm{d}}{\mathrm{d}z}\langle x^2\rangle = \frac{1}{2\sigma_x}2\langle xx'\rangle = \frac{\sigma_{xx'}}{\sigma_x} \\ \frac{\mathrm{d}^2\sigma_x}{\mathrm{d}z^2} &= \frac{\mathrm{d}}{\mathrm{d}z}\frac{\sigma_{xx'}}{\sigma_x} = \frac{1}{\sigma_x}\frac{\mathrm{d}\sigma_{xx'}}{\mathrm{d}z} - \frac{\sigma_{xx'}^2}{\sigma_x^3} = \frac{1}{\sigma_x}\left(\langle x'^2\rangle + \langle xx'\rangle\right) - \frac{\sigma_{xx'}^2}{\sigma_x^3} = \frac{\sigma_{x'}^2 + \langle xx''\rangle}{\sigma_x} - \frac{\sigma_{xx'}^2}{\sigma_x^3}\end{aligned} \quad . \tag{16}$$

Rearranging the second derivative in Eq. (16), we obtain a second-order non-linear differential equation for the beam envelope evolution,

$$\sigma_x'' = \frac{\sigma_x^2\sigma_{x'}^2 - \sigma_{xx'}^2}{\sigma_x^3} + \frac{\langle xx''\rangle}{\sigma_x} \quad , \tag{17}$$

or, in a more convenient form, using the rms emittance definition Eq. (10),

$$\sigma_x'' - \frac{1}{\sigma_x}\langle xx''\rangle = \frac{\varepsilon_{\mathrm{rms}}^2}{\sigma_x^3} \quad . \tag{18}$$

In Eq. (18), the emittance term can be interpreted physically as an outward pressure on the beam envelope produced by the rms spread in trajectory angle, which is parameterized by the rms emittance.

Let us now consider, for example, the simple case with $\langle xx''\rangle = 0$, describing a beam drifting in free space. The envelope equation reduces to

$$\sigma_x^3\sigma_x'' = \varepsilon_{\mathrm{rms}}^2 \quad . \tag{19}$$

With initial conditions σ_0, σ'_0 at z_0, depending on the upstream transport channel, Eq. (19) has a hyperbolic solution:

$$\sigma(z) = \sqrt{\left(\sigma_0 + \sigma_0'(z - z_0)\right)^2 + \frac{\varepsilon_{\mathrm{rms}}^2}{\sigma_0^2}(z - z_0)^2} \quad . \tag{20}$$

Considering the case of a beam at waist ($\langle xx'\rangle = 0$) with $\sigma'_0 = 0$, using Eq. (5), the solution Eq. (20) is often written in terms of the β function as

$$\sigma(z) = \sigma_0\sqrt{1 + \left(\frac{z - z_0}{\beta_{\mathrm{w}}}\right)^2} \quad . \tag{21}$$

This relation indicates that without any external focusing element the beam envelope increases from the beam waist by a factor √2 with a characteristic length $\beta_w = \sigma_0^2/\varepsilon_{\mathrm{rms}}$, as shown in Fig. 6.

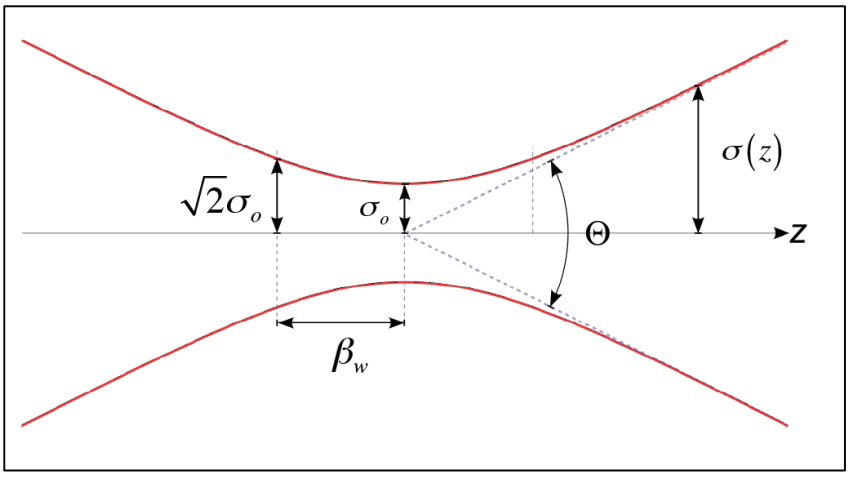


Fig. 6: Schematic representation of the beam envelope behaviour near the beam waist

At the waist, the relation $\varepsilon^2_{\text{rms}} = \sigma^2_{0,x}\sigma^2_{0,x'}$ also holds, which can be inserted into Eq. (20) to give $\sigma^2_x(z) = \sigma^2_{0x'}(z - z_0)^2$. Under this condition, Eq. (15) can be written as

$$\varepsilon^2_{n,\text{rms}}(z) = \langle\gamma^2\rangle\left(\sigma^2_\gamma\sigma^4_{x'}(z - z_0)^2 + \varepsilon^2_{\text{rms}}\right),$$

showing that beams with large energy spread and divergence undergo a significant normalized emittance growth even in a drift of length $(z - z_0)$ [8, 12].

Notice also that the solution Eq. (21) is exactly analogous to that of a Gaussian light beam for which the beam width $w = 2\sigma_{\text{ph}}$ increases away from its minimum value at the waist w_0 with characteristic length $Z_R = \pi w_0^2/\lambda$ (Rayleigh length) [4]. This analogy suggests that we can identify an effective emittance of a photon beam as $\varepsilon_{\text{ph}} = \lambda/4\pi$.

For the effective transport of a beam with finite emittance, it is mandatory to make use of some external force providing beam confinement in the transport or accelerating line. The term $\langle xx''\rangle$ accounts for external forces when we know x'', given by the single particle equation of motion:

$$\frac{\mathrm{d}p_x}{\mathrm{d}t} = F_x \,. \tag{22}$$

Under the paraxial approximation $p_x << p = \beta\gamma mc$, the transverse momentum p_x can be written as $p_x = px' = \beta\gamma m_0 cx'$, so that

$$\frac{\mathrm{d}p_x}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(px') = \beta c\frac{\mathrm{d}}{\mathrm{d}z}(px') = F_x \,, \tag{23}$$

and the transverse acceleration results in

$$x'' = -\frac{p'}{p}x' + \frac{F_x}{\beta cp} \,. \tag{24}$$

It follows that

$$\langle xx''\rangle = -\frac{p'}{p}\langle xx'\rangle + \frac{\langle xF_x\rangle}{\beta cp} = \frac{p'}{p}\sigma_{xx'} + \frac{\langle xF_x\rangle}{\beta cp} \,. \tag{25}$$

Inserting Eq. (25) into Eq. (18) and recalling Eq. (16), $\sigma_x' = \sigma_{xx'}/\sigma_x$, the complete rms envelope equation is:

$$\sigma_x'' + \frac{p'}{p}\sigma_x' - \frac{1}{\sigma_x}\frac{\langle xF_x \rangle}{\beta cp} = \frac{\varepsilon_{n,\mathrm{rms}}^2}{\gamma^2\sigma_x^3} \,, \tag{26}$$

where we have included the normalized emittance $\varepsilon_{n,\mathrm{rms}} = \gamma\varepsilon_{\mathrm{rms}}$. Notice that the effect of longitudinal accelerations appears in the rms envelope equation as an oscillation damping term, called 'adiabatic damping', proportional to p'/p. The term $\langle xF_x \rangle$ represents the moment of any external transverse force acting on the beam, such as that produced by a focusing magnetic channel.

5 External forces

Let's now consider the case of an external linear force acting on the beam in the form $F_x = \mp kx$. It can be focusing or defocusing, according to the sign. The moment of the force is

$$\langle xF_x \rangle = \mp k\langle x^2 \rangle = \mp k\sigma_x^2 \tag{27}$$

and the envelope equation becomes

$$\sigma_x'' + \frac{\gamma'}{\gamma}\sigma_x' \mp k_{\mathrm{ext}}^2\sigma_x = \frac{\varepsilon_{n,\mathrm{rms}}^2}{\gamma^2\sigma_x^3} \,, \tag{28}$$

where we have explicitly used the momentum definition $p = \gamma mc$ for a relativistic particle with $\beta \approx 1$ and defined the wavenumber

$$k_{\mathrm{ext}}^2 = \frac{k}{\gamma m_0 c^2} \,.$$

Typical focusing elements are quadrupoles and solenoids [3]. The magnetic quadrupole field is given in Cartesian coordinates by

$$\begin{cases} B_x = B_0\dfrac{y}{d} = B_0' y \\ B_y = B_0\dfrac{x}{d} = B_0' x \end{cases} , \tag{29}$$

where d is the pole distance and B_0' is the field gradient. The force acting on the beam is $\vec{F}_\perp = qv_z B_0'\left(y\hat{j} - x\hat{i}\right)$ and, when B_0 is positive, is focusing in the x direction and defocusing in the y direction. The focusing strength is

$$k_{\mathrm{quad}} = \frac{qB_0'}{\gamma m_0 c} = k_{\mathrm{ext}}^2 \,.$$

In a solenoid the focusing strength is given by

$$k_{\mathrm{sol}} = \left(\frac{qB_0}{2\gamma m_0 c}\right)^2 = k_{\mathrm{ext}}^2 \,.$$

Notice that the solenoid is always focusing in both directions, an important property when the cylindrical symmetry of the beam must be preserved. However, being a second-order quantity in γ, it is more effective at low energy.

It is interesting to consider the case of a uniform focusing channel without acceleration described by the rms envelope equation

$$\sigma_x'' + k_{\text{ext}}^2 \sigma_x = \frac{\varepsilon_{\text{rms}}^2}{\sigma_x^3} \quad . \tag{30}$$

By substituting $\sigma_x = \sqrt{\beta_x \varepsilon_{\text{rms}}}$ into Eq. (30) one obtains an equation for the 'betatron function' $\beta_x(z)$ that is independent of the emittance term:

$$\beta_x'' + 2k_{\text{ext}}^2 \beta_x = \frac{2}{\beta_x} + \frac{\beta_x'^2}{2\beta_x} \quad . \tag{31}$$

Equation (31) contains just the transport channel focusing strength and, being independent of the beam parameters, suggests that the meaning of the betatron function is to account for the transport line characteristic. The betatron function reflects exterior forces from focusing magnets and is highly dependent on the particular arrangement of the quadrupole magnets. The equilibrium, or matched, solution of Eq. (31) is given by $\beta_{\text{eq}} = \frac{1}{k_{\text{ext}}} = \frac{\lambda_\beta}{2\pi}$, as can be easily verified. This result shows that the matched β_x function is simply the inverse of the focusing wavenumber or, equivalently, is proportional to the 'betatron wavelength' λ_β. The corresponding envelope equilibrium condition, i.e., a stationary solution of Eq. (30), is given by: $\sigma_{\text{eq},x} = \sqrt{\frac{\varepsilon_{\text{rms}}}{k_{\text{ex}}}}$.

In analogy with the kinetic theory of gases we can define the beam temperature in a transverse direction at equilibrium and without correlations as

$$k_{\text{B}} T_{\text{beam},x} = \gamma m_0 \left\langle v_x^2 \right\rangle = \frac{\sigma_{p_x}^2}{\gamma m_0} = m_0 c^2 \frac{\varepsilon_{n,\text{rms}}^2}{\gamma \sigma_{\text{eq},x}^2} = \gamma m_0 \beta^2 c^2 \frac{\varepsilon_{\text{rms}}}{\beta_{\text{eq},x}},$$

where k_{B} is the Boltzmann constant and we have used Eq. (12), showing that the conditions for a cold beam are typically: low emittance, low energy, high betatron function.

By means of the beam temperature concept one can also define the beam emittance at the source called the thermal emittance. Assuming that electrons are in equilibrium with the cathode temperature $T_{\text{c}} = T_{\text{beam}}$ and γ=1, the thermal emittance is given by $\varepsilon_{\text{th,rms}}^{\text{cat}} = \sigma_x \sqrt{\frac{k_{\text{B}} T_{\text{c}}}{m_0 c^2}}$ which, per unit rms spot size at the cathode, is $\varepsilon_{\text{th,rms}} = 0.3$ μm/mm at $T_{\text{c}} = 2500$ K. For comparison, in a photocathode illuminated by a laser pulse with photon energy $\hbar\omega$ the expression for the variance of the transverse momentum of the emitted electrons is given by $\sigma_{p_x} = \sqrt{\frac{m_0}{3}\left(\hbar\omega - \phi_{\text{eff}}\right)}$, where $\phi_{\text{eff}} = \phi_{\text{w}} - \phi_{\text{Schottky}}$, ϕ_{w} being the material work function and ϕ_{Schottky} the Schottky work function [19]. The corresponding thermal emittance is $\varepsilon_{\text{th,rms}}^{\text{ph}} = \sigma_x \sqrt{\frac{\hbar\omega - \phi_{\text{eff}}}{3 m_0 c^2}}$ that, with the typical parameters of a Copper photocathode illuminated by a UV laser, gives a thermal emittance per unit spot size of about 0.5 μm/mm.

6 Space charge forces

Another important force acting on the beam is the one produced by the beam itself due to the internal Coulomb forces. The net effect of the Coulomb interaction in a multiparticle system can be classified into two regimes [3]:

i) *collisional regime*, dominated by binary collisions caused by close particle encounters;

ii) *collective regime* or *space charge regime*, dominated by the self-field produced by the particles' distribution, which varies appreciably only over large distances compared with the average separation of the particles.

A measure for the relative importance of collisional versus collective effects in a beam with particle density n is the relativistic *Debye length*,

$$\lambda_{\mathrm{D}} = \sqrt{\frac{\varepsilon_0 \gamma^2 k_{\mathrm{B}} T_{\mathrm{b}}}{e^2 n}} \quad . \tag{32}$$

As long as the Debye length remains small compared with the particle bunch transverse size, the beam is in the space charge dominated regime and is not sensitive to binary collisions. Smooth functions for the charge and field distributions can be used in this case, and the space charge force can be treated as an external applied force. The space charge field can be separated into linear and non-linear terms as a function of displacement from the beam axis. The linear space charge term defocuses the beam and leads to an increase in beam size. The non-linear space charge terms also increase the rms emittance by distorting the phase space distribution. Under the paraxial approximation of particle motion, we can consider the linear component alone. We shall see next that the linear component of the space charge field can also induce emittance growth when correlations along the bunch are taken into account.

For a bunched beam of uniform charge distribution in a cylinder of radius R and length L, carrying a current $\hat{I}$ and moving with longitudinal velocity $v_z = \beta c$, the linear component of the longitudinal and transverse space charge field are given approximately by [13]

$$E_z(\xi) = \frac{\hat{I} L}{2\pi\varepsilon_0 R^2 \beta c} h(\xi) \quad , \tag{33}$$

$$E_{\mathrm{r}}(r,\xi) = \frac{\hat{I} r}{2\pi\varepsilon_0 R^2 \beta c} g(\xi) \quad . \tag{34}$$

The field form factor is described by the functions:

$$h(\xi) = \sqrt{A^2 + (1-\xi)^2} - \sqrt{A^2 + \xi^2} - |1-\xi| + |\xi| \tag{35}$$

$$g(\xi) = \frac{(1-\xi)}{2\sqrt{A^2 + (1-\xi)^2}} + \frac{\xi}{2\sqrt{A^2 + \xi^2}} \quad , \tag{36}$$

where $\zeta = z/L$ is the normalized longitudinal coordinate along the bunch, $\zeta = 0$ being the bunch tail, and $A = R/\gamma L$ is the beam aspect ratio. The field form factors account for the variation of the fields along the bunch and outside the bunch for $\zeta<0$ *and* $\zeta>L$. As γ increases, $g(\xi) \to 1$ and $h(\xi) \to 0$, thus showing that space charge fields mainly affect transverse beam dynamics. It shows also that an energy increase corresponds to a bunch lengthening in the moving frame $L' = \gamma L$, leading to a vanishing longitudinal field component, as in the case of a continuous beam in the laboratory frame.

To evaluate the force acting on the beam, one must also account for the azimuthal magnetic field associated with the beam current, which, in cylindrical symmetry, is given by

$$B_\vartheta = \frac{\beta}{c} E_r \ .$$

Thus, the Lorentz force acting on each single particle is given by

$$F_r = e\left(E_r - \beta c B_\vartheta\right) = e\left(1-\beta^2\right)E_r = \frac{eE_r}{\gamma^2} \ . \tag{37}$$

The attractive magnetic force, which becomes significant at high velocities, tends to compensate for the repulsive electric force. Therefore, space charge defocusing is primarily a non-relativistic effect and decreases as γ^{-2}.

To include space charge forces in the envelope equation, let us start by writing the space charge forces produced by the previous fields in Cartesian coordinates:

$$F_x = \frac{e\hat{I}x}{8\pi\gamma^2\varepsilon_0\sigma_x^2\beta c} g\left(\xi\right) \ . \tag{38}$$

Then, computing the moment of the force, we need

$$x'' = \frac{F_x}{\beta cp} = \frac{eIx}{8\pi\varepsilon_0\gamma^3 m_0\beta^3 c^3\sigma_x^2} = \frac{k_{\mathrm{sc}}\left(\xi\right)}{\left(\beta\gamma\right)^3\sigma_x^2} \tag{39}$$

where we have introduced the generalized beam perveance,

$$k_{\mathrm{sc}}\left(\xi\right) = \frac{\hat{I}}{2I_{\mathrm{A}}} g\left(\xi\right) \tag{40}$$

where $I_{\mathrm{A}} = 4\pi\varepsilon_0 m_0 c^3/e = 17$ kA is the Alfvén current for electrons. Notice that in this case the perveance in Eq. (40) explicitly depends on the slice coordinate ζ. We can now calculate the term that enters the envelope equation for a relativistic beam,

$$\left\langle xx'' \right\rangle = \frac{k_{\mathrm{sc}}}{\gamma^3\sigma_x^2}\left\langle x^2 \right\rangle = \frac{k_{\mathrm{sc}}}{\gamma^3} \ , \tag{41}$$

leading to the complete envelope equation

$$\sigma_x'' + \frac{\gamma'}{\gamma}\sigma_x' + k_{\mathrm{ext}}^2\sigma_x = \frac{\varepsilon_{n,\mathrm{rms}}^2}{\gamma^2\sigma_x^3} + \frac{k_{\mathrm{sc}}}{\gamma^3\sigma_x} \ . \tag{42}$$

From the envelope equation Eq. (42), we can identify two regimes of beam propagation: *space charge dominated* and *emittance dominated*. A beam is space charge dominated as long as the space charge collective forces are largely dominant over the emittance pressure. In this regime, the linear component of the space charge force produces a quasi-laminar propagation of the beam, as one can see by integrating one time Eq. (39) under the paraxial ray approximation $x' \ll 1$.

A measure of the relative importance of space charge effects versus emittance pressure is given by the *laminarity parameter*, defined as the ratio between the space charge term and the emittance term:

$$\rho = \frac{\hat{I}}{2I_A\gamma}\frac{\sigma^2}{\varepsilon_n^2} \quad . \tag{43}$$

When ρ greatly exceeds unity, the beam behaves as a laminar flow (all beam particles move on trajectories that do not cross), and transport and acceleration require a careful tuning of focusing and accelerating elements to keep laminarity. Correlated emittance growth is typical in this regime, which can be made reversible if proper beam matching conditions are fulfilled, as discussed next. When $\rho < 1$, the beam is emittance dominated (thermal regime) and space charge effects can be neglected. The transition to the thermal regime occurs when $\rho \approx 1$, corresponding to the transition energy

$$\gamma_{tr} = \frac{\hat{I}}{2I_A}\frac{\sigma^2}{\varepsilon_n^2} \quad . \tag{44}$$

For example, a beam with $\hat{I} = 100$ A, $\varepsilon_n = 1$ μm, and $\sigma = 300$ μm is leaving the space charge dominated regime and is entering the thermal regime at the transition energy of 131 MeV. From this example, one may conclude that the space charge dominated regime is typical of low-energy beams. Actually, for such applications as linac-driven free electron lasers, peak currents exceeding kA are required. Space charge effects may recur if bunch compressors are active at higher energies and a new energy threshold with higher $\hat{I}$ must be considered.

7 Correlated emittance oscillations

When longitudinal correlations within the bunch are important, like that induced by space charge effects, beam envelope evolution is generally dependent also on the coordinate along the bunch ζ. In this case, the bunch should be considered as an ensemble of n longitudinal slices of envelope $\sigma_s(z,\zeta)$, whose evolution can be computed from n slice envelope equations equivalent to Eq. (42), provided that the bunch parameters refer to each single slice: γ_s, γ'_s, $k_{sc,s} = k_{sc}g(\zeta)$. Correlations within the bunch may cause emittance oscillations that can be evaluated, once an analytical or numerical solution [13] of the slice envelope equation is known, by using the following correlated emittance definition:

$$\varepsilon_{rms,cor} = \sqrt{\langle\sigma_s^2\rangle\langle\sigma_s'^2\rangle - \langle\sigma_s\sigma_s'\rangle^2} \quad , \tag{45}$$

where the average is performed over the entire slice ensemble, assuming uniform charge distribution within each slice. In the simplest case of a two-slice model, the previous definition reduces to

$$\varepsilon_{rms,cor} = \left|\sigma_1\sigma_2' - \sigma_2\sigma_1'\right| \quad , \tag{46}$$

which represents a simple and useful formula for an estimation of the emittance scaling [14].

The total normalized rms emittance is given by the superposition of the correlated and uncorrelated terms as

$$\varepsilon_{rms,cor} = \langle\gamma\rangle\sqrt{\varepsilon_{rms}^2 + \varepsilon_{rms,cor}^2} \quad . \tag{47}$$

An interesting example to consider here, showing the consequences of non-perfect beam matching, is the propagation of a beam in the space charge dominated regime nearly matched to an external focusing channel, as illustrated in Fig. 7. To simplify our computations, we can neglect acceleration, as in the case of a simple beam transport line made by a long solenoid ($k^2_{ext} = k_{sol}$). The envelope equation for each slice, indicated as σ_s, reduces to

$$\sigma_s'' + k_{ext}^2 \sigma_s = \frac{k_{sc,s}}{\gamma^3 \sigma_s} \ . \tag{48}$$

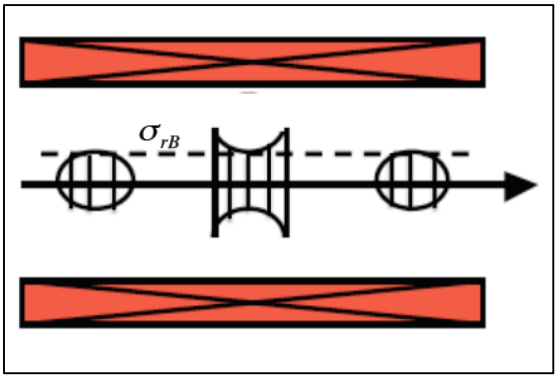


Fig. 7: Schematic representation of a nearly matched beam in a long solenoid. The dashed line represents the reference slice envelope matched to the Brillouin flow condition. The other slice envelopes are oscillating around the equilibrium solution.

A stationary solution corresponding to slice propagation with constant envelope, called *Brillouin flow*, is given by

$$\sigma_{r,B} = \frac{1}{k_{ext}} \sqrt{\frac{\hat{I} g(\xi)}{2\gamma^3 I_A}} \ , \tag{49}$$

where the local dependence of the current $\hat{I}_s = \hat{I} g(\zeta)$ within the bunch has been explicitly indicated. This solution represents the matching conditions for which the external focusing completely balances the internal space charge force. Unfortunately, since k_{ext} has a slice-independent constant value, the Brillouin matching condition is different for each slice and usually cannot be achieved at the same time for all of the bunch slices. Assuming that there is a reference slice perfectly matched (49) with an envelope $\sigma_{r,B}$ and negligible beam energy spread, the matching condition for the other slices can be written as:

$$\sigma_{sB} = \sigma_{rB} + \frac{\sigma_{rB}}{2} \left(\frac{\delta I_s}{\hat{I}} \right) , \tag{50}$$

with respect to the reference slice. Considering a slice with a small perturbation δ_s with respect to its own equilibrium Eq. (50) in the form

$$\sigma_s = \sigma_{s,B} + \delta_s \ , \tag{51}$$

and substituting into Eq. (48), we can obtain a linearized equation for the slice offset

$$\delta_s'' + 2k_{ext}^2 \delta_s = 0 \ , \tag{52}$$

which has a solution given by

$$\delta_s = \delta_0 \cos\left(\sqrt{2} k_{ext} z \right) \ , \tag{53}$$

where $\delta_0 = \sigma_{so} - \sigma_{sB}$ is the amplitude of the initial slice mismatch, which we assume, for convenience, is the same for all slices. Inserting Eq. (53) into Eq. (51) we get the perturbed solution:

$$\sigma_s = \sigma_{s,B} + \delta_0 \cos\left(\sqrt{2} k_{ext} z \right) \ . \tag{54}$$

Equation (54) shows that slice envelopes oscillate together around the equilibrium solution with the same frequency for all slices ($\sqrt{2}k_{ext}$, often called the plasma frequency) dependent only on the external focusing forces. This solution represents a collective behaviour of the bunch, similar to that of the electrons subject to the restoring force of ions in a plasma. Using the two-slice model and Eq. (54), the emittance evolution Eq. (46) results in

$$\varepsilon_{rms,cor} = \frac{1}{4} k_{ext} \sigma_{rB} \left| \frac{\Delta I}{\hat{I}} \delta_0 \sin\left(k_{ext} z\right) \right|, \tag{55}$$

where $\Delta I = \hat{I}_1 - \hat{I}_2$. Notice that, in this simple case, envelope oscillations of the mismatched slices induce correlated emittance oscillations that periodically return to zero, showing the reversible nature of the correlated emittance growth. It is, in fact, the coupling between transverse and longitudinal motion induced by the space charge fields that allows reversibility. With proper tuning of the transport line length or of the focusing field, one can compensate for the transverse emittance growth.

At first, it may seem surprising that a beam with a single charge species can exhibit plasma oscillations, which are characteristic of plasmas composed of two-charge species. However, the effect of the external focusing force can play the role of the other charge species, providing the necessary restoring force that is the cause of such collective oscillations, as shown in Fig. 8. The beam can actually be considered as a single-component, relativistic, cold plasma.

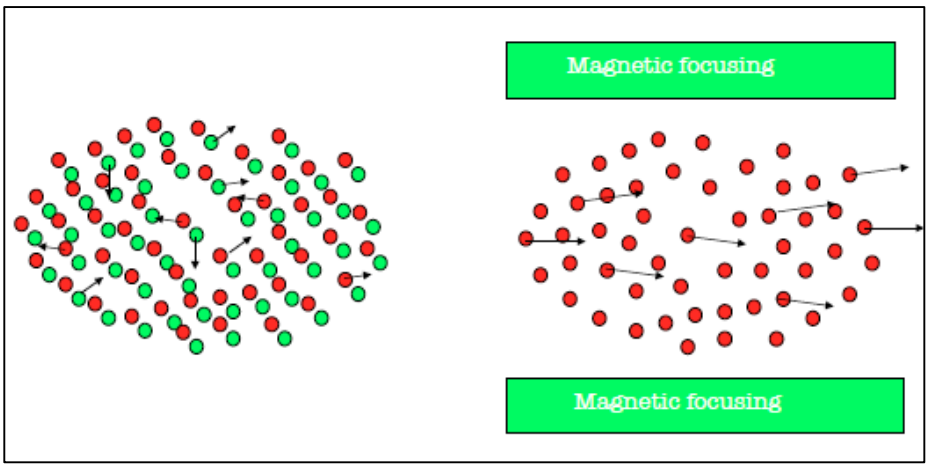


Fig. 8: The restoring force produced by the ions (green dots) in a plasma may cause electron (red dots) oscillations around the equilibrium distribution. In a similar way, the restoring force produced by a magnetic field may cause beam envelope oscillations around the matched envelope equilibrium.

It is important to bear in mind that beams in linacs are also different from plasmas in some important respects [5]. One is that beam transit time through a linac is too short for the beam to reach thermal equilibrium. Also, unlike a plasma, the Debye length of the beam may be larger than, or comparable to, the beam radius, so shielding effects may be incomplete.

8 Matching conditions in a radiofrequency linac

In order to prevent space charge induced emittance growth in a radiofrequency (rf) linac, as in the case of a high brightness photoinjector, and to drive a smooth transition from the space charge to the thermal regime, space charge induced emittance oscillations have to be damped along the linac in such a way that an emittance minimum is obtained at the transition energy Eq. (44). To this end the beam has to be properly matched to the accelerating sections with a Brillouin like flow in order to keep under control emittance oscillations that in this case are provided by the ponderomotive rf focusing force [2] acting in the rf structures. In some case rf focusing is too weak to provide sufficient beam containment. A long solenoid around the accelerating structure is a convenient replacement to provide the necessary focusing.

The matching conditions for a beam subject to acceleration (assuming $\gamma(z) = \gamma_o + \gamma' z$ and $\gamma'' = 0$) can be obtained following the previous example (Brillouin flow). This process can be described using the envelope equation (42) for a generic slice σ_s with external focusing provided by $k_{\text{ext}}^2 = k_{\text{sol}} + k_{\text{rf}}^2$, where $k_{rf}^2 = \frac{\eta}{8}\left(\frac{\gamma'}{\gamma}\right)^2$ and $\gamma' = \frac{eE_{acc}}{mc^2}$. The quantity η is a measure of the higher spatial harmonic amplitudes of the rf wave and it is generally quite close to unity in standing wave (SW) structures and close to 0 in travelling wave (TW) structures [15].

Being now γ(*z*) a time-dependent function, a stationary solution of Eq. (42) cannot be found by simply looking for a constant envelope solution. A possible way to find an 'equilibrium' solution is described hereafter. By substituting the reduced variable $\hat{\sigma} = \sqrt{\gamma}\sigma_s$ [16] in the envelope equation (42) we obtain

$$\hat{\sigma}'' + \hat{k}_{ext}^2 \hat{\sigma} = \frac{\hat{K}_{sc}}{\hat{\sigma}} + \frac{\varepsilon_n^2}{\hat{\sigma}^3} \tag{56}$$

with the scaled parameters $\hat{k}_{\text{ext}}^2 = k_{\text{ext}}^2 + \frac{1}{4}\left(\frac{\gamma'}{\gamma}\right)^2 = k_{\text{sol}} + \frac{1}{4}\left(\frac{\gamma'}{\gamma}\right)^2\left(1+\frac{\eta}{2}\right)$ and $\hat{K}_{sc} = k_{sc}/\gamma^2$. Eq. (56) is equivalent to Eq. (42) but the damping term has disappeared and the $\hat{k}_{ext}^2$ and $\hat{K}_{sc}$ parameters have the same γ^{-2} dependence. In the space charge regime the emittance term can be neglected in Eq. (56) and an equilibrium solutions in the reduced variables (called the 'invariant envelope' in the literature [2]) is given by $\hat{\sigma}_{sc} = \frac{\sqrt{\hat{K}_{sc}}}{\hat{k}_{ext}}$, corresponding to the matching conditions for the beam envelope:

$$\sigma_{sc} = \sqrt{\frac{2\hat{I}}{\gamma I_A\left(\Theta^2 + \gamma'^2\left(\frac{\eta}{2}+1\right)\right)}} \quad \text{for } \rho > 1 \tag{57}$$

where $\Theta = \frac{eB}{mc}$.

The expression for the emittance oscillation in the space charge dominated regime, i.e. when $\gamma < \gamma_{tr}$, can be obtained from Eq. (55) using reduced variables and results:

$$\varepsilon_n = \frac{1}{\gamma}\sqrt{\frac{\hat{I}}{34 I_A}}\left|\frac{\Delta I}{\hat{I}}\delta_o \sin\left(\frac{\left(\Theta^2 + \gamma'^2\left(\frac{\eta}{2}+1\right)\right)^{1/2}}{2\gamma} z\right)\right| \quad . \tag{58}$$

Before the transition energy is achieved the emittance performs damped oscillations with wavelength depending on the external fields and with amplitude depending on the current profile. A careful tuning of the external fields and bunch charge profile can minimize the value of the emittance at the injector extraction. A successful application of the emittance compensation technique can be seen in [17, 18].

When the beam enters in the thermal regime an equilibrium solution can be found directly from Eq. (42) neglecting the space charge term. The result is

$$\sigma_{\text{th}} = \sqrt{\frac{2\varepsilon_n}{\left(\Theta^2 + \frac{\eta}{2}\gamma'^2\right)^{1/2}}} \quad \text{for } \rho < 1 \tag{59}$$

and no correlated emittance oscillations are expected. Note also that Eq. (57) scales like $\gamma^{-1/2}$ while Eq. (59) is independent of γ.

Acknowledgements

I wish to thank A. Cianchi, E. Chiadroni, J.B. Rosenzweig, A.R. Rossi, A. Bacci, and L. Serafini for the many helpful discussions and suggestions.

References

[1] T. Shintake, Review of the worldwide SASE FEL development, Proc. 22nd Particle Accelerator Conf., Albuquerque, NM, 2007 (IEEE, New York, 2007), p. 89. https://doi.org/10.1109/PAC.2007.4440331

[2] L. Serafini and J.B. Rosenzweig, *Phys. Rev. E* **55** (1997) 7565. https://doi.org/10.1103/PhysRevE.55.7565

[3] M. Reiser, *Theory and Design of Charged Particle Beams* (Wiley, New York, 1994). https://doi.org/10.1002/9783527617623

[4] J.B. Rosenzweig, *Fundamentals of Beam Physics* (Oxford University Press, Oxford, 2003). https://doi.org/10.1093/acprof:oso/9780198525547.001.0001

[5] T. Wangler, *Principles of RF Linear Accelerators* (Wiley, New York, 1998). https://doi.org/10.1002/9783527618408

[6] S. Humphries, *Charged Particle Beams* (Wiley, New York, 2002).

[7] N. Pichoff, Beam dynamics basics in RF linacs, Proc. CERN Accelerator School: Small Accelerators, Zeegse, The Netherlands, 2005, CERN-2006-012. https://doi.org/10.5170/CERN-2006-012

[8] M. Migliorati *et al*., *Phys. Rev. ST Accel. Beams* **16** (2013) 011302. https://doi.org/10.1103/PhysRevSTAB.16.011302

[9] M. Ferrario *et al*., *Nucl. Instrum. Methods Phys. Res. B* **309** (2013) 183. https://doi.org/10.1016/j.nimb.2013.03.049

[10] A. Cianchi *et al*., *Nucl. Instrum. Methods Phys. Res. A* **720** (2013) 153. https://doi.org/10.1016/j.nima.2012.12.012

[11] F.J. Sacherer, *IEEE Trans. Nucl. Sci*. **NS-18** (1971) 1105. https://doi.org/10.1109/TNS.1971.4326293

[12] K. Floettmann, *Phys. Rev. ST Accel. Beams* **6** (2003) 034202. https://doi.org/10.1103/PhysRevSTAB.6.034202

[13] M. Ferrario *et al*., *Int. J. Mod. Phys. A* **22** (2007) 4214. https://doi.org/10.1142/S0217751X07037779

[14] J. Buon, Beam phase space and emittance, Proc. CERN Accelerator School: 5th General Accelerator Physics Course, Jyvaskyla, Finland, 1992, CERN-94-01. https://doi.org/10.5170/CERN-1994-001

[15] J. B. Rosenzweig, L. Serafini, Phys. Rev. E 49, 1599 (1994). https://doi.org/10.1103/PhysRevE.49.1599

[16] C. Wang, Phys. Rev. E 74, 046502 (2006). https://doi.org/10.1103/PhysRevE.74.046502

[17] R. Akre et al., *Phys. Rev. ST Accel. Beams* 11, 030703 (2008). https://doi.org/10.1103/PhysRevSTAB.11.030703

[18] M. Ferrario et al., Phys. Rev. Lett. 99, 234801 (2007). https://doi.org/10.1103/PhysRevLett.99.234801

[19] D. H. Dowell et al., *Phys. Rev. ST Accel. Beams* **12**, 074201 (2009). https://doi.org/10.1103/PhysRevSTAB.12.074201

Proceedings of the CAS–CERN Accelerator School: Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018)

Introduction to Beam Transfer

B. Goddard
CERN, Geneva, Switzerland

Abstract
Beam transfer refers to the process of moving a collection of particles between accelerators or onto external targets, dumps or measurement devices. A recall of the salient features of transverse beam dynamics is given, and some common manipulations portrayed in normalized phase space. Different beam transfer processes and techniques are introduced, including key generic requirements, constraints, and performance indicators

Keywords
Injection; extraction; kickers; septa; losses; emittance.

1 Introduction

There are, by some counts, over 18 000 particle accelerators worldwide, almost all of which require the injection and/or extraction of the particles for beam transfer. High-energy particles have been widely used over the past 100 years since Rutherford's seminal scattering experiments, to extend the knowledge of the fundamental nature of matter and forces at smaller and smaller scales. Today, accelerators have become a common tool in society, and alongside the rare behemoths like the Large Hadron Collider (LHC), there are linacs, cyclotrons, and synchrotrons used in materials, medical, nuclear, high-energy physics, and other domains.

Beam transfer can be from a linear to a circular accelerator, or between circular accelerators. Extraction is the process of removal of a beam from an accelerator, which can either be for the transfer to another accelerator or for use on a target, dump, or measurement system. Injection is the insertion of a beam into another accelerator, usually (but certainly not always) to accelerate to higher energy. The specific equipment systems used for injection and extraction need to be designed for several constraints: to transfer the beam with minimum beam loss; to achieve the desired beam parameters; and frequently to minimize dilution of emittance.

The different injection and extraction techniques described in this paper share many similarities and often use the same hardware concepts. One important difference is that extraction is often at higher beam rigidity, which reduces the effects of space charge but requires stronger and normally longer deflecting systems, with a consequent need for a larger footprint in the lattice and insertions.

1.1 Single-turn injection and extraction

Single-turn injection and extraction methods are used for both lepton and hadron machines, see e.g. [1–5], and generally involve a septum system to deflect the beam into (or out of) the accelerator aperture and a kicker system to deflect the beam onto (or away from) the closed orbit. A closed orbit bump is sometimes used to reduce the kicker strength required. For these methods, the beam losses can be very low. The emittance dilution is defined by the delivery precision, the kicker flat top ripple, and septum stability. For both injection and extraction, the circulating beam quality can be degraded by septum stray fields and by the kicker field rise time. Injecting a bunched beam into another accelerator also generally requires that the momentum spread and phase be matched to the radiofrequency (RF) bucket.

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-005.109

1.2 Multiple-turn injection and extraction

Injection can also take place over multiple turns [6–8] to fill the circumference of a receiving accelerator, in order to increased the obtained bunch intensity. Multiple-turn extraction [9–11] is used to produce a spill length longer than the circumference of the accelerator.

A variety of multiple-turn injection and extraction schemes exist, and these can be very different between lepton and hadron machines. Lepton injection can take advantage of synchrotron radiation damping to achieve high beam brightness, while for hadron machines space charge effects dominate, especially at low energy. High brightness proton injection can make use of phase-space painting, particularly with H^- charge exchange injection, while resonant multiple turn extraction can provide quasi-continuous particle fluxes for periods which range from milliseconds to hours. The additional systems required for these more advanced injection and extraction techniques include multiple RF systems, programmed fast closed orbit bumps, stripping foils, and non-linear lattice elements.

2 Recall of transverse phase space representation

A particle accelerator beam typically consists of 10^6–10^{13} particles. We can apply the methods of statistical mechanics and treat properties of the statistical ensemble, not of individual particles. This is a useful concept for representation of many common beam manipulations and hence for helping the understanding of injection and extraction processes.

2.1 Phase space

In relativistic classical mechanics, the motion of a single particle is fully defined at any instant t, if the position r and momentum p of the particle are given with the forces (fields) F acting on the ith particle:

$$\begin{aligned} \vec{r_i} &= x_i\hat{x} + y_i\hat{y} + z_i\hat{z}, \\ \vec{p_i} &= p_{xi}\hat{x}_i + p_{yi}\hat{y} + p_{zi}\hat{z}, \\ \vec{F_i} &= F_{xi}\hat{x}_i + F_{yi}\hat{y} + F_{zi}\hat{z}. \end{aligned} \tag{1}$$

It is convenient to use a six-dimensional phase space representation, where the ith particle has coordinates

$$P_i \equiv \{x_i, p_{xi}, y_i, p_{yi}, z_i, p_{zi}\}. \tag{2}$$

Generally, the three planes can be considered decoupled (with some important exceptions), and it is very common to study particle evolution separately in each plane:

$$\{x_i, p_{xi}\}, \{y_i, p_{yi}\}, \{z_i, p_{zi}\}. \tag{3}$$

For the two transverse planes, we generally use a modified phase space, see Fig. 1, where the transverse momentum coordinates p_{xi} and p_{yi} are replaced by angles x' and y':

$$\begin{aligned} p_x \rightarrow x' &= \frac{\mathrm{d}x}{\mathrm{d}s} = \tan\theta_x, \\ p_y \rightarrow y' &= \frac{\mathrm{d}y}{\mathrm{d}s} = \tan\theta_y. \end{aligned} \tag{4}$$

The angle is related to the momentum by

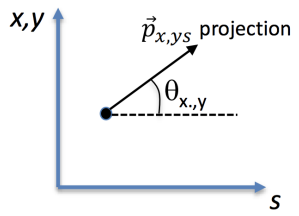


Fig. 1: x' and y' from projection on the $x(y)$–s plane

$$p_x = \gamma_r m_0 \frac{\mathrm{d}x}{\mathrm{d}t} = \gamma_r m_0 v_s \frac{\mathrm{d}x}{\mathrm{d}s} = \beta_r \gamma_r m_0 c x',$$
$$p_y = \gamma_r m_0 \frac{\mathrm{d}y}{\mathrm{d}t} = \gamma_r m_0 v_s \frac{\mathrm{d}y}{\mathrm{d}s} = \beta_r \gamma_r m_0 c y'. \tag{5}$$

One important note is that (unlike x and p_x) x and x' are not canonical conjugate variables unless γ_r and β_r remain constant (i.e. no acceleration).

2.2 Poincaré map

A first recurrence map, or Poincaré map, is the intersection of a periodic orbit in the state space of a continuous dynamical system with a certain lower-dimensional subspace, called the Poincaré section, transversal to the flow of the system [12]. More simply, to generate a Poincaré map, we plot the evolution of the phase space coordinates at one physical location over time (turns), see Fig. 2.

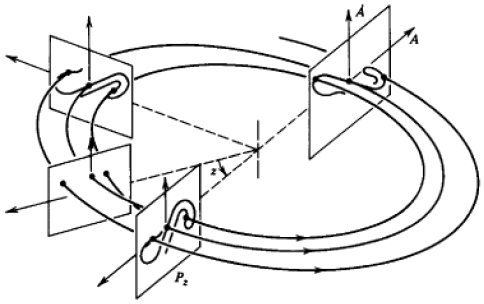


Fig. 2: A Poincaré map is generated by plotting the phase space coordinates at one physical location over time

The Poincaré map is a common representation in non-linear systems, where characteristic behaviour can become apparent. Some examples are shown in Fig. 3.

2.3 Phase space ellipses and emittance

For a circular accelerator, if we make the $\{x, x'\}$ or $\{y, y'\}$ Poincaré map for any single particle at one location over many turns, it traces an ellipse, see Fig. 4. This ellipse at this location always has the same shape and orientation for all particles—the local ellipse form is a property of the focussing lattice. The ellipse changes shape and orientation as it propagates along the accelerator, but the area is preserved.

For many calculations and applications, the canonical conjugate variable pair called the *action-angle variables* (J_x, ϕ_x) are a practical concept. The relation between Cartesian and action-angle variables is given in Fig. 4. The figure shows the well-known elliptic particle trajectory in horizontal phase space. The area A of the phase space ellipse is proportional to action J_x via $A = 2\pi J_x$.

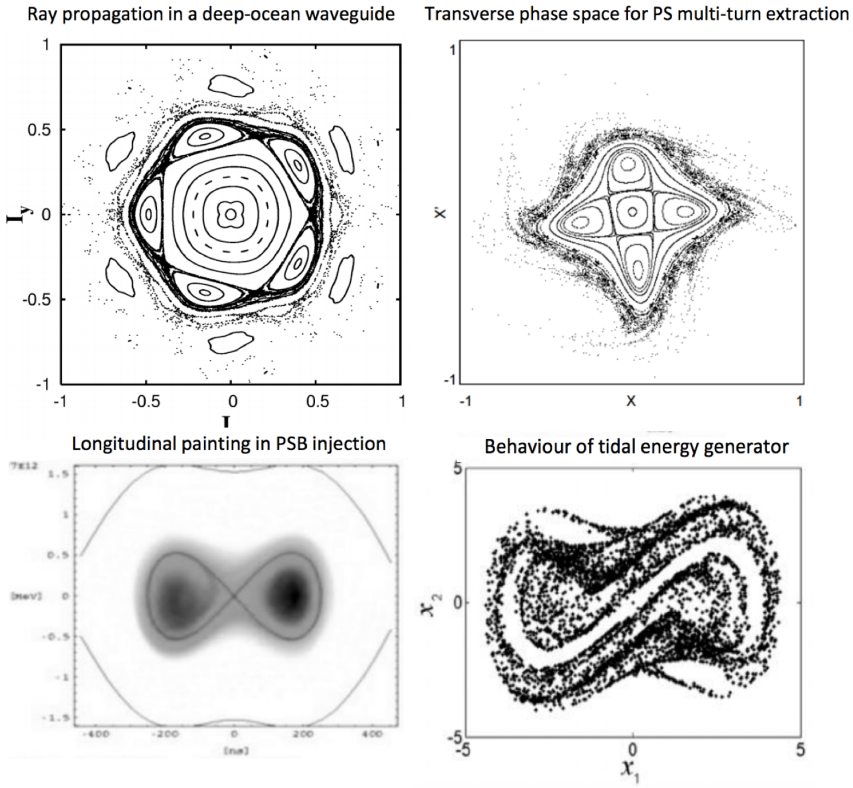


Fig. 3: Example Poincaré maps from different non-linear physical systems, including CERN's Proton Synchrotron (PS) and Proton Synchrotron Booster (PSB).

The transformation between action-angle variables and Cartesian variables is summarized in equations (6) and (7). The action J_x is a measure of the particle amplitude of motion, see Eq. (7):

$$\begin{aligned} 2J_x &= \gamma_x x^2 + 2\alpha_x x' x + \beta_x x'^2, \\ \tan\phi_x &= -\beta_x x \frac{x}{x'} - \alpha_x, \end{aligned} \tag{6}$$

$$\begin{aligned} x &= \sqrt{2\beta_x J_x}\cos\phi_x, \\ x' &= -\sqrt{\tfrac{2J_x}{\beta_x}}(\sin\phi_x + \alpha_x \cos\phi_x). \end{aligned} \tag{7}$$

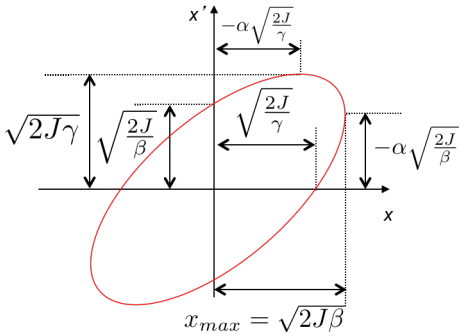


Fig. 4: Phase space in the horizontal plane. The area of the ellipse defines the action J_x, with $A = 2\pi J_x$

The characteristics of the ellipse at a certain location are parameterized by the Twiss parameters, which describe the aspect ratio and orientation. The size of the ellipse (area) is defined by the distribution of angles and positions in the beam. It is very important to recall the following relation, which defines the Courant–Synder invariant as the area of the ellipse $A(s)$, a conserved quantity:

$$A(s) = 2\pi J_x = \pi(\gamma x^2 + 2\alpha x x' + \beta x'^2 = x_{\max}^2/\beta). \tag{8}$$

Very importantly, the emittance is defined as the mean action of all particles in the beam:

$$\varepsilon_x = \langle J_x \rangle. \tag{9}$$

2.4 Normalized phase space

Ellipses are mathematically more difficult to manipulate than circles, and the concept of normalized phase space is used to simplify the representation and highlight the key physical processes. The normalized phase space coordinates are obtained from the real phase space coordinates by a simple linear transformation:

$$\begin{bmatrix} \overline{\mathbf{X}} \\ \overline{\mathbf{X}'} \end{bmatrix} = \mathbf{N} \begin{bmatrix} x \\ x' \end{bmatrix} = \sqrt{\frac{1}{\beta_s}} \begin{bmatrix} 1 & 0 \\ \alpha_s & \beta_s \end{bmatrix} \begin{bmatrix} x \\ x' \end{bmatrix} \tag{10}$$

From this the transformations for the individual coordinates are easily written down as

$$\overline{\mathbf{X}} = \sqrt{\frac{1}{\beta_s}} x, \qquad \overline{\mathbf{X}'} = \sqrt{\frac{1}{\beta_s}} \alpha_s x + \sqrt{\beta_s} x'. \tag{11}$$

In normalized phase space the elliptical trajectories from real phase space become circles, see Fig. 5.

While the equation of the ellipse for real phase space is given by Eq. (6), in normalized phase space it becomes the equation of a circle:

$$2J_x = \overline{x}^2 + \overline{x}'^2. \tag{12}$$

The area enclosed by the trajectory in real as well as normalized phase space is $A = 2\pi J_x$.

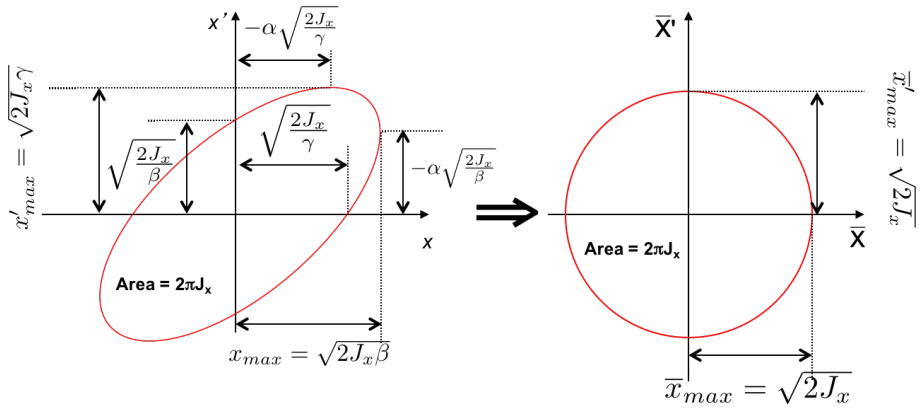


Fig. 5: In normalized phase space the trajectories are circles. The area stays the same and is linked to the action of the particle.

2.5 Common phase space manipulations

Some commonly occurring phase space transformations are illustrated as a precursor to the discussions of the different injection and extraction techniques.

In normalized phase space this is just a clockwise rotation of the distribution around the origin by a phase advance θ, Fig. 6, maintaining the radius of each particle:

$$\begin{bmatrix} x_1 \\ x_1' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_0 \\ x_0' \end{bmatrix}. \tag{13}$$

A dipole deflection (kick) k is a vertical displacement in normalized phase space. The two are combined together to described a kick followed by a transport.

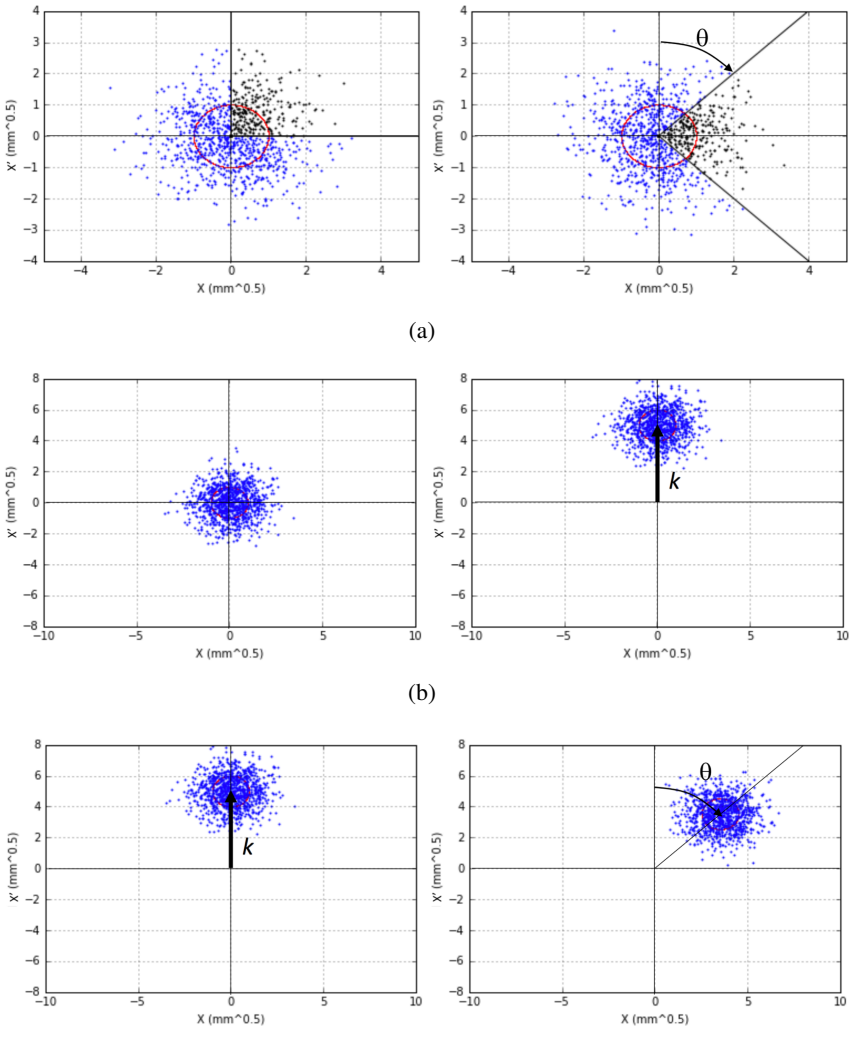
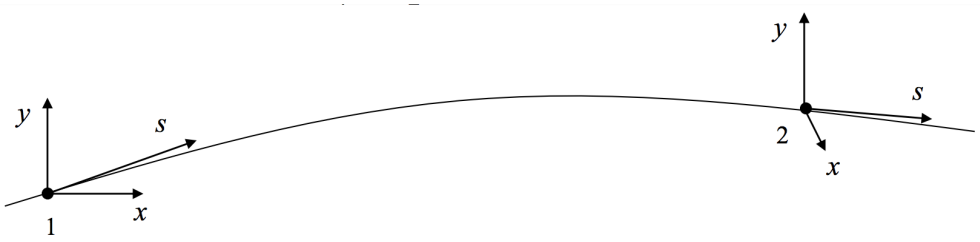


Fig. 6: Manipulations of particle distributions in normalized phase space: (a) rotation of the particle distribution corresponding to transport along a lattice by phase advance θ; (b) displacement of the particle distribution corresponding to a dipole deflection k; (c) a dipole kick followed by transport.

3 Beam transport and difference between rings and transfer lines

As will be already firmly ingrained in the attentive student's mind, each accelerator element has an associated matrix (with the cos-like and sin-like terms) which describes how the input particle coordinates

are transferred from the entrance to the exit of the element. The generalized beam transport through a number of n elements moving from s_1 to s_2, each with transfer matrix M_i, is then simply the product of all the individual transfer matrices. This is illustrated for transport in a beam line in Fig. 7.



$$\begin{bmatrix} x_2 \\ x_2' \end{bmatrix} = \mathbf{M}_{1\to 2} \cdot \begin{bmatrix} x_1 \\ x'_1 \end{bmatrix} = \begin{bmatrix} C & S \\ C' & S' \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x'_1 \end{bmatrix} \qquad \mathbf{M}_{1\to 2} = \prod_{i=1}^{n} \mathbf{M}_n$$

Twiss parameterisation

$$\mathbf{M}_{1\to 2} = \begin{bmatrix} \sqrt{\beta_2/\beta_1}\,(\cos\Delta\mu + \alpha_1 \sin\Delta\mu) & \sqrt{\beta_1\beta_2}\,\sin\Delta\mu \\ \sqrt{1/\beta_1\beta_2}\,[(\alpha_1 - \alpha_2)\cos\Delta\mu - (1+\alpha_1\alpha_2)\sin\Delta\mu] & \sqrt{\beta_1/\beta_2}\,(\cos\Delta\mu - \alpha_2 \sin\Delta\mu) \end{bmatrix}$$

Fig. 7: Generalized beam transport in a *beam line* using the product of individual transfer matrices

In a circular accelerator, the transport is similarly generalized, with the important difference that the one-turn matrix must result in a unique solution for the different Twiss parameters at any s location, since a periodic boundary condition exists. This periodicity condition for one turn in a closed ring imposes $\alpha_1 = \alpha_2$, $\beta_1 = \beta_2$, and $D_1 = D_2$, which uniquely determines $\alpha(s)$, $\beta(s)$, $\mu(s)$, and $D(s)$ around the whole ring, see Fig. 8. The one-turn phase advance $\mu_2 - \mu_1$ reduces to the machine tune Q.

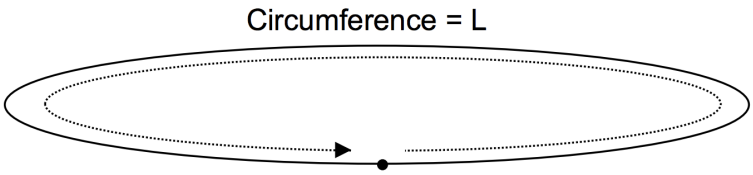
In a transfer line this periodicity condition is absent: the Twiss parameters are still of great use in the description of the beam transport, but the parameters are propagated from the initial conditions at the start of the line. In a sense, they are therefore dependent on the beam distribution, since for non-Gaussian distributions (as obtained in slow extraction, for example), the best-fitting ellipse at the extraction point for the extracted beam distribution may differ greatly from the periodic conditions defined by the circular lattice.

3.1 Effect of dipole kick in a transfer line

In a transfer line or over one turn in a ring we can calculate the effect of a dipole kick k at one location. Using the Twiss parameters, we calculate the downstream position x (or y) and angle $x'_0 = k$ (or $y'_0 = k$) as a function of initial coordinates and lattice functions. We can often, for beam transfer, assume an initial x (or y) $= 0$, as we are deflecting from or onto the closed orbit, which gives

$$x_s = x'_0 \sqrt{\beta_0 \beta_s} \sin 2\pi(\mu_s - \mu_0). \tag{14}$$

Two points are immediately obvious by inspection of (14). First, a 90 degree phase advance $\Delta\mu$ and large β at the deflecting element and observation point give maximum deflection x_s. This is important in designing an injection or extraction system, where the optimization of the system will depend on the location of the different elements in the lattice. Second, the beam as a whole performs



$$\text{One turn} \qquad \mathbf{M}_{1\to 2} = \mathbf{M}_{0\to L} = \begin{bmatrix} \cos 2\pi Q + \alpha \sin 2\pi Q & \beta \sin 2\pi Q \\ -1/\beta \left(1+\alpha^2\right) \sin 2\pi Q & \cos 2\pi Q - \alpha \sin 2\pi Q \end{bmatrix}$$

Fig. 8: Beam transport in a *ring* using the product of individual transfer matrices, where the periodicity condition fixes the Twiss parameters around the whole ring.

transverse betatron oscillations after a kick. These oscillations can be a short fraction of one wavelength, or can stretch for many wavelengths over a large distance, see Fig. 9.

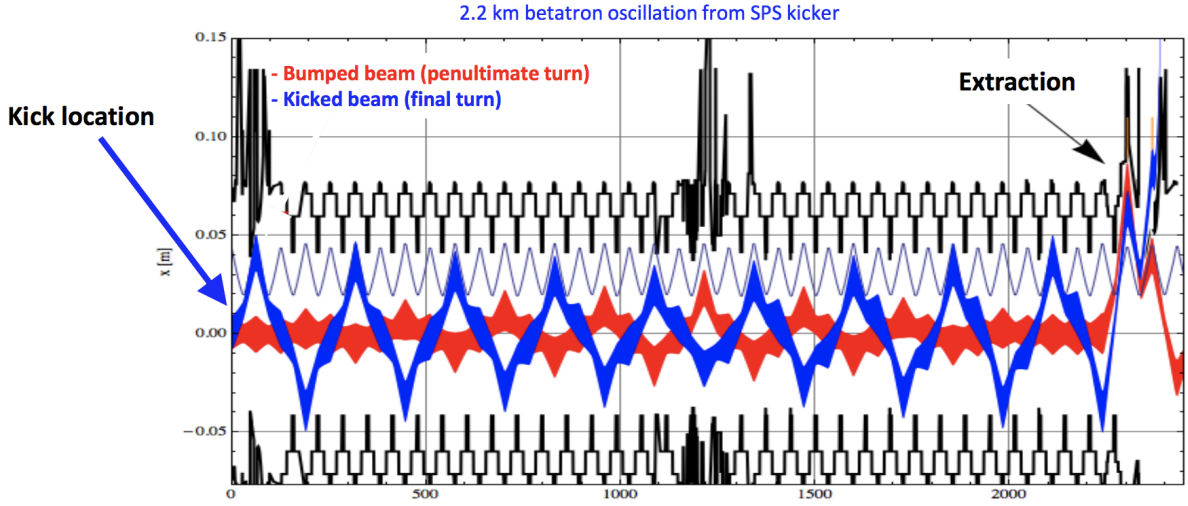


Fig. 9: Betatron oscillation (blue) over several kilometres, from a kicker in one part of the SPS used to extract the beam in a different straight section.

4 The beam transfer process

Almost by definition, beam transfer is a pulsed or transient process, although this can be stretched to several hours for the longest slow extraction techniques. It is often a critical event (or series of events) in the accelerator cycle, requiring specialized equipment not used at other times in the cycle. If badly designed or not optimized it can have a large negative impact on subsequent beam performance, in terms of emittance, intensity, and stability. The beam transfer between machines or onto a target often needs regular operational adjustment to optimize these performance criteria.

For high-energy/high-intensity beams in particular, control of beam losses becomes critical, to avoid excessive irradiation and activation of the hardware, but also to avoid direct damage from beam impact. For both injection and extraction there are a number of desiderata common to most techniques and applications:

– a low fraction of beam lost (for injected/extracted beam and any already circulating beam);

– the need to maintain a large filling factor for injection (i.e. the injection process does not cause too much empty ring circumference);
– precise beam delivery onto the reference orbit;
– minimum perturbation to any circulating beam;
– for injection, as high an energy as possible (sometimes full ring energy);
– a minimum effect on the overall ring performance (e.g. impedance, aperture, etc.);
– a simple setup and operation;
– the need for reliability, reproducibility, and stability.

For transfer lines, some additional design factors need to be taken into account:

– a precise delivery onto the target, dump, or the next injection system;
– precise optical matching to the ring optics (and minimum optical perturbation);
– easy steering and trajectory control;
– filtering of different beam energies or ion species;
– achromatic transport;
– accurate beam characterization.

5 Types of injection and extraction

A wide variety of injection and extraction techniques are deployed in accelerators. A brief summary of the main features of a few of the most common is given below, as an introduction to the concepts explored in detail in the other contributions to this CAS.

5.1 Fast injection and extraction

Fast injection and extraction are typically used to fill or empty a machine, which is often with bunch-to-bucket transfer, with one or many injections (boxcar stacking) [13, 14]. This is typically accomplished with fast kickers and septa. Very fast kicker rise times are required to maximize the amount of beam which can be injected, especially in machines with small circumferences, since the kicker rise and fall times must be significantly shorter than the revolution time. Orbit bumpers can move the beam closer to the septum to reduce the kicker strength needed. Schematic layouts for fast injection and extraction are shown in Fig. 10.

Injection of leptons can take advantage of the strong damping, present from the emission of synchrotron radiation, to accumulate intensity [15]. This is very commonly used at synchrotron radiation light sources, where the top-up operation consists of frequently injecting small amounts of beam to replace losses [16].

5.2 H^- charge-exchange injection

High-brightness proton machines frequently make use of H^- charge exchange injection [17–19]. A linac accelerates negative hydrogen ions which are then merged with a circulating proton beam in a dipole magnet, Fig. 11, before the two electrons are stripped. This technique allows the accumulation of high-brightness beams, since the method allows injection into the already occupied phase space area. Transverse particle distributions can be controlled using phase space painting, to ameliorate space charge effects, reduce beam losses, and increase accumulated intensity. The stripping is achieved with thin foils of carbon or other materials, with a thickness typical in the micrometre range to minimize losses from scattering.

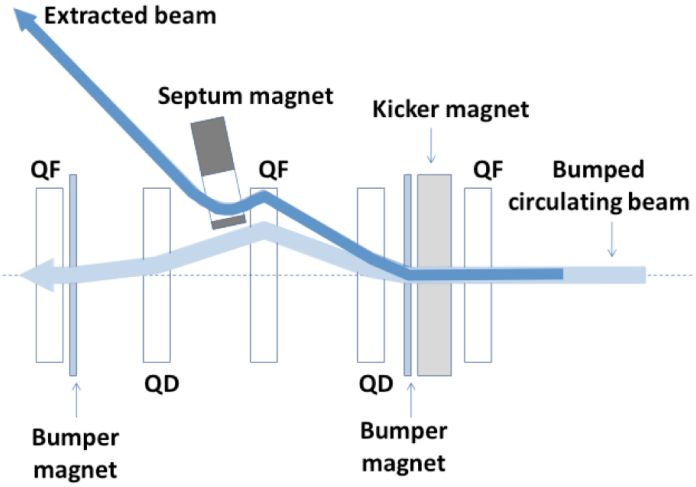


Fig. 10: Schematic of a fast extraction with kicker, septum, and orbit bumpers

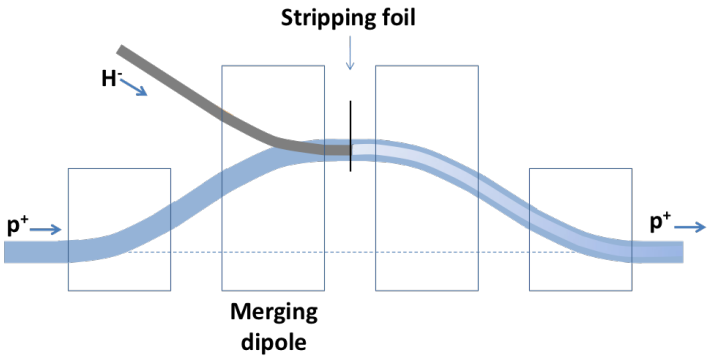


Fig. 11: Merging H^- and p^+ beams in charge exchange injection

5.3 Resonant extraction

Resonant extraction using the third integer is the most common method of providing such uniform spills. A triangular stable area in phase space (usually horizontal) is defined by exciting sextupole elements, and by moving the machine tune close to the third integer resonance. The beam is extracted by driving some particles unstable in a controlled way. The unstable particle amplitudes increase rapidly, following the outward-going separatrix every three turns, and the particles eventually move into the high-field region of a very thin electrostatic septum and are extracted, Fig. 12. Spills can be very long if stochastic excitation is used [20].

6 Slip-stacking injection

In slip-stacking injection, two trains of bunches are merged to increase the bunch intensity, using three separate RF systems (or a single RF system that can operate in a sufficiently flexible manner at two frequencies). A first train of bunches is injected on the closed orbit and captured by the first RF system. This train of bunches is then decelerated, and circulates on a different orbit. A second batch is then injected on the closed orbit and captured by the second RF system. The two trains of bunches have slightly different energies and move relative to each other in phase. When the phase difference reaches zero, both sets of bunches are captured together by the third RF system (or a fast frequency change of one of the two initial systems).

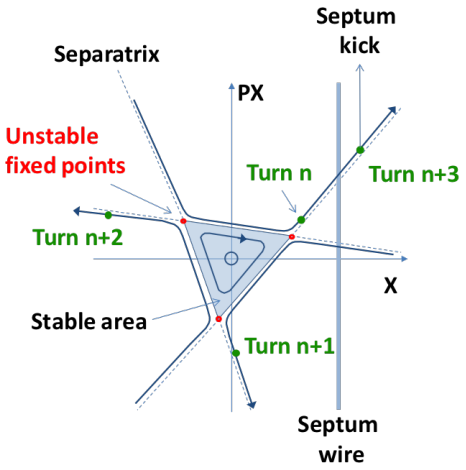


Fig. 12: Resonant extraction in normalized phase space. The amplitudes of particles outside the stable area grow rapidly, following the outward-going separatrix lines every three turns until they reach the electrostatic septum.

7 Conclusion

Beam transfer is ubiquitous to accelerators and complexes of accelerators. It is a pulsed/transient process which, because the equipment is often pulsed, requires specific attention on the timing, controls, and machine protection. The process requires specialized equipment, such as kickers and septa, and has many associated constraints associated with the performance requirements and the space limitations in the accelerators. Ultimately it is a challenging domain which gives scope for very creative and elegant solutions [22–27], some of which will be explored in the companion lecture on Exotic Injection and Extraction Techniques at this CAS.

Acknowledgement

The contributions from W. Bartmann, C. Bracco, M.A. Fraser, V. Kain, and F.M. Velotti, together with the input and discussions with many other colleagues, are greatly appreciated. In particular, the author would like to thank Marcel Gyr for continuing to motivate new generations of beam physicists in beam transfer.

References

[1] S.Myers, a possible new injection and accumulation scheme for LEP, CERN LEP Note 334, 1981.

[2] J. Uythoven et al., Fast injection into the PS2, Proc. PAC09, Vancouver, BC, Canada, 2009.

[3] A.Hilaire, V.Mertens and E.Weiss, Beam transfer to and injection into LHC, Proc. 6th European Particle Accelerator Conference, Stockholm, Sweden, 1998, pp. 2117–2119.

[4] M.Q. Barton, Beam extraction from synchrotrons, Proc. VIIIth Int. Conference on High Energy Accelerators, CERN, Geneva, 1971, pp. 85-88.

[5] O.D. Kazachkovskii, *Atomic Energy* **84** (1998) 345, https://doi.org/10.1007/BF02414883.

[6] G.I. Budker *et al.*, *Atomnaya Énergiya* **19** (1965) 507, https://doi.org/10.1007/BF02153693.

[7] J.W.G. Thomason *et al.*, Injection upgrades for the ISIS synchrotron, Proc. IPAC'10, Kyoto, Japan, 2010, pp. 705–707.

[8] Y. Yamazaki (ed.), Accelerator Technical Design Report for J-PARC, KEK Report 2002-13; JAERI-Tech 2003-044, 2003.

[9] A. Miyamoto *et al.*, Study of slow beam extraction through the third order resonance with transverse phase space manipulation by a mono-frequency RFKO, Proc. 2005 Particle Accelerator Conference, Knoxville, Tennessee, 2005, pp. 1892–1894, https://doi.org/10.1109/PAC.2005.1590948.

[10] L. Badano and S. Rossi, Characteristics of a betatron core for extraction in a proton-ion medical synchrotron, CERN-PS-97-019-DI (CERN, Geneva, 1997).

[11] S. van der Meer, Stochastic extraction, a low-ripple version of resonant extraction, CERN/PS/AA 78-6 (CERN, Geneva, 1978).

[12] G. Teschl, *Ordinary Differential Equations and Dynamical Systems*, Vol. 140 (American Mathematical Society, Providence, RI).

[13] B. Goddard and V. Mertens, Moving the beam into and out of the LHC, in*The Large Hadron Collider: A Marvel of Technology*, Ed. L. Evans (EPFL Press, Laussane, 1999), Chap. 4.4, pp. 114–133.

[14] W. Bartmann, B. Goddard, and C. Hessler, A doublet-based injection-extraction straight for PS2, Proc. of IPAC'10, Kyoto, Japan, 2010, pp. 3945–3947.

[15] P. Collier, Synchrotron phase space injection into LEP, Proc. 16th IEEE Particle Accelerator Conference, Dallas, TX, USA, 1995, pp. 551–553, https://doi.org/10.1109/PAC.1995.504716.

[16] T.S. Ueng *et al.*, Topping up experiments at SRRC, Proc. 5th European Particle Accelerator Conference, 10–14 June 1996.

[17] J. Wei *et al.*, Evolution of the Spallation Neutron Source Ring Lattice, Proc. 20th ICFA Advanced Beam Dynamics Workshop, Chicago IL, 2002.

[18] J. Beebe-Wang *et al.*, Transverse phase space painting for SNS accumulator ring injection, Proc. 1999 Particle Accelerator Conference, New York, 1999, pp. 1743–1745.

[19] V. Danilov *et al.*, *Phys. Rev. ST Accel. Beams* **6** (2003).

[20] S. van der Meer, *IEEE Trans. Nucl. Sci.* **NS-28** (1981), https://doi.org/10.1109/TNS.1981.4331574.

[21] K. Seiya *et al.*, Slip stacking, Proc. CARE-HHH-APD Workshop on Finalizing the Roadmap for the Upgrade of the CERN and GSI Accelerator Complex, CERN, Geneva, Switzerland, 2007, pp. 66–68.

[22] R. Cappi and M. Giovannozzi, *Phys. Rev. ST Accel. Beams* **7** (2004).

[23] C. Bovet *et al.*, The fast shaving ejection for beam transfer from the CPS to the CERN 300 GeV machine, Proc. 1973 Particle Accelerator Conference, San Francisco, California, 1973, p. 438.

[24] K. Johnsen, *Proc. Natl Acad. Sci. USA* **70** (1973) 619, https://doi.org/10.1073/pnas.70.2.619.

[25] D. Neuffer, *IEEE Trans. Nucl. Sci.* **NS-28** (1981).

[26] G.N. Vialov, Y.C. Oganessian and G.N. Flerov, Method for heavy ion beam extraction from a cyclotron with azimuthal variation of the magnetic field, JINR Preprint 1884, Dubna, 1964.

[27] V. Biryukov, *Phys. Rev. Lett.* **74** (1995) 2471, https://doi.org/10.1103/PhysRevLett.74.2471.

Proceedings of the CAS–CERN Accelerator School: Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018)

Injection: Electron Beams

M. Aiba
Paul Scherrer Institute, Villigen, Switzerland

Abstract
Beam injection into electron (or positron) rings is the subject of this paper. Through synchrotron radiation damping, top-up injection is realized: the beam current is kept essentially constant by injecting electrons or positrons atop of the stored beam to compensate for particle loss. This top-up operation mode maximizes the output of the accelerator, namely the integrated luminosity in lepton colliders and the photon flux in light sources. The history of top-up injection is briefly reviewed, and Liouville's theorem and synchrotron radiation damping are expounded from the viewpoint of top-up injection. Finally, various top-up injection schemes are explained, together with advantages and disadvantages.

Keywords
Beam injection; electron beam; lepton circular colliders; light sources; top-up injection.

1 Introduction

In lepton circular colliders and light sources, beam particles are continuously lost through collisions, Touschek scattering, and so on. The outputs of these accelerators, namely the luminosity and the photon beam flux, decrease as the beam current drops. To maximize the output, the beam current is kept essentially constant by injecting electrons or positrons atop of the stored beam, i.e. *top-up injection*. Top-up injection is important, not only to maximize the output but also to stabilize the beams. As discussed in the following, top-up injection is enabled through synchrotron radiation damping. The synchrotron radiation, conversely, heats up the accelerator components; thus top-up injection increases the beam stability by keeping the thermal load constant.

This paper specifically covers top-up injections. For those accelerators that do not necessitate top-up injection, other lectures in this series may be referred to; for most of these beam injections, injection schemes used there are commonly applicable to lepton and hadron beams. The technical and engineering aspects of injection devices, namely kickers and septa, are also not covered here but may be found in the other lectures and in the literature elsewhere.

2 Brief history of top-up injection

In early times, the beams in lepton colliders were dumped for the next filling, as in hadron colliders, when the beam current significantly decreased. It took some time to refill the machine and re-establish stable collisions after the beam dump. The luminosity production was not very efficient then.

In the 1980s, an improved operation mode was established in lepton colliders, e.g. PEP [1], where the beams were no longer dumped but electrons and positrons were injected atop of the stored beam left in the machine so as to compensate for the missing beam current. It is noted that a full-energy injector was necessary for this. Although the physics detectors had to be turned off during the injection to avoid possible damage, the turn-around time was successfully reduced. The injection was typically performed when the beam current dropped by about 30%.

In the 2000s, the technique was further developed, mainly at the KEKB [2] and PEP colliders [3], to top-up operation mode. It was possible to keep the physics detectors turned on during the top-up

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-005.121

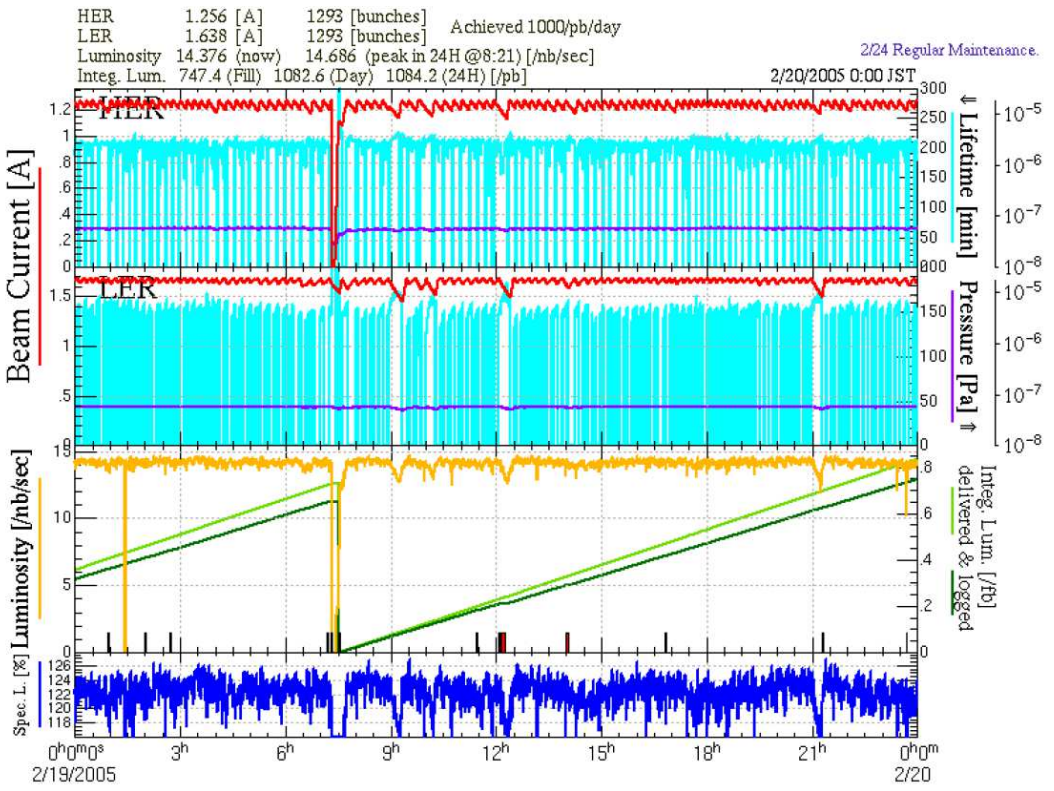


Fig. 1: Top-up operation at KEKB in 2005. Top and middle plots show beam currents in high-energy ring (HER, electron) and low-energy ring (LER, positron), respectively. The beam currents are kept essentially constant with a regular sawtooth pattern, where top-up injection is repeated approximately every 12 min. It can be seen that the current drops from time to time for various reasons. The bottom plot shows approximately constant luminosity with its integral linearly increasing. The beam current stability has been further improved after an injector linac upgrade in 2009; see Ref. [2], from where the figure is taken.

injection, with a mask to the data acquisition for a short time [3]. Without turning the detectors on and off, the injection could be repeated frequently, and the beam current could be kept essentially constant. Figure 1 shows the various parameters during top-up operation of the KEKB.

In light sources too, the stored beam was dumped and refilled before the emergence of top-up injection. A number of light sources were operated in energy-ramping mode, where the beam was injected with lower energy and accelerated to the operation beam energy in the storage ring. In 1990, top-up injection was first applied to the light source SORTEC [4]. The top-up operation mode was able to maintain the full beam current and so the maximum photon flux. Furthermore, the electron beam stability is of importance in providing a stable photon beam to the beamlines. Constant beam current helps to achieve micrometre level electron-beam position stability with orbit feedback, see e.g. Ref. [5]. The top-up operation became standard for light sources in the 2000s [6].

3 Physics behind top-up injection

According to *Liouville's theorem*—'Under the influence of conservative forces the density of the particles in phase space stays constant,' [7]—the injection beam particles cannot be put into the phase space volume occupied by the stored beam particles by an injection kicker. At the same time, it is preferable, if not essential, that the injection process does not disturb the stored beam. Therefore, the injection beam is placed with finite initial betatron or synchrotron oscillation amplitude, being separated from the stored beam at the origin of these oscillations.

Betatron and synchrotron oscillations are damped through *synchrotron radiation damping* in the ring. When a charged particle is accelerated, photons are emitted. The longitudinal acceleration, for instance by RF field, is normally negligible in terms of photon emission. For transverse acceleration with a bending magnetic field, the energy loss for the charged particle due to the emission is proportional to γ^4/ρ, where γ is the Lorentz factor corresponding to the beam energy and ρ is the bending radius. For electron and positron beams, it can be significant.

The energy loss is compensated by RF acceleration turn by turn. Because the radiation energy loss reduces the transverse and longitudinal momenta by the same fraction, whereas the RF acceleration only compensates for the longitudinal component, the betatron oscillation is damped. At the same time, the energy loss depends on the particle energy, i.e., there is greater energy loss for higher-energy particles and vice versa. Hence, synchrotron oscillation is also damped.

The injection beams initially separated in the betatron or synchrotron phase space are merged to the stored beam after several damping times. Since the synchrotron radiation, i.e., photon emission, acts on beam particles as a non-conservative force, Liouville's theorem does not apply here. Hence, the density of the particles in phase space is increased by the injected beam particles. Without synchrotron radiation damping, the entire phase space volume within the machine aperture is to be filled, and thus the top-up injection cannot be repeated indefinitely.

In fact, in hadron machines, top-up injection is not applicable because the radiative energy loss is typically marginal; thus the damping time is too long. It is, however, worth noting that the synchrotron radiation damping can be replaced by other means; for instance, electron cooling is applied to hadron beams, enabling an injection atop of the stored beam. Interested readers are led to literature elsewhere, e.g. Ref. [8].

The oscillations of the injected beam are not harmful for user experiments in light sources, since the injection beam current is negligible compared with the stored beam current. However, betatron oscillations are unfavourable for physics detectors in the collider, as discussed shortly.

4 Top-up injection schemes

There are several ways to perform top-up injection, depending on how the injection beam is separated from the stored beam and which type of kicker is employed. This section describes various injection schemes, together with their advantages and disadvantages.

4.1 Conventional injection scheme

The conventional injection scheme is performed with a static septum and a dynamic *kicker bump* (a series of dipole kickers). The former can be a pulsed magnet but is regarded as a static element for the time scale where the injection beam passes. The latter is turned on and off within one or a few revolution periods, depending on the ring circumference and the field required. The stored beam is brought to the vicinity of the septum wall by the kicker bump, and the injection beam is synchronized to arrive at the septum at the same time. The stored beam is brought back to the closed orbit when the kicker bump is turned off, while the injection beam is brought from the inside of septum channel to the outside. The radio frequency is normally much higher than the ring revolution frequency, accommodating a large number of RF buckets and beam bunches in lepton colliders and light sources. To place the injection beam in the target RF bucket, the precision of synchronization in time should be much shorter than the RF period. For a typical RF frequency of 500 MHz, corresponding to 2 ns RF period, the required precision would be about 100 ps or less. The conventional injection scheme has been used widely, including the first realizations of top-up injection mentioned in Section 3.

Figure 2 shows the configuration of the injection devices and the beam orbits. In Fig. 2, all kickers are located within the straight section. Other configuration, for instance two kickers with several quadrupoles in between, may be employed when a sufficiently long straight section is not afford-

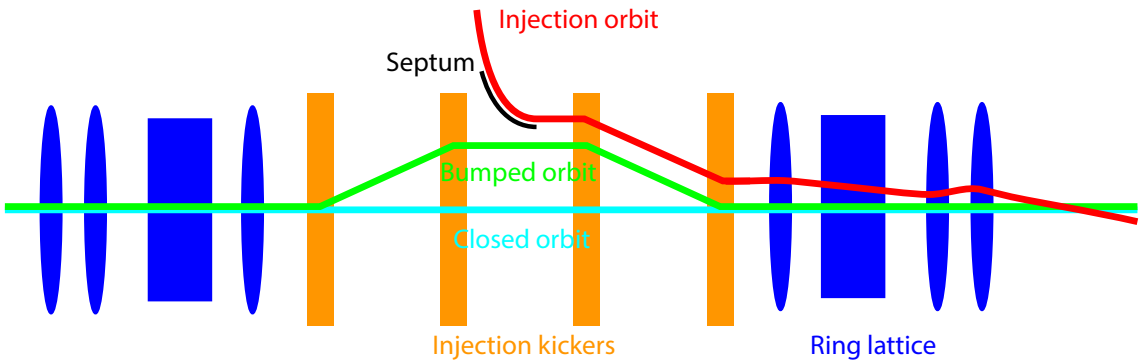


Fig. 2: Conventional injection scheme

able. Also, in high-energy rings, the orbit bump may be formed with the latter configuration, since the quadrupole field is useful to enhance the bump height efficiently for the high-energy beams.

The bump height is determined by the geometry in the former configuration, while it is expressed in the latter case as

$$x_{\mathrm{bump}} = \theta\sqrt{\beta_{\mathrm{k}}\beta_{\mathrm{s}}}\sin\mu, \tag{1}$$

where θ is the kick angle of the first kicker, β_{k} is the beta function at the first kicker, β_{s} is the beta function at the septum location between the two kickers, and μ is the betatron phase advance between the first kicker and the septum. The beta functions at the two kickers may not always be of the same value, and the kick angle of the second kicker is then different from that of the first kicker and adjusted to close the orbit bump. When the second kicker cannot be located at the phase advance of 180° (or its multiple), a third kicker is necessary to close the orbit bump.

The orbit bump excited by the kicker may not, in practice, be fully closed, despite being provided by design. Hence, the stored beam centroid is transversely misplaced. Such a beam disturbance (or transient) is minimized by tuning the kicker strengths. Additional kickers may also be installed to improve the closure of the orbit bump. It is noted that the bump closure deteriorates when sextupoles are located within the orbit bump, which distort the orbit bump through the off-axis field proportional to the horizontal position squared. Such a configuration should be avoided if at all possible.

The injection beam performs a betatron oscillation for many turns until damped through synchrotron radiation. The kicker bump must therefore be turned off before the injection beam hits the septum wall after a few turns, at which point the betatron phase advance at the septum approximately corresponds to a multiple of 360°.

Figure 3 shows the beams positions during the injection process; σ_{s} and σ_{i} are the r.m.s. beam sizes of the stored and injection beams, respectively, S is the septum wall thickness, and n_{s} and n_{i} are the clearances for the stored and injection beams, respectively, i.e., how much the beam centroids are separated from the septum wall in units of r.m.s. beam size.

It can be seen in Fig. 3 that the bump height of $2n_{\mathrm{i}}\sigma_{\mathrm{i}} + S$ and the separation of $n_{\mathrm{s}}\sigma_{\mathrm{s}} + S + n_{\mathrm{i}}\sigma_{\mathrm{i}}$ are required to achieve the given clearances. From the bottom figure, the machine aperture required to accept the injection beam particles within $n_{\mathrm{i}}\sigma_{\mathrm{i}}$ is found to be $n_{\mathrm{s}}\sigma_{\mathrm{s}} + S + 2n_{\mathrm{i}}\sigma_{\mathrm{i}}$. The injection beam size, σ_{i}, can be optimized to minimize the required machine aperture, or conversely to maximize n_{s} and n_{i} for a given machine aperture and septum thickness. Note that it is the injection beam shape in the betatron phase space (Twiss parameters) that is optimized [9], and not the beam size itself. The bump height may be set to a higher value so as to keep the septum away from the closed orbit, in order to, e.g., reduce the beam impedance due to the septum wall if needed.

Instead of placing the injection beam with a separation from the stored beam in the betatron phase

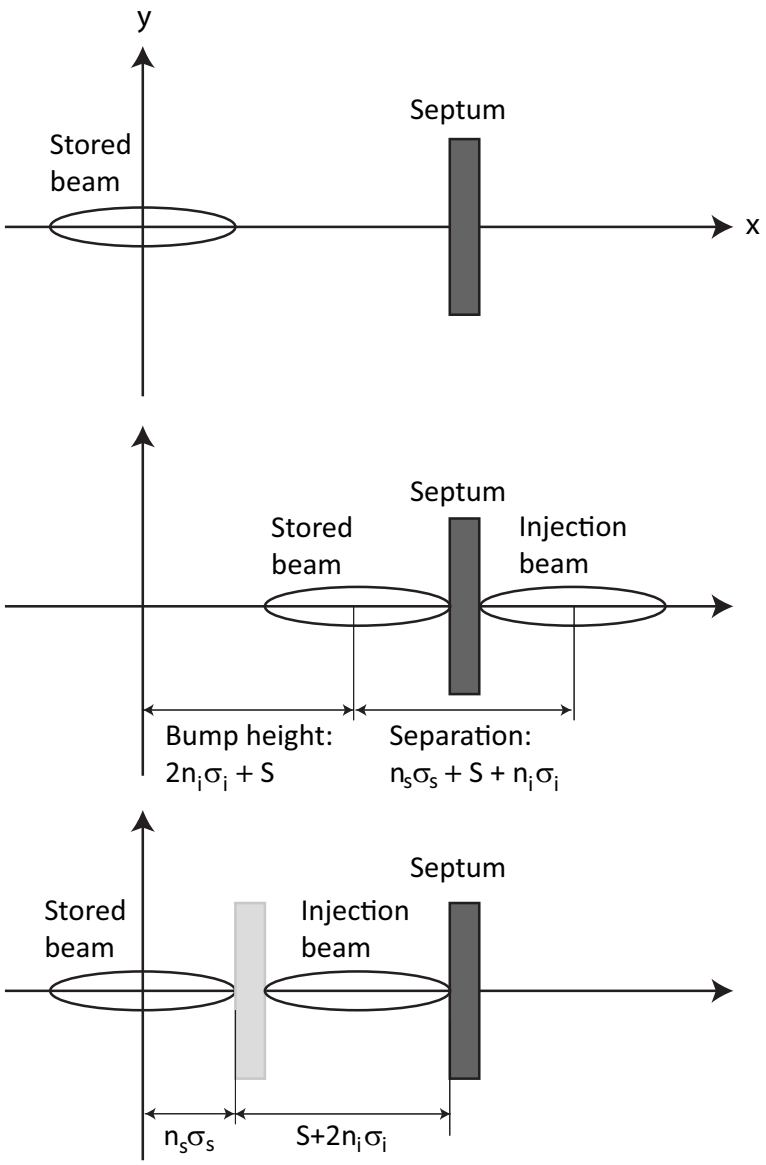


Fig. 3: Beam positions in real space at the exit of the septum: (top) during the injection process before the injection; (centre) at the time of injection with maximum bump height; (bottom) after the injection. The beam position after injection corresponds to the critical case, where the injected beam is closest to the septum, i.e., the number of turns multiplied by the fractional part of horizontal betatron tune is an integer. See text for symbols.

space, the beam energy can be slightly shifted, resulting in a separation in the longitudinal phase space. This alternative conventional injection scheme, *synchrotron phase space injection* [10], was applied to the LEP collider. The dispersion function at the septum location and the injection beam energy were adjusted such that the off-energy injection beam was placed in the corresponding off-energy closed orbit. Therefore, the injected beam performed a synchrotron oscillation rather than a betatron oscillation. The dispersion function at the collision points were essentially zero; thus, adverse radiation doses to the detectors were reduced. Moreover, a higher injection efficiency was achieved [10]. This is attributed to the fact that the injection efficiency may be less sensitive to the injection beam position and angle jitters, since the injection beam is placed transversely on-axis.

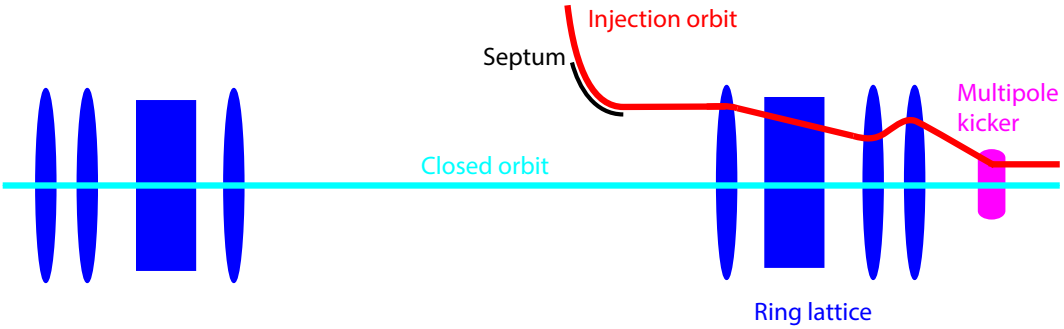


Fig. 4: Multipole kicker injection scheme

4.2 Multipole kicker injection

Multipole kicker injection [11] is developed with the motivation of eliminating the beam disturbance in the conventional scheme. A multipole kicker—quadrupole or higher multipole—excites a magnetic field proportional to x^n, where x is the horizontal position with respect to the magnetic axis of the kicker, and n is an integer: $n = 1$ represents quadrupole, $n = 2$ sextupole, and so on. Since the field is zero on-axis ($x = 0$), the stored beam centroid is not affected, whereas the injection beam passing through off-axis ($x \neq 0$) is deflected. The injection kicker must be turned off within one or a few revolutions in order not to kick the injection beam particles out of the machine aperture. Figure 4 shows the configuration of the injection devices and the beam orbits.

The disturbance to the stored beam is brought to higher order, depending on the kicker employed. A sextupole kicker may be the preferred choice; the disturbance arising from a quadrupole kicker may still be visible as a quadrupolar oscillation (beam size oscillation) while the field strength is rather limited in higher multipoles than sextupole.

The injection beam passing the kicker off-axis is defocused by the feed-down quadrupole components, which may potentially deteriorate the injection efficiency. A *non-linear kicker* that has a plateau in its field profile is proposed to avoid this issue. When the plateau covers the injection beam orbit, the injection beam receives approximately a dipole kick and is not defocused. At the same time, the region with zero or negligibly small field around the magnetic axis can be expanded in the non-linear kicker. It is, however, to be noted that the physical aperture of the non-linear kicker becomes smaller when the field plateau is brought closer to the axis. This may cause an aperture bottle-neck or beam impedance issues, and the plateau may not be set to overlap with the injection orbit in such cases. Figure 5 shows a non-linear kicker and typical field profile, compared with sextupole and octupole kicker profiles.

Like synchrotron phase space injection, a multipole kicker injection can be performed with off-energy injection beams. The dispersion function is adjusted at the location of the kicker, where the injection beam is placed in the corresponding off-energy closed orbit.

4.3 Resonance injection

The kicker in the conventional and multipole kicker injection schemes must be turned off after injection within a few revolution periods, as already discussed. In small rings, however, the fall time of the kicker cannot be so short. Resonance injection may be employed in such small rings (see e.g. Ref. [14]).

When the betatron tune is close to a resonance condition, the betatron phase space topology can be arranged such that the injected beam placed off-axis moves first towards the origin in the following turns and then gets away afterwards. The phase space topology can be varied during this movement, either by moving the working point away from the resonance condition or exciting the resonance less. The quadrupoles to control the tune or multipole magnets to control the resonance excitation are used as the injection kicker equivalently.

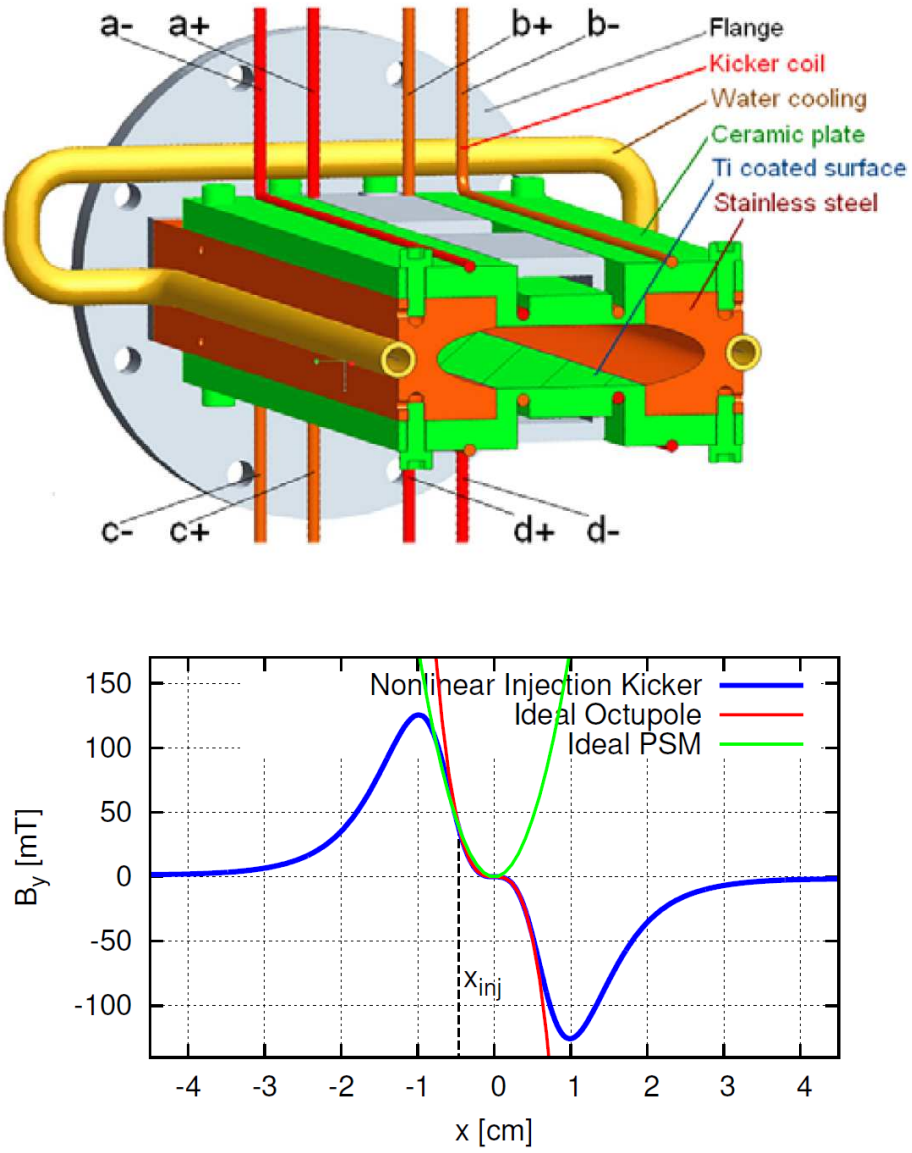


Fig. 5: Non-linear kicker [12] and its field profile [13] compared with sextupole and octupole kickers

4.4 Swap-out injection

For rings with a very small dynamic aperture, where the injected beam separated in the betatron phase space from the stored beam will be partly or fully lost, it is proposed to *swap out* the injection and stored beams [15]. The injection beam is prepared with the design charge and injected into the storage ring. At the same time, the stored beam that is missing a fraction of charge is kicked out from the storage ring.

Such a small aperture ring may be a consequence of aiming at the best performance of the accelerator, e.g., very small equilibrium emittance in a light source. To apply the aforementioned off-energy injections, where the off-energy injection beam is placed in the corresponding off-energy closed orbit (and thus accepted within the small aperture), a finite dispersion function at the septum or the kicker is required. It may be difficult, if not impossible, to realize this in the very low emittance ring. Swap-out injection, however, has no requirement on the dispersion function, since the stored beam is kicked out and the injection beam can then be injected into the nominal-energy closed orbit that *was* occupied by the stored beam.

Swapping-out is performed either bunch by bunch or bunch-train by bunch-train. The former re-

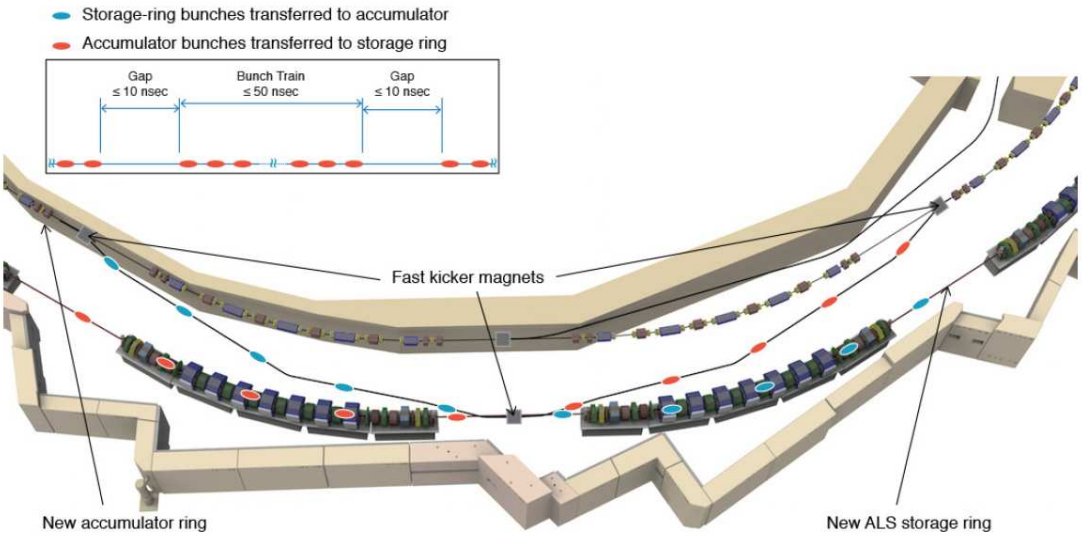


Fig. 6: Swap-out injection (bunch-train swapping) [16]

quires a short pulse kicker (shorter than twice the bunch spacing), which kicks out only one bunch and inserts the injection bunch. The latter may require an accumulator ring to prepare the injection beam and a flat-top kicker. The stored beam kicked out from the storage ring may be treated in various ways, e.g. re-injected into the accumulator ring for recycling or brought to a beam dump. Figure 6 shows the swap-out injection planned for the ALS upgrade project.

The injection kicker strength (kick angle) is adjusted to minimize or, ideally, eliminate the betatron oscillation of the injected beam. In the bunch-by-bunch swapping-out, the residual oscillation may hardly affect the photon beam of light source when the number of bunches in the storage ring is high. However, bunch-train swapping-out requires a stable flat-top kicker to minimize the oscillation.

4.5 Longitudinal injection

Longitudinal injection is also proposed for small aperture rings [17]. The injection beam energy is shifted, and the off-energy injection beam is placed in the corresponding off-energy closed orbit to fit the limited aperture by a dipole kicker. As in the swap-out injection, no requirement on the dispersion function is imposed. The injection timing is also shifted by a fraction of the RF period, typically one-half: the injection beam is inserted between two stored bunches (or two RF buckets). When the pulse length of the dipole kicker is shorter than the bunch spacing, the stored beam remains undisturbed.

Owing to the energy dependence of synchrotron radiation loss, there is a narrow channel of the longitudinal acceptance that extends to the space between RF buckets. An injection beam, with adjusted energy and timing, fitting this channel can be trapped into the RF bucket and merged to the stored beam.

Figure 7 shows the result of a tracking study, where the longitudinal injection scheme is applied to the MAX IV storage ring, equipped with a 100 MHz RF system. It is seen that the injection beam is trapped in the RF bucket and its synchrotron oscillation is almost damped at 18,000 turns (the radiation damping time is about 6900 turns). The injection efficiency strongly depends on the longitudinal injection beam emittance, since the injection beam should fit the narrow channel of the longitudinal acceptance. In this simulation, 100% injection efficiency was observed for the beams with small longitudinal emittance, generated by linac injector.

Most light sources are equipped with a 500 MHz RF system, requiring a kicker pulse length shorter than 2 ns. It is noted that such a short pulse kicker is technically challenging.

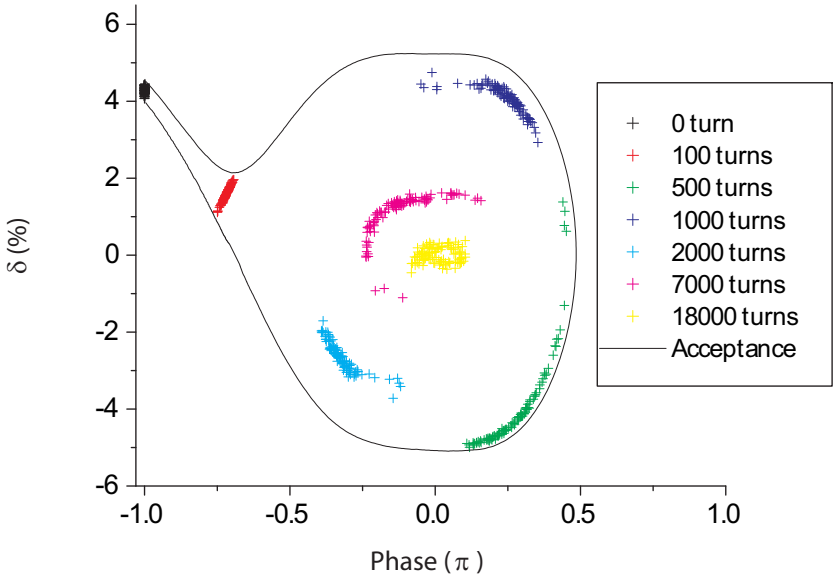


Fig. 7: Injected beam particles in longitudinal phase space, tracked through the MAX IV storage ring lattice up to 18,000 turns. The black solid line represents the longitudinal acceptance, taking into account the energy dependence of synchrotron radiation energy loss. The tracking is six-dimensional, including transverse planes.

Table 1: Top-up injection schemes

Injection scheme	Kicker	Separation
Conventional scheme	Kicker bump	Transverse position
Synchrotron phase space injection	Kicker bump	Beam energy[a]
Multipole kicker injection	Multipole or non-linear kicker	Transverse position
Multipole kicker injection, off-energy	Multipole or non-linear kicker	Beam energy[a]
Resonance injection	Multipole kicker	Transverse position
Swap-out injection	Dipole kicker[b]	Transverse position[c]
Longitudinal injection	Dipole kicker (short pulse)	Beam energy + timing

[a] The beam is also separated transversely at dispersive sections.

[b] Pulse length depends on whether swapping-out is performed bunch by bunch or bunch-train by bunch-train.

[c] Stored beam is kicked out.

4.6 Remarks

The top-up injection schemes described in Section 4 are summarized in Table 1. They are also categorized as off-axis injection or on-axis injection, depending on whether the separation is in betatron phase space (transverse position) or longitudinal phase space (beam energy and/or timing).

So far, the conventional injection scheme has been widely used in colliders and light sources. Among the new schemes, for instance, the multipole kicker injection has been used at the KEK-PF storage ring since 2011 [18] and also implemented in MAX IV as the design injection scheme [19]. The upgrade projects of APS [20] and ALS [21], and the HEPS project [22] are based on swap-out injection to enable beam injection into a small dynamic aperture. Longitudinal injection has been further studied, and possible variants or improvements are proposed [23, 24].

Acknowledgements

I would like to thank my colleagues, Jan Chrin and Andreas Streun, for proofreading.

References

[1] J.T. Seeman, Injection issues of electron-positron storage rings, Proc. B Factories: The State of the Art in Accelerators, Detectors and Physics, SLAC-400 (1992), p. 233

[2] M. Satoh, Simultaneous top-up injection for three different rings in KEK injector linac, Proc. IPAC'10 (2010), p. 4287.

[3] J.T. Seeman, Top-up injection at PEP-II and applications to a circular e^+e^- Higgs factory, Proc. HF2014 (2014), p. 205.

[4] S. Nakamura *et al.*, Present status of the 1 GeV synchrotron radiation source at SORTEC, Proc. EPAC'90 (1990), p. 472.

[5] T. Schilcher *et al.*, Commissioning of the fast orbit feedback at SLS, Proc. PAC'03 (2003), p. 3386.

[6] H. Ohkuma, Top-up operation in light sources, Proc. EPAC'08 (2008), p. 36.

[7] H. Wiedemann, *Particle Accelerator Physics I—Basic Principles and Linear Beam Dynamics*. (Springer, Berlin, 1993).

[8] T. Uesugi *et al.*, *Nucl. Instrum. Methods Phys. Res. A* **545** (2005) 45, https://doi.org/10.1016/j.nima.2004.12.038.

[9] R.P. Fliller III, Optimum Twiss parameters for top off injection in a synchrotron light source, Proc. IPAC'10 (2010), p. 1814.

[10] P. Collier, Synchrotron phase space injection into LEP, Proc. PAC'95 (1995), p. 551.

[11] H. Takaki *et al.*, *Phys. Rev. ST Accel. Beams* **13** (2010), 020705, https://doi.org/10.1103/PhysRevSTAB.13.020705.

[12] T. Atkinson *et al.*, Development of a non-linear kicker system to facilitate a new injection scheme for the BESSY II storage ring, Proc. IPAC'11 (2011), p. 3394.

[13] S.C. Leemann and L.O. Dalin, Progress on pulsed multipole injection for the MAX IV storage rings, Proc. NA-PAC'13 (2013), p. 1052.

[14] T. Takayama, *Nucl. Instrum. Methods Phys. Res. B* **24/25** (1987) 420, https://doi.org/10.1016/0168-583X(87)90674-4.

[15] L. Emery and M. Borland, Possible long-term improvements to the advanced photon source, Proc. PAC'03 (2003), p. 256.

[16] C. Steier *et al.*, On-axis swap-out R&D for ALS-U, Proc. IPAC'17 (2017), p. 2821.

[17] M. Aiba *et al.*, *Phys. Rev. ST Accel. Beams* **18** (2015) 020701, https://doi.org/10.1103/PhysRevSTAB.18.020701.

[18] R. Takai, Beam profile measurement during top-up injection with a pulsed sextupole magnet, Proc. DIPAC'11 (2011), p. 305.

[19] S.C. Leemann, *Phys. Rev. ST Accel. Beams* **15** (2012) 050705, https://doi.org/10.1103/PhysRevSTAB.15.050705.

[20] S. Henderson, Status of the APS upgrade project, Proc. IPAC'15 (2017), p. 1791.

[21] C. Steier *et al.*, ALS-U: a soft X-ray diffraction limited light source, Proc. NA-PAC'16 (2016), p. 263.

[22] Y. Jiao *et al.*, *Phys. Procedia* **84** (2016) 40, https://doi.org/10.1016/j.phpro.2016.11.008.

[23] G. Xu *et al.*, On-axis beam accumulation enabled by phase adjustment of a double-frequency RF system for diffraction-limited storage rings, Proc. PAC'16 (2016), p. 2032.

[24] G. Xu, Longitudinal accumulation in Triple RF systems, Topical Workshop on Injection and Injection Systems (2017), https://indico.cern.ch/event/635514/contributions/2660492.

Proceedings of the CAS–CERN Accelerator School: Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018)

Injection: Hadron Beams

C. Bracco
CERN, Geneva, Switzerland

Abstract

The injection process involves placing a particle beam in a circular hadron accelerator or accumulator at the right time, in the correct orbit, with the right phase space parameters, and minimizing beam losses. Injection insertions are among the most complex parts of an accelerator and, if not properly designed, can compromise the beam quality and, in the worst case, cause machine damage. An overview of the main principles and hardware used for single, multi-turn, and charge exchange injection systems is presented. Common injection errors and their consequences are explained, together with the definition of the machine aperture and the concept of protection devices. Examples of unconventional injection processes are introduced.

Keywords

Single turn; multiturn; charge exchange; mismatch; aperture; machine protection.

1 Introduction

The performance of a hadron collider or accumulator depends on the minimum achievable beam size and the maximum storable intensity. The injection process has a strong influence on both these parameters and, if not properly performed, it might downgrade the effective achievement of a machine. Minimization of the beam size involves positioning the injected beam on the right orbit, with the correct timing, and with matched optics parameters with respect to the ring lattice at the injection point. Small injection losses are then a basic requirement for an efficient transfer into a ring and to avoid any risk of beam-induced machine damage.

Different challenges are encountered, depending on the energy of the injected beams. In the low-energy regime (i.e., $\sim$10 MeV for hadrons), the main concern comes from space charge. Particles in a beam feel electromagnetic forces, which are produced by the beam itself. Assuming an unbunched beam, with a uniform charge density and a circular cross-section, each particle will experience an incoherent transverse tune shift $\Delta Q_{x,y}$ due to space charge, which is proportional to [1]:

$$\Delta Q_{x,y} \propto \frac{N}{\gamma^2 \beta \varepsilon_N}\,, \tag{1}$$

where N is the number of particles per unit length, γ and β the relativistic factors and ε_N the r.m.s. normalized transverse emittance. The resulting tune spread can lead to an emittance blow-up and losses. The effect is stronger at low energy and for large energy densities and hence imposes a limit on the maximum achievable brightness.

Injections of $\sim$100 GeV hadron beams are critical because of the large amount of stored energy (up to 2.4 MJ, which is equivalent to 0.5 kg of TNT, for the Large Hadron Collider (LHC) beams) and the consequent disruptive power in the case of wrong handling. A proper choice and correct design of the injection system, which depends on the beam conditions and the main requirements of the receiving machine, enable minimization of the aforementioned undesired effects and thus maximization of the machine performance while guaranteeing the required protection.

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-005.131

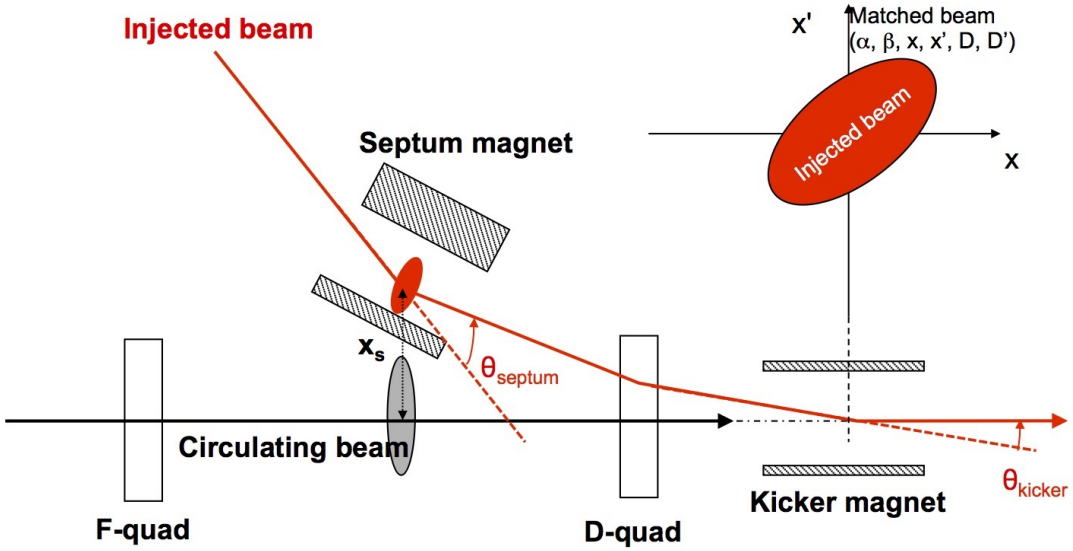


Fig. 1: System used for single-turn injection in one plane

2 Single-turn injection

In the 'single-turn' injection process, one or more bunches of particles are transferred, all at once, into the receiving ring and, as stated in the label, the procedure is completed in one machine revolution. Each transferred bunch must be injected into an empty RF bucket; this implies that the momentum spread $\Delta p/p$ and the phase $\Delta\phi$ of the incoming beam must be matched to those of the bucket defined by the RF cavity system of the ring [2]. Single-turn injection is performed by bringing the beam onto the central closed orbit of the machine by means of septum magnets [3] and fast deflectors or kickers [4], as shown in Fig. 1. These elements are often installed at each side of a defocusing quadrupole to profit from its additional deflection and hence minimize the strength of the kicker (i.e., the required voltage and the cost of the kicker). The position of the injected beam at the septum x_s must be larger than the sum of both the injected and circulating beam envelopes, taking into account the beam size increase due to the energy spread and the thickness of the septum blade, plus closed-orbit distortions and alignment errors. The injected beam must get an angular deflection θ from the kicker equal to

$$\theta = \frac{x_\mathrm{s}}{\sqrt{\beta_\mathrm{k}\beta_\mathrm{s}}\sin(\mu)}\,, \tag{2}$$

where β_k and β_s are the β-functions at the kicker and the septum, respectively, and μ is the relative phase advance between these two elements. Large values of β are favourable to reduce the kick angle. Septa can be DC magnetic or electrostatic elements. The leakage into the field-free region of the septa must be adequately shielded to avoid perturbing the circulating beam, in particular at low energy. Different techniques can be used to reduce the stray fields, as explained in Ref. [3].

In general, in an accelerator chain, particles are transferred to rings with larger radius to cope with the increasing rigidity (larger $B\rho$) of the accelerated beams. This allows the accumulation of more bunches at each acceleration stage by several consecutive injections of bunch trains until the receiving machine is filled. The number of bunches that can be stored in a circular accelerator can ideally be defined as the ratio between the ring circumference and the bunch spacing, which is determined by the RF structure. In reality, a number of gaps with no particles must be guaranteed between trains to allow loss-free injections and extractions. Circulating bunches passing through the injection and extraction kickers, when they are pulsing, would be kicked and lost in the machine aperture. Two consecutive trains must be spaced by at least the rise time (t_rise) of the injection kickers (as shown in Fig. 2). The so-called abort gap must then be synchronized with the rise time (t_abort) of the extraction kickers. As a first

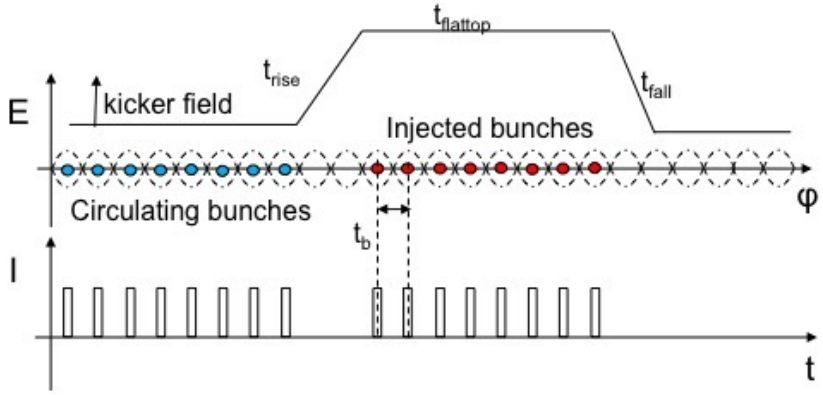


Fig. 2: The full waveform of the injection kickers is overlapped to the structure of the circulating (blue) and injected (red) bunches. Particle-free gaps must be guaranteed to allow loss-free injection and extractions.

approximation, the number of bunches that can be stored in a ring (n_b) can be expressed as

$$n_\mathrm{b} = \frac{n \cdot t_\mathrm{flattop} - [(n-1) \cdot t_\mathrm{rise} + t_\mathrm{abort}]}{t_\mathrm{b}}\,, \tag{3}$$

where n is the number of injections to fill the ring, t_flattop is the length of the flat top of the injection kicker waveform, which defines the number of bunches that can be injected at the time (also dependent on the size of the upstream accelerator), and t_b is the bunch spacing. It is then clear that kickers with fast rise times (from hundreds of nanoseconds up to a few microseconds) are needed to maximize the filling factor that is the ratio between the ideal and the real number of stored bunches.

In ideal conditions, the injected beam should be perfectly superimposed on the closed orbit (x,x',y,y') and fully matched to the optics parameters ($\alpha_{x,y}$, $\beta_{x,y}$, $D_{x,y}$, and $D'_{x,y}$) of the ring at the injection point [5]. Errors in the septum or kicker angle, optical mismatch, and inaccurate steering of the transfer lines will lead to an emittance blow-up through filamentation, owing to non-linear effects (e.g., high-order field components), which introduce amplitude-dependant effects to the particle motion [6]. Algorithms, based on the readings from beam position monitors, can be used to identify and correct possible magnetic errors affecting the transfer process and thus minimize the injection oscillations in the receiving machine [7]. This is fundamental to the reduction of both reduce emittance blow-up and injection losses, owing to the interception of particles with large-amplitude oscillations, which represent a potential risk of damage for the machine mechanical aperture.

3 Aperture and protection devices

Machine protection becomes one of the largest concerns when dealing with high-energy hadron beams. Failures in the transfer and injection process can occur within a few microseconds and active protection systems (e.g., interlocks, magnet current monitors) might not be able to detect the faults and react in time. In this case, a significant amount of beam can be lost on the machine geometrical aperture $A_{x,y}^\mathrm{geom}$ and, in the worst case, cause damage. Dedicated protection elements are hence installed at specific locations to intercept miskicked beams and protect the accelerator components. The design of a passive protection system starts with the definition of the minimum aperture that these collimators have to shade and protect ($A_{x,y}^\mathrm{prot}$, where x and y represent the horizontal and vertical plane, see Fig. 3). In particular, they must be set at an aperture $n_{x,y}$, which, in beam betatron units ($\sigma_{\beta_{x,y}} = \sqrt{\varepsilon_{x,y}\beta_{x,y}}$ with $\varepsilon_{x,y}$ r.m.s. geometric emittance), is

$$n_{x,y} \leq A_{x,y}^\mathrm{prot} \leq \min_{s\in[0,L]} \frac{A_{x,y}^\mathrm{geom}(s) - \Delta_{x,y}(s)}{\sigma_{\beta_{x,y}}} \tag{4}$$

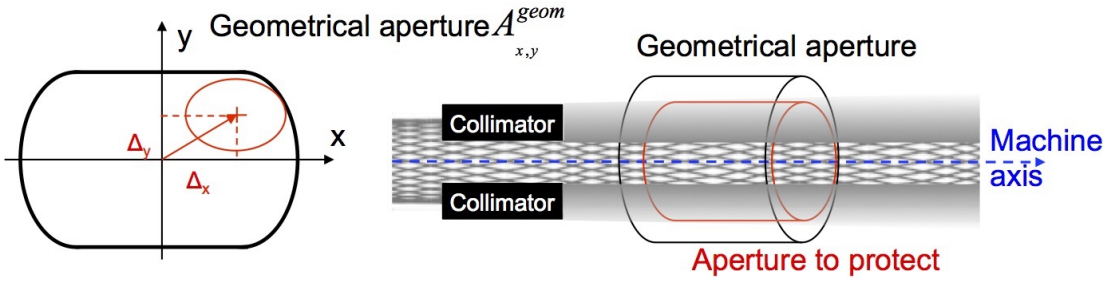


Fig. 3: Beam envelope with respect to machine's geometrical aperture. The injection protection collimators must shadow the aperture of the machine equipment, withstand direct beam impact, and reduce the downstream energy deposition to less than the damage limit.

where s is the longitudinal co-ordinate along a machine of length L and $\Delta_{x,y}(s)$ is the maximum beam transverse displacement:

$$\Delta_{x,y}(s) = CO^{\mathrm{peak}}_{x,y}\sqrt{\frac{\beta_{x,y}(s)}{\beta^{\mathrm{max}}_{x,y}}} + \left[\delta^{\mathrm{mech}}_{x,y}(s) + \delta^{\mathrm{align}}_{x,y}(s)\right] + k_\beta D_{x,y}(s)\delta_p + d^{\mathrm{sep}}_{x,y}(s) + d^{\mathrm{inj}}_{x,y}(s) + d^{\mathrm{axis}}_{x,y}(s)\,. \quad (5)$$

In Eq. 5, mechanical ($\delta^{\mathrm{mech}}_{x,y}$) and alignment ($\delta^{\mathrm{align}}_{x,y}$) tolerances are added to the peak ring closed orbit (CO_{peak}), which is normalized with respect to the local and the maximum β-functions. A dispersive term is also considered where δ_p is the sum of the beam momentum offset and the maximum momentum spread, k_β is the β-beating correction factor, and $D_{x,y}$ is the dispersion. The contribution of orbit displacements due to crossing and separation ($d^{\mathrm{sep}}_{x,y}$), injection oscillations ($d^{\mathrm{inj}}_{x,y}$), and offsets of the magnets with respect to the ring axis ($d^{\mathrm{axis}}_{x,y}$) are also taken into account [8]. As a general rule, the higher the beam stored energy, which is defined as the number of particles multiplied by the energy per particle, the larger the margin between the area occupied by the beam and $A^{\mathrm{geom}}_{x,y}$. Collimators should be set to intercept only miskicked or large-amplitude particles without interfering with the unperturbed injected and circulating beams. Special attention must be used for superconducting machines to limit the number of magnet quenches induced by beam losses.

The passive protection devices must withstand direct beam impacts and reduce energy deposition, owing to secondary showers, on the downstream elements to below the damage limit. The choice of the material of the absorbing blocks and the active length depend on the beam energy and intensity. Low-Z materials, like graphite, provide the best performance in terms of robustness but are quite transparent to secondary showers. Protection collimators must be set at well-defined longitudinal positions to provide a full coverage of the transverse phase space. As an example, a system of three collimators (each made of two 1 m long graphite jaws) per plane, located at 60° relative phase advance, is installed at the end of the transfer lines connecting the CERN Super Proton Synchrotron (SPS) to the LHC. This layout allows the required phase space coverage to be guaranteed while minimizing the number of collimators (Fig. 4) [9]. A >4 m long two-sided injection dump, composed of a sequence of low- and high-Z materials [9] to absorb secondary showers, is installed in the LHC injection region at 90° phase advance from the injection kicker. In the case of kicker failures, this dump cuts the particles with the largest amplitudes. The two-sided injection dump will be upgraded to cope with operation with high-brightness beams corresponding to 5 MJ stored energy per injection, which is at the limit of what materials can deal with, in terms of robustness and transmission. Material limitations might constrain the operation of future machines (e.g., the Future Circular Collider (FCC)), limiting the beam intensity that can be transferred at once. This will increase the number of injections required to fill the machine and, therefore, depending on the achievable rise time of the injection and extraction kickers, impact the total stored intensity [10, 11].

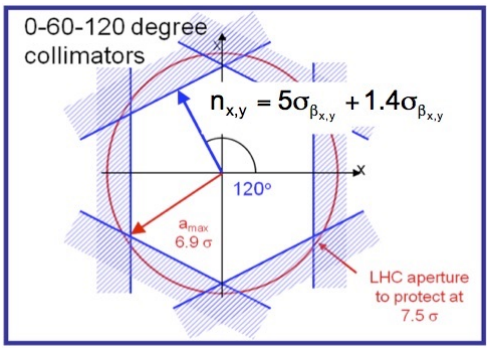


Fig. 4: Transverse phase space coverage provided by the collimators installed in the SPS-to-LHC transfer line to intercept large-amplitude oscillation particles. These collimators are set at $5\sigma_{\beta_{x,y}}$ to protect the $7.5\sigma_{\beta_{x,y}}$ LHC aperture, assuming a $1.4\sigma_{\beta_{x,y}}$ error in positioning due to alignment or optics errors.

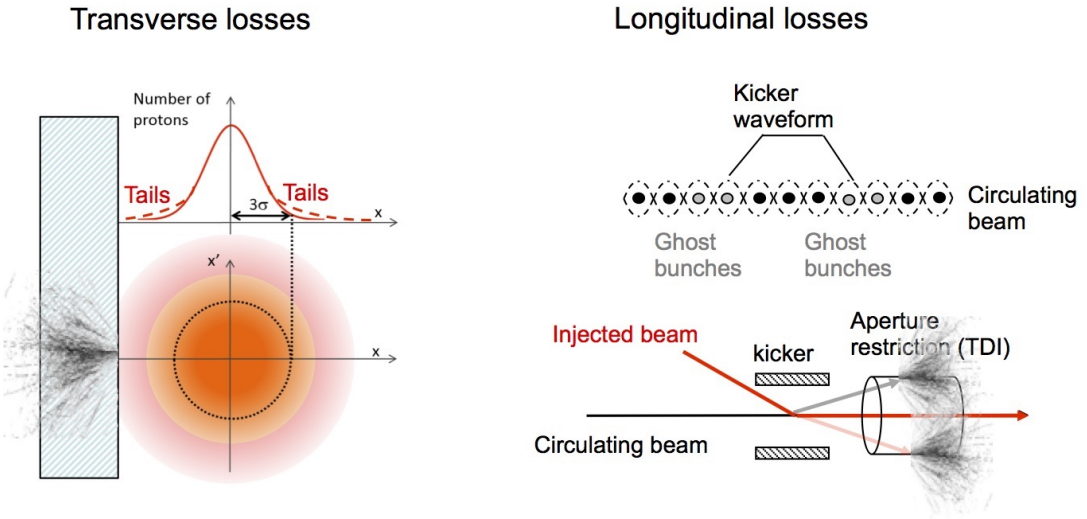


Fig. 5: Left: transverse losses mainly occur when largely populated transverse tails of particles are intercepted by aperture bottlenecks, ideally collimators. Right: longitudinal losses are due to ghost low-intensity bunches (grey), located near the main trains (black), which are miskicked by the injection kickers. TDI, two-sided injection dump.

4 Injection losses and mitigation

Injection losses are caused by some beam being intercepted at the machine aperture restrictions, which, ideally, correspond to the protection elements. These losses must be controlled and kept as small as possible to prevent damage and minimize the activation of the machine components. Beam losses can be divided into transverse [12] and longitudinal [13] losses (Fig. 5) and detected by means of conventional ionization chambers or diamond beam loss monitors, which are characterized by nanosecond resolutions. Hadron beams can be reasonably well approximated by a transverse Gaussian distribution of particles. The tails of this distribution are often overpopulated and extend well beyond the conventional 3σ. A non-negligible number of particles can then be intercepted by the protection collimators and create losses even during nominal operation without faults. These losses can be minimized through correct steering of the beam in the transfer lines and in the injection region. If this is not sufficient, the beam can be transversely scraped in the pre-injector [14] to remove the tails before the transfer in the next machine. In this case, losses are not suppressed but gathered in dedicated locations, allowing better monitoring and control.

Longitudinal losses occur when, as a consequence of a non-perfect RF capture, some low-intensity

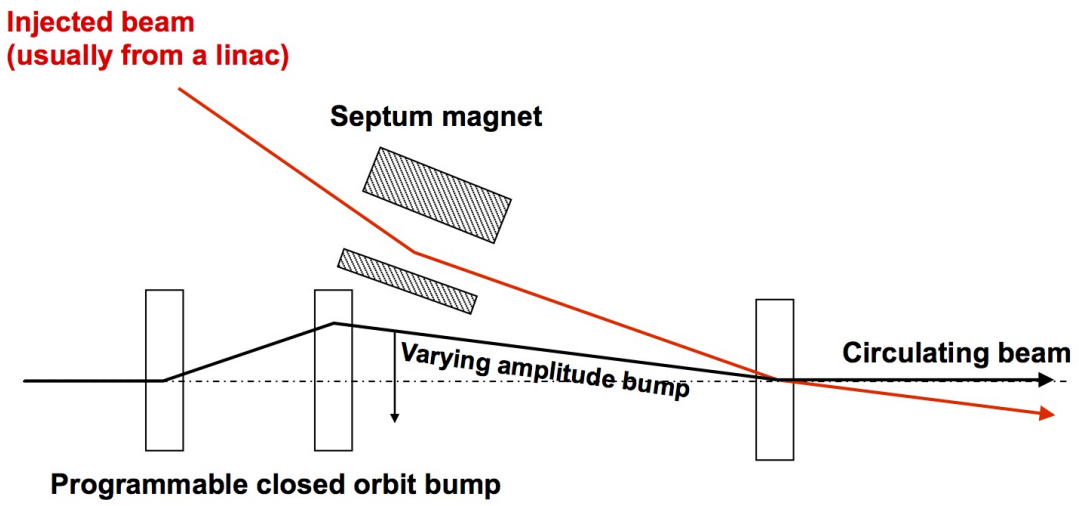


Fig. 6: System used for multiturn injection

ghost bunches populate the buckets surrounding the main trains. The ghost bunches can be either kicked by the rising or falling edge of the injection kicker magnetic field and lost at the next aperture bottleneck, or be injected on the closed orbit. In the second case, these bunches will be overkicked by the kicker at the next injection and finally lost. If no further tuning of the RF system is possible to mitigate this problem, the ghost bunches can be cleaned by exciting betatron oscillations (i.e., with a transverse damper) until they are lost in dedicated cleaning insertions of the machine.

5 Multiturn injection

Multiturn injection is a process that allows the injection of more bunches into a bucket. This technique is used for hadrons when the beam density at injection is limited by either space charge effects or the injector capacity. When the charge density from the injector cannot be increased, it is possible to fill the horizontal phase space and augment the overall intensity by injecting in the same bucket over several turns [15]. The condition that the acceptance of the receiving machine is larger than the delivered beam emittance must be fulfilled to be able to accumulate intensity via a multiturn injection process. Conventional multiturn injection is performed using a septum and a set of three or four bumpers which, at first, bring the orbit close to the septum blade and then decay in time to zero in a programmed way (Fig. 6). In particular, the bump amplitude is reduced, at each turn, such that the first beam occupies the central region and the later bunches the periphery of the transverse phase space acceptance (Fig. 7). Multiturn injection is often performed in the horizontal plane, owing to the usually larger acceptance. The final particle distribution is defined by the details of the multiturn painting process and by the filamentation, often space charge driven, which finally determines a quasi-uniform charge density. Owing to the finite thickness of the septum blade and the elliptical phase space contours of the injected beam, even in the absence of space charge, the resulting emittance in the receiving machine is

$$\varepsilon_x > 1.5 N \varepsilon_{\mathrm{i}} \,, \tag{6}$$

where ε_{i} is the emittance of the injected beam and N the number of turns over which the injection occurred. Multiturn injection is essential to accumulate high intensity at low energy, when space charge plays a dominant effect. Unfortunately it presents several disadvantages, mainly inherent to the use of the injection septum. The width of the septum blade reduces the available aperture; high losses (up to 30–40%) are produced by the circulating beam hitting the septum and, owing to the limited acceptance, the injection can be performed over a maximum of 10–20 turns.

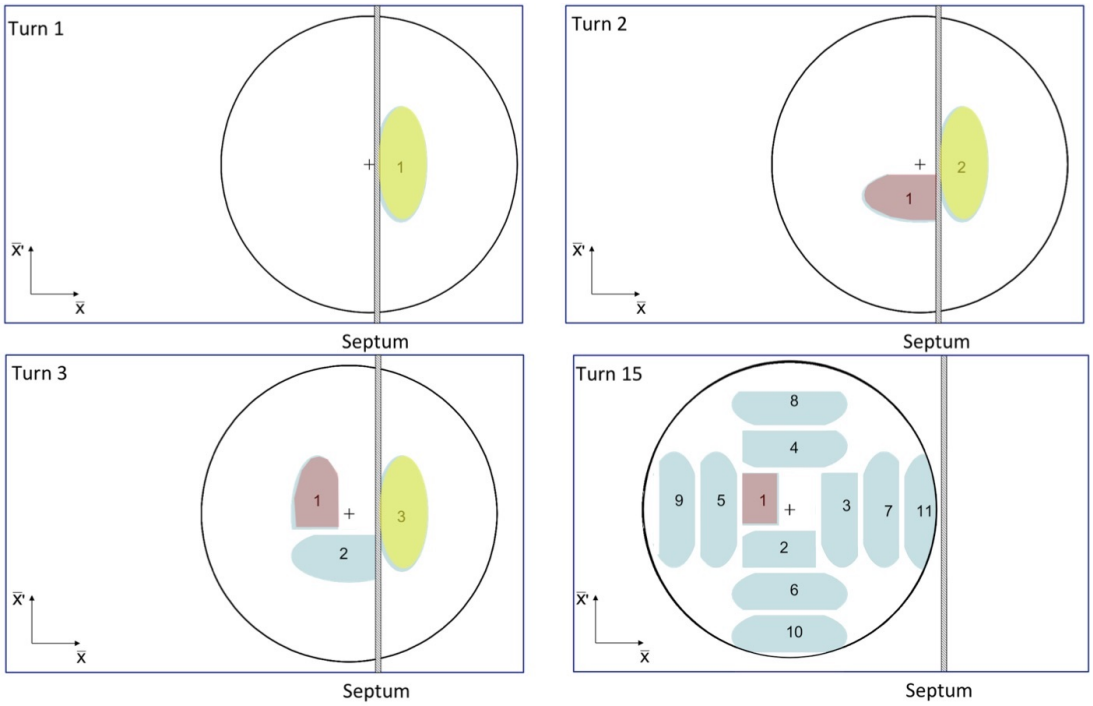


Fig. 7: Example of transverse painting in phase space over 15 turns, assuming a fractional tune $Q_{\mathrm{h}} = 0.25$

5.1 Charge exchange

Charge exchange injection provides an elegant alternative to the conventional multiturn injection process [16]. It involves converting H^- ions into protons by using a thin stripping foil, allowing injection in the same phase space area. A decaying magnet chicane is then used to modify the closed-orbit bump and paint the beam uniformly in the transverse phase space (Fig. 8). The thickness of the stripping foils must be carefully chosen to maximize the stripping efficiency, while minimizing emittance blow-up and losses. Typical foils are made of carbon and have a thickness of 50–200 $\mathrm{g/cm^2}$ for energies varying between 50 MeV and 800 MeV. The chicane is switched off at the end of the injection, to avoid excessive foil heating and further emittance blow-up. In some cases, longitudinal phase space painting can be performed by varying the energy of the injected beam turn-by-turn, by scaling the voltage of the linac accordingly [17]. A chopper system is then used to match the length of the injected batch to the bucket.

The absence of the septum and the nature of this method allow the injection losses to be reduced to a few per cent. Moreover the process can continue over tens, up to a hundred, of turns, allowing the accumulation of significant charge densities. Also, in this case, the final particle distribution is determined by the painting and the filamentation due to space charge.

6 Unconventional injection techniques

In the following, examples of unconventional single and multiturn injection techniques are briefly presented.

6.1 Slip stacking

Slip stacking is a configuration that is used to store and accelerate particle beams with different momenta in the same accelerator. This technique was demonstrated for the first time at the SPS to increase the production of anti-protons for p–$\bar{\mathrm{p}}$ physics [18] and later used at Fermilab for $\bar{\mathrm{p}}$ production and to double the power of the proton beam for the target, and, in particular neutrino physics [19]. It involves having two

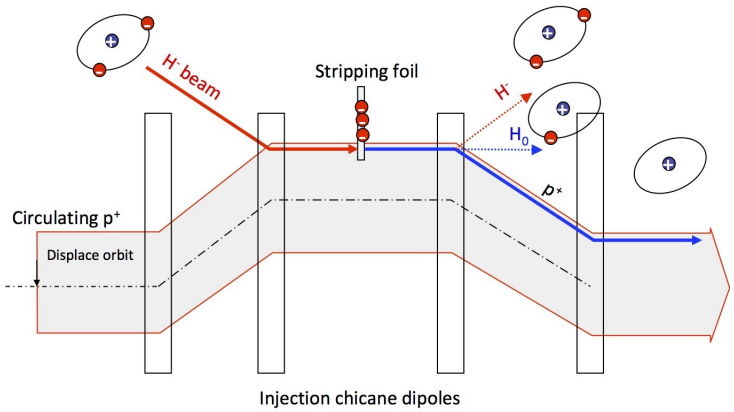


Fig. 8: The principle of the H^- charge exchange injection: a magnetic chicane, installed around a thin stripping foil, allows the beam to be painted in the transverse phase space and avoid excessive crossing, which might induce foil heating and emittance blow-up.

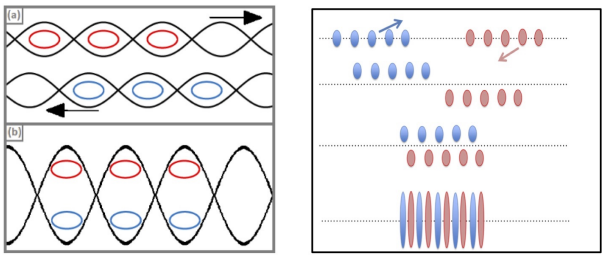


Fig. 9: Two sets of bunches slipping in opposite directions and then either (left) combined in the the same RF bucket once azimuthally overlapping or (right) interleaved at an average frequency.

beams (b_1 and b_2), which are bunched by two RF cavities with a small frequency difference $\Delta f = f_2 - f_1$. Each beam is synchronized to one RF cavity (i.e., b_1 to f_1 and b_2 to f_2) and, at the same time, perturbed by the other (i.e., b_1 by f_2 and b_2 by f_1). It is possible to define the so-called slip stacking parameter $\alpha = \Delta f / f_\mathrm{s}$, where f_s is the synchrotron frequency, which determines the level of perturbation felt by each beam due to the presence of the other RF system. When $\alpha \to 1$, the perturbation is large, while the condition $\alpha = 4$ corresponds to tangent boundaries for the two stationary buckets, that is, the lower limit for a stable motion. The two sets of bunches drift longitudinally with respect to each other and periodically coincide azimuthally, doubling the local line density (Fig. 9, left). These bunches can then be captured in a common larger bucket with an average frequency $(f_1 + f_2)/2$. The larger the bucket separation (i.e., large α) the bigger the final emittance due to recombination of the initial set of bunches. The critical point of this technique is to find the correct value of α, which allows enough separation, to limit the mutual perturbation, but, at the same time, minimizes the emittance blow-up.

Studies are ongoing to apply slip stacking in the SPS to increase the number of bunches for operation of the High Luminosity LHC (HL-LHC) with ions [20]. Two super-batches, made of 24 bunches separated by 100 ns, will be injected into the SPS and captured by two pairs of independently controlled 200 MHz cavities. The RF frequency will be varied to accelerate the first batch while decelerating the second one. The two batches will first be let slip and then brought back together by decelerating the first and accelerating the second. Once the bunches are interleaved, they will be recaptured at an average RF frequency (Fig. 9, right). Also, in this case, the final emittance will be blown up, owing to the filamentation of the bunch in the new RF bucket.

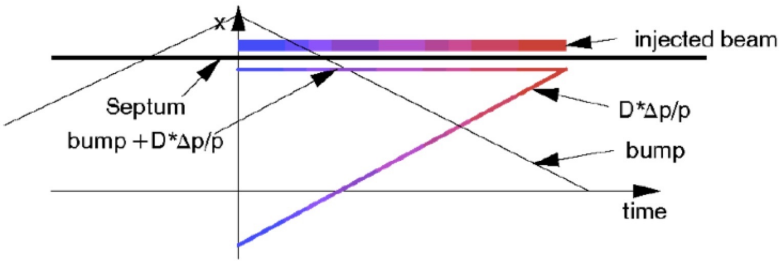


Fig. 10: Horizontal position of injected and circulating beam during injection process. The combination of the decaying bump and the displacement defined as $D\Delta p/p$ determines an instantaneous closed orbit.

6.2 Combined longitudinal and transverse multiturn injection

An accumulation technique that combines multiturn injection and stacking by electron cooling is used in the CERN Low-Energy Ion Ring to increase the number of injected particles [21]. As during conventional multiturn injection, a decreasing closed-orbit bump is used. Injection occurs in a section of the machine characterized by a large normalized dispersion and the energy in the linac is ramped. The bump decrease and the energy ramp are adjusted, such that the instantaneous closed orbit, given by the sum of the bump and the displacement, defined as $D\Delta p/p$, stays constant at injection (Fig. 10). In this way, both the transverse and longitudinal phase space are exploited and injection is performed, keeping the same betatron amplitude but stacking in momentum. Additional stacking in vertical phase space can be performed by using a combination of a magnetic and an inclined electrostatic septum [3]. This implies a complicated geometry of the injection line and additional constraints on the working point to avoid touching the septum blade at each turn. Also, in this case, the injection bump is applied only in the horizontal plane.

The injection of one full pulse from the linac takes approximately 200 µs, corresponding to 76 turns. In total, seven injections are needed to fill the machine and one pulse is injected every 200 ms, to allow for electron cooling, which takes about 100–150 ms. After each injection, the momentum offset is reduced, through cooling, from 0.4% down to 0.1%; this permits the orbit of the stacked beam to be moved and leaves room for the injection of the next pulse from the linac. Efficiencies of the order of 50–70% are expected with this method.

7 Main lessons

Injections of low-energy beams are dominated by space charge, which makes the storage of high-intensity and high-brightness beams challenging. Conventional multiturn injection with phase space painting allows this limit to be partially overcome but implies a non-negligible emittance growth and large losses at the septum. Moreover, the process is limited to 10–20 turns, owing to the finite acceptance of the receiving machine. Charge exchange H^- injection is an elegant alternative, which permits these limitations to be overcome and high-brightness beams to be produced.

The fast processes involved in the transfer of high-energy beams define a major concern in terms of machine protection. A proper design of the injection insertions and a careful handling of these potentially harmful beams is fundamental. Moreover, the preservation of a small emittance is crucial to maximize the quality of the beam and the machine performance (e.g., the luminosity in a collider). Mismatch and injection errors must be minimized to reduce losses, injection oscillations, and emittance blow-up. Passive protection elements must be installed at strategic locations to reduce the energy deposition on sensitive machine components and avoid damage while limiting activation.

Acknowledgements

Special thanks to M.J. Barnes, W. Bartmann, M. Fraser, B. Goddard, and V. Kain for their inputs, the fruitful exchange of information, and stimulating discussions.

References

[1] D.A. Edwards and M.J. Syphere, in *An Introduction to the Physics of High Energy Accelerators*, (Wiley, New York, 1993), Chap. 6, pp. 172–186, https://doi.org/10.1002/9783527617272.

[2] F. Tecker, Review of longitudinal beam dynamics, these proceedings.

[3] M. Paraliev, Septa I and Septa II, these proceedings.

[4] M.J. Barnes, Kicker systems. Part 1: introduction and hardware, these proceedings.

[5] W. Bartmann, Transfer line design, matching, and design tools I, these proceedings.

[6] V. Kain, Emittance, these proceedings.

[7] W. Bartmann, Transfer line design, matching, and design tools II, these proceedings.

[8] J.B. Jeanneret and R. Ostojic, Geometrical acceptance in LHC Version 5.0, LHC project note 111, 1997.

[9] V. Kain, Machine protection and beam quality during the LHC injection process, Ph.D. thesis, Technische Universität Wien, 2005, CERN-THESIS-2005-047.

[10] W. Bartmann *et al.*, Beam transfer to the FCC-hh collider from a 3.3 TeV booster in the LHC tunnel, Proc. IPAC2015, Richmond, VA, USA.

[11] M. Barnes, Kicker systems part 2, hardware: existing and future, these proceedings.

[12] W. Bartmann *et al.*, LHC injection and extraction losses, Proc. Evian Workshop, 2010.

[13] O. Stein *et al.*, Investigation of injection losses at the Large Hadron Collider with diamond based particle detectors, Proc. IPAC2016, Busan, 2016.

[14] L. Drösdal *et al.*, SPS transverse beam scraping and LHC injection losses, Proc. IPAC 2012, New Orleans, LA, 2012.

[15] C. Prior, Ring injection for high intensity accelerators, CAS on High Power Hadron Machines, Bilbao, 2011.

[16] C. Bracco, Phase space painting and H^- stripping injection, these proceedings.

[17] V. Forte, Performance of the CERN PSB at 160 MeV with H^- charge exchange injection, Ph.D. thesis, CERN and Université Blaise Pascal, 2016.

[18] D. Boussard and Y. Mizumachi, *IEEE Trans. Nucl. Sci.* **NS-26** (1979) 3623, https://doi.org/10.1109/TNS.1979.4330122.

[19] J. Eldred and R. Zwaska, Slip-stacking dynamics and the 20 Hz booster, Fermilab-conf-14-565-APC, Batavia, IL, 2015.

[20] T. Argyroupoulos, Slip stacking in the SPS (mode of operation with beam, simulations, scheduled tests and implementation), LIU Day, 2014.

[21] C. Carli *et al.*, Combined longitudinal and transverse multi-turn injection in a heavy ion accumulator, Proc. PAC 1997, Vancouver, 1997.

Proceedings of the CAS–CERN Accelerator School: Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018)

Phase Space Painting and H^- Stripping Injection

C. Bracco
CERN, Geneva, Switzerland

Abstract
The achievable hadron beam density at injection can be limited by space charge or by the capacity of the injectors. Conventional multiturn injection allows the total intensity to be augmented but also determines a linear increase of the beam emittance with the number of injected turns and is characterized by high losses at the septum. These limits can be overcome with the H^- charge exchange injection technique and high-power beams can be produced. Combination with transverse or longitudinal phase space painting allows for control of the size and distribution of the injected particles and minimizes the number of foil hits for the circulating beam. In the following, the H^- charge exchange technique is described in detail, together with the related hardware, main achievements, and complications.

Keywords
Multiturn injection; charge exchange; high power; stripping foil; phase space painting.

1 Introduction

As explained in Ref. [1], the beam density in hadron storage rings can be limited at injection by space charge or by the injector capacity. A way to overcome this limitation and attain higher bunch current is to accumulate particles by injecting several beam pulses, generally from a linac, into the same RF bucket. This method is called multiturn injection and can be applied when the acceptance of the receiving machine is larger than the delivered beam emittance. The injected and circulating beam cannot be overlapped in phase space. Multiturn injection is then often also described as 'phase space painting', since the injected bunches are positioned in different spots of the transverse or longitudinal phase space of the ring to generate the desired beam distribution. Conventional multiturn injection employs a septum with a programmed orbit bump, which brings the circulating beam close to the septum blade. The closed orbit is then adiabatically (i.e., slowly with respect to the synchrotron oscillations) moved away from the septum so that the early beam occupies the central region and the later beam the outer part of the acceptance. The main disadvantages of this technique consist in high beam losses due to the circulating beam hitting the septum and the small number of injected turns due to the limited acceptance. The achievable brightness is then restrained by the total accumulated intensity and the fact that the final emittance increases linearly with the number of injection turns.

In addition to conventional multiturn schemes, there are several more exotic injection techniques, devised to control and improve the properties of the stored beam. Charge exchange injection allows the injected (H^-) and circulating (p^+) beam to be overlapped in the same phase space area and then provides the only way to achieve low-loss multiturn injections (from 10% in the best case with conventional multiturn injection down to 0.02% for charge exchange injection) and stack many small bunches over several turns without a linear increase of the emittance, e.g., allowing the production of high-density beams.

2 H^- charge exchange injection

Charge exchange injection was applied for the first time at the end of the 1960s; examples of pioneering experiments can be found in Refs. [2–4]. The roadmap of the achieved and target beam powers for the

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-005.141

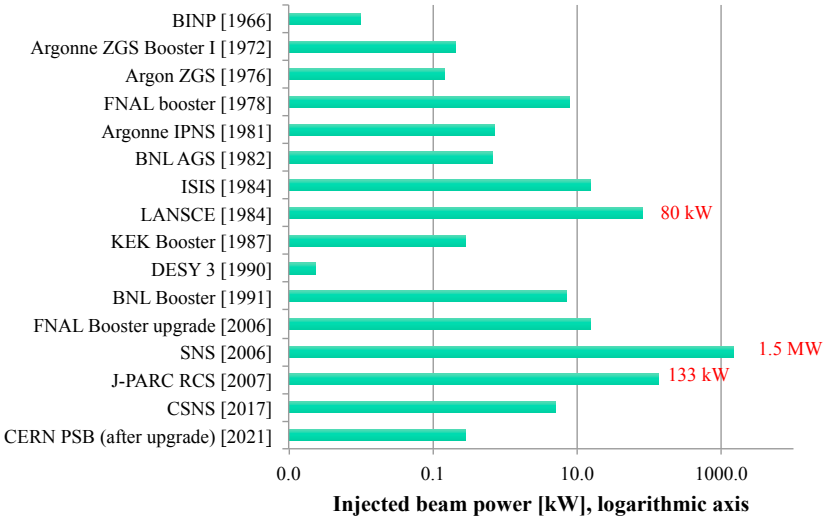


Fig. 1: Roadmap of the achieved and goal beam power for the main facilities using H^- charge exchange injection

main facilities that use H^- charge exchange injection is shown in Fig. 1. In particular, the Spallation Neutron Source (SNS) at the Oak Ridge National Laboratory leads, with 1.5 MW power beams.

In general, the charge exchange is performed by stripping two electrons, by means of thin stripping foils, from an H^- ion and transforming it into a proton (p^+). The closed orbit of the circulating beam is bumped onto the injection orbit using a set of chicane magnets located around the stripping foil, as shown in Fig. 2. The fully stripped p^+ follows the closed orbit while the partially (H^0) and unstripped (H^- ions are dismissed. Additional fast bumpers, installed in the ring lattice, are used to perform transverse phase space painting by changing the closed orbit, in a programmed way, during the injection process. This allows the size and the distribution of the beam to be controlled. The final size of the beam is defined by the distance between the injected H^- beam and the closed orbit at the location of the foil: the closer the foil, the smaller the beam. Ideally, in the absence of space charge, perfectly hollow, very dense, or smooth and uniform distributions could be obtained. In reality, the final particle distribution is determined by both the painting and the filamentation due to space charge. The fact of being able to overlap the injected and circulating beams in phase space and the space charge mitigation obtained by painting the beam in an almost uniform way allow much higher densities to be achieved than with conventional multiturn injection. The closed-orbit bump decay is also important to minimize the foil hits, which could determine foil heating, losses, or emittance blow-up due to scattering. Longitudinal painting can also be performed by modulating the energy and the energy spread of the injected beam to further reduce the charge density.

In the following, two examples of H^- charge exchange injection systems are described: those of the SNS and the future CERN Proton Synchrotron Booster (PSB).

2.1 SNS injection system

The SNS accelerator complex comprises an H^- source, a 1 GeV linac, an accumulator ring, and a liquid mercury target for spallation neutron production [5]. The accumulator ring delivers 0.7 μs pulses of $\sim 2\times10^{14}$ particles and delivers them at a rate of 60 Hz to the target. Owing to the high power of the beam, care must be taken to ensure hands-on maintainability of the ring components by minimizing uncontrolled losses of the proton beam. The H^- beam is transferred from the linac, stripped by the foil, and then injected into the ring. A fixed horizontal chicane is installed around the stripping foil and assures adequate clearance during the full injection process. Eight symmetrically placed horizontal (four) and vertical (four) dynamic kickers are installed, in pairs, immediately upstream and downstream of the

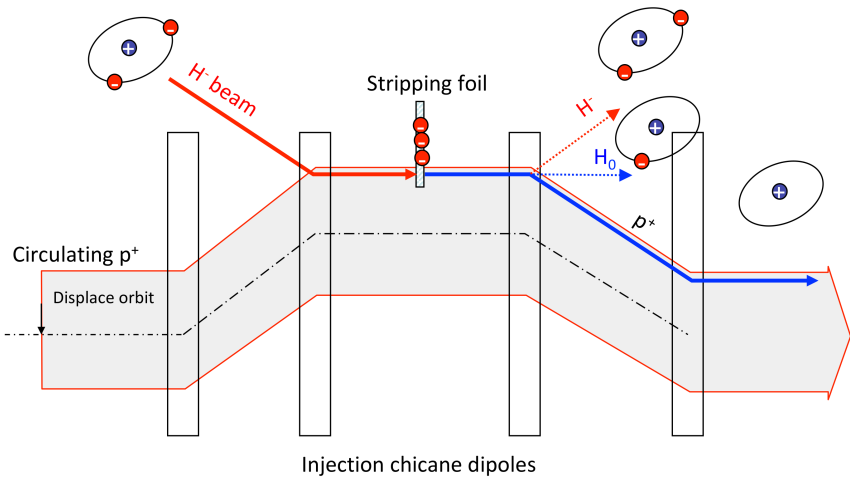


Fig. 2: Magnetic chicane installed around stripping foil. Varying the amplitude of the orbit bump during the injection process enables transverse phase space painting while avoiding excessive crossings, which might induce losses, foil heating, or emittance blow-up.

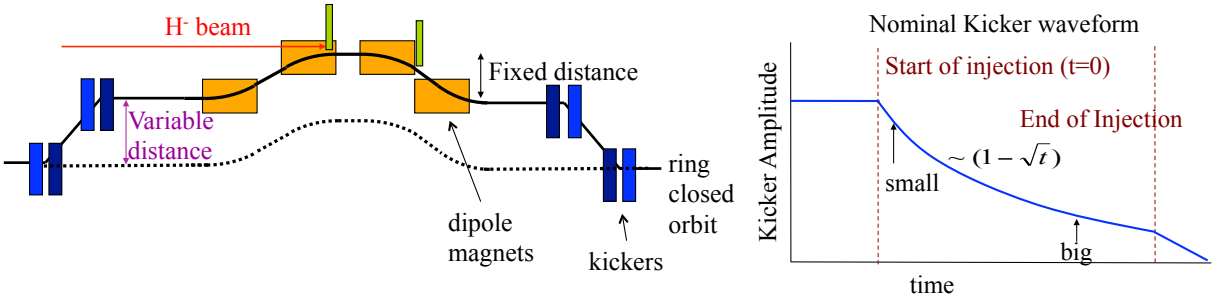


Fig. 3: Layout of the H$^-$ charge exchange injection system used at SNS (left). Four dipoles (orange) are used to create the fixed injection chicane. Four horizontal and four vertical fast bumpers (light and dark blue) are powered with a current decaying in time proportionally to $1 - \sqrt{t}$ (right) and provide a correlated transverse painting in both planes.

chicane, as shown in the left panel of Fig. 3. They provide the variable amplitude bump (Fig. 3, right) to perform the transverse phase space painting, in both planes, which gives the desired transverse density and beam profile at the target. The painting allows an average of eight foil hits per particle to be achieved during the full accumulation (>1000 turns). Longitudinal painting is also performed, by manipulating both the energy and the energy spread, to fill the phase space as uniformly as possible and reduce the longitudinal halo [6]. The SNS system separates transverse and longitudinal phase space, since injection occurs in a dispersion-free section of the ring.

2.2 PSB injection system

The PSB at CERN has to provide beams with different emittance and intensity to several users. The goal is to produce, after the upgrade, which will occur during the 2019–2021 long shutdown [7], high-brightness beams for the LHC (1.7×10^{12} p$^+$ in <1 µm r.m.s. normalized emittance) and high-intensity beams for fixed-target experiments (up to 2×10^{13} p$^+$ in 8×13 µm r.m.s. normalized vertical and horizontal emittance respectively). Linac2, which is currently injecting 50 MeV p$^+$ into the PSB, will be replaced with Linac4, with the goal of providing 600 µs pulses of up to 5×10^{13} 160 MeV H$^-$ ions. The PSB consists of four superimposed rings and, for each ring, the present conventional multiturn injection

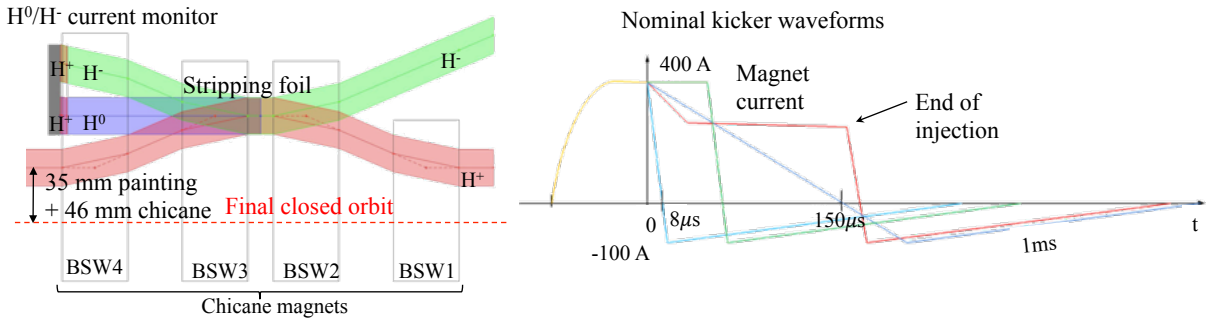


Fig. 4: Layout of the future H^- charge exchange injection system for the PSB (left). In this case, the injection bump is defined, only in the horizontal plane, by the chicane magnets and four kickers installed in the ring lattice. Different painting schemes can be obtained, by variable modulations of the current of the kickers, for the various users (right).

system will be replaced with charge exchange injection. This will contain a horizontal fixed chicane and a stripping foil (Fig. 4, left). A current monitor will be installed in front of the injection dump to measure the stripping efficiency of the foil. Four horizontal kickers, installed in the ring lattice, will be used for the transverse phase space painting [8]. No painting will be applied in the vertical plane but the injected beam will be steered, in the transfer line, to a fixed vertical offset. Different waveforms will be used for the horizontal painting, according to the requirements of the various users (Fig. 4, right). The bump will stay constant for small-emittance beams, while a multilinear decay will be applied for the fixed-target experiments. An initially fast decay must be followed by an almost constant slope to fill first the centre and then the outer area of the phase space, reducing the charge density in the core and thus the space charge effects. The accumulation will be completed in a maximum of 150 turns. Once injection is finished, the circulating beam must be moved away from the foil, as quickly as possible. A negative bump is implemented to limit the interactions with the foil during the slow decay of the injection chicane. Longitudinal painting will be performed only for the high-intensity beams. The dispersion is non-zero in the PSB injection region; the longitudinal painting affects the transverse painting and constrains the minimum size of the stripping foils [9].

3 Stripping foils

Stripping foils are the best technology available today for charge exchange injection into storage rings and synchrotrons. Standard foils are commercially available but, in some cases, as for SNS, owing to the extremely high power of the handled beams, special designs are needed, which require in-house fabrication. The foils must be at the same time thick enough to ideally strip off all the electrons from the H^- ions but also thin enough to minimize the effect on the circulating beam (i.e., minimize scattering and energy loss). Foils can sometimes get very hot so they should be made of materials with a high melting point. When foils can be supported only from one side, to avoid any extra material that might be intercepted by the beam, they must be self-supporting. Amorphous or diamond carbon, with a density range of ρ = 1.7–2 g/cm^2, provides a good material for stripping foils, owing to its thermal stability, high sublimation temperature, and radiation and mechanical resistance. The stripping efficiency of a carbon foil depends on the energy of the H^- ions and the foil thickness. The absolute yield of H^-, H^0, and p^+ can be calculated as [10]:

$$H^-(x) = \exp[-n(\sigma_{-0} + \sigma_{-+})x]; \tag{1}$$

$$H^0(x) = \frac{\sigma_{-0}}{\sigma_{-0} + \sigma_{-+} - \sigma_{0+}} \left\{ \exp(-n\sigma_{0+}x) - \exp[-n(\sigma_{-0} + \sigma_{-+})x] \right\}; \tag{2}$$

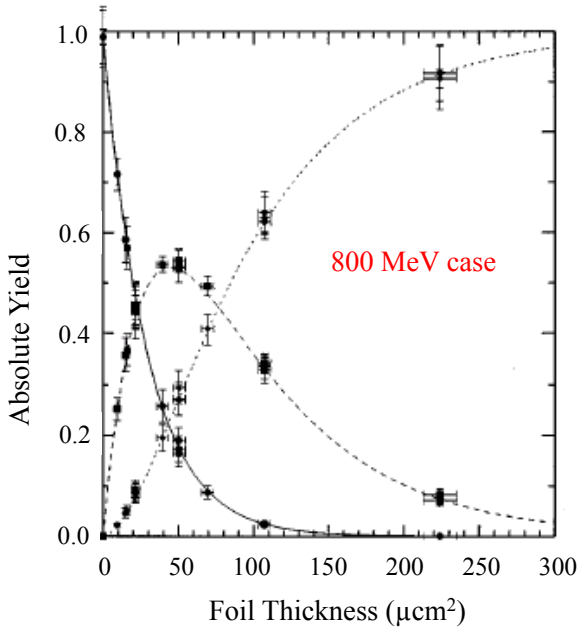


Fig. 5: Absolute yield of H^-, H^0, and p^+ as a function of the carbon foil thickness for an 800 MeV beam [10]

$$H^-(x) = 1 - H^-(x) - H^0(x)\ . \tag{3}$$

Here, x is the foil thickness and σ_{-0} and σ_{-+} are the cross-sections for one- and two-electron stripping of H^- respectively, σ_{0+} is the cross-section of one-electron stripping for H^0 and $n = \rho N_0/A$ is the foil density in atoms per unit volume, N_0 being Avogadro's number and A the atomic mass number of the foil atoms. The relative curves, for an 800 MeV beam, are shown in Fig. 5. In the PSB, owing to the lower energy, 200 μg/cm^2 thick foil should guarantee that ∼100% of the electrons are stripped. Foils with a thickness of 300 μg/cm^2 are used at SNS and provide ∼97% stripping efficiency for the 1 GeV beam. Despite the lower stripping efficiency, thicker foils are not used, since they would increase the foil temperature and the radio-activation. Both inelastic and elastic nuclear scattering processes are assumed to result in uncontrolled loss of the impacting protons. The probability of scattering per foil traversal is:

$$P = \left(\frac{2Zm_{\mathrm{e}}r_{\mathrm{e}}}{\gamma M\beta^2}\right)^2 N_0\left(\frac{\rho t}{A}\right)\left[\frac{1}{\theta_{xl}\theta_{yl}} + \frac{1}{\theta_{xl}^2}\tan^{-1}\left(\frac{\theta_{yl}}{\theta_{xl}}\right) + \frac{1}{\theta_{yl}^2}\tan^{-1}\left(\frac{\theta_{xl}}{\theta_{yl}}\right)\right]; \tag{4}$$

where Z is the atomic number, t the foil thickness, m_{e} and r_{e} the electron mass and classical radius, respectively, and M the rest mass of the incident particles. In this formula, θ_{xl}^2 and θ_{yl}^2 define the limiting angles above which a scattered particle is lost and can be written as:

$$\theta_{xl}^2 = \left(\frac{X_A}{\beta_{fx}}\right) \tag{5}$$

and

$$\theta_{yl}^2 = \left(\frac{Y_A}{\beta_{fy}}\right). \tag{6}$$

X_A and Y_A are the horizontal and vertical acceptance of the machine and $\beta_{fx,fy}$ the β-functions at the foil. For very thin foils, losses due to single large-angle scattering have a significant probability (much higher than from a Gaussian distribution).

Beside the losses, thick foils might induce unwanted emittance blow-up through multiple-Coulomb scattering (mainly an issue for high-brightness beams, such as those at LHC) and foil heating

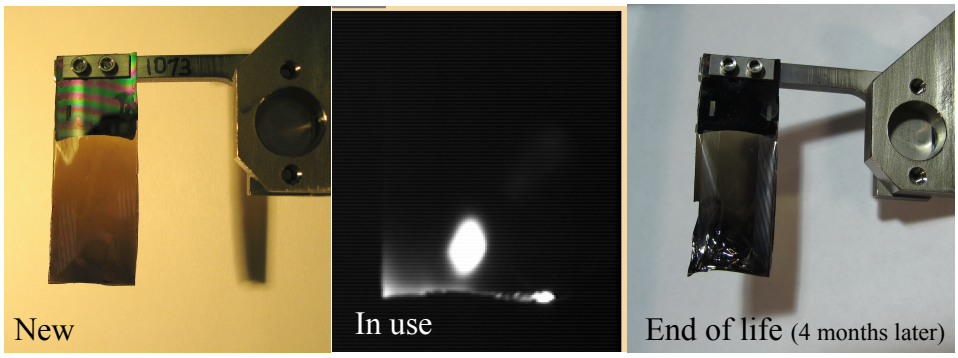


Fig. 6: Nanocrystalline diamond foil used at SNS when new, during, and after 4 months of operation

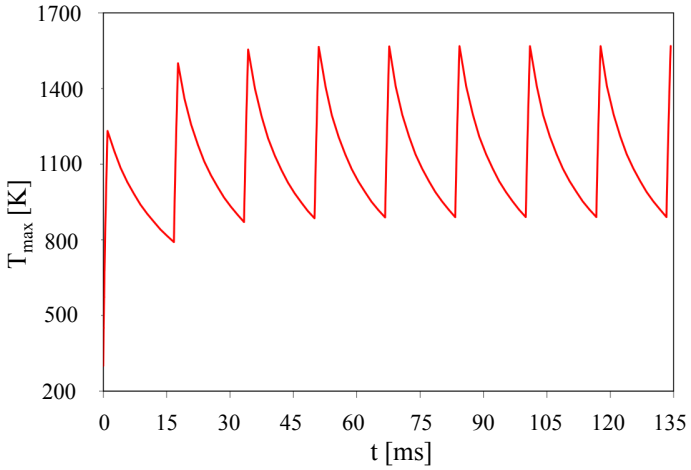


Fig. 7: Calculated temperature fluctuations at the SNS foil when operating with a 1 GeV, 1.5 MW beam

via ionization and electron excitation energy loss of the impacting protons and stripped electrons [10]. High temperature and thermal stress constitute two of the biggest issues for the foil lifetime and pose limits to the maximum achievable beam power. Figure 6 shows images of a nanocrystalline diamond foil, new, during, and after four months of operation in the SNS facility. It is calculated that temperature fluctuations of the order of 600 K can be achieved when operating with the 1.5 MW SNS beam (Fig. 7). In reality, foil temperatures are very hard to model, calculate, or measure. In fact, emissivity, heat capacity, and thermal conductivity are not well known at these high temperatures. The energy deposition is not well known either. The exact number of foil hits per injected proton, the deposited energy when H^- particles break up while passing through the foil, and effects such as knock-out electrons (delta rays) caused by relativistic p^+ striking the foil need to be known to correctly evaluate the total energy deposition. Measurement techniques are under development but are also challenging, since the heating process is very fast (the peak temperature only lasts for tens of microseconds). It was possible to observe, while operating at 800 kW, that the SNS foil was turning white, indicating a temperature $>$1700 K, i.e., higher than calculations. Sublimation is one of the biggest risks for stripping foils. At 2200 K, carbon sublimates at a rate of about 1 μm/h and foils have a typical thickness of 1–2 μm. The sublimation rate is a strong function of the temperature and small increases determine a dramatic change in the foil lifetime (four orders of magnitude for a 300 K temperature increase).

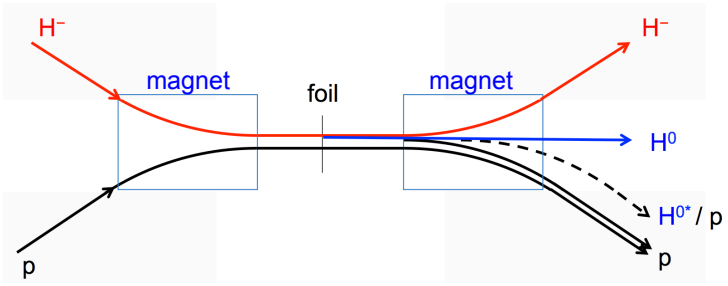


Fig. 8: Possible trajectories of H^{0*} atoms when then decay into p^+, owing to the effect of the magnetic field. If their trajectory is outside of the ring acceptance, these particles are lost [11].

4 Beam losses caused by H^0 excited states

The neutral hydrogen atoms emerging from the foil are in a distribution of excited quantum states H^{0*} according to the $n^{-2.8}$ law where $n = 1, 2, 3, \ldots$ is the principle quantum number of the H^0 atoms. When they pass through a magnetic field, for example of the injection chicane dipoles, they see an electric field due to the relativistic transformation $E = \gamma\beta cB_{\text{lab}}$ (Fig. 8). This electric field can strip off the electron and, if the newly created proton is outside of the acceptance of the machine, it will create beam losses. This phenomenon was observed for the first time in 1993 by Hutson and Macek at the Los Alamos Proton Storage Ring (PSR) [12] and causes, at present, 15–20% of the total beam loss, corresponding to 23–40 W. Such a loss level would be catastrophic for the SNS, since it would correspond to 2.85 kW. The SNS design was made taking this process into account; the stripping foil is located in the falling fringe field of a magnet. In this way, high n states ($n \geq 6$) are short-lived and are Lorentz-stripped immediately after the foil and captured into the ring acceptance, along with the fully stripped protons. The low n states ($n \leq 3$) are long-lived; they can be transported along with the ground state H^0 into the injection dump and are controlled losses. The intermediate, $n = 4$ and $n = 5$, states have the potential of decaying, in the magnetic field, far enough downstream from the injection foil that their resultant deflection puts them on trajectories that do not lie within the ring acceptance. About 0.01% of them are lost. For these considerations, the choice of foil thickness should also take into account the H^0 excited states [10].

5 Control of 'convoy' electrons

Positioning the foil inside the fringe field of a magnet also has the advantage of directing the stripped, also called 'convoy', electrons to a dedicated electron catcher (Fig. 9) [11]. At J-PARC the convoy electrons carry a design power of about 145 W; these are directed to a water-cooled block of graphite by the fringe field of one of the chicane dipole magnets. At the SNS, these particles carry 1.6 kW and hence must be properly controlled to avoid damaging the equipment surrounding the foil. A water-cooled carbon–carbon electron collector is installed at the bottom of the vacuum chamber to intercept and trap these electrons. In reality, owing to a combination of fabrication errors and modifications of the foil mount position, in the original design some electrons were reflected back up and hit the foil and the bracket, which, after 3 months of operation, was damaged (Fig. 10). Evidence of softening, sag, and sputtering of the titanium bracket were observed. A new design was proposed, with brackets made of titanium-zirconium-molybdenum alloy. This alloy has a high sublimation temperature, low sputtering yield, and high sputtering threshold; no damage was observed after 16 days of operation with 1.3–1.4 MW beams. The main disadvantage of this material is a higher residual radioactivity. The electron catcher also showed signs of damage, confirming the hazard of the convoy electrons; a new upgraded design is presently under development.

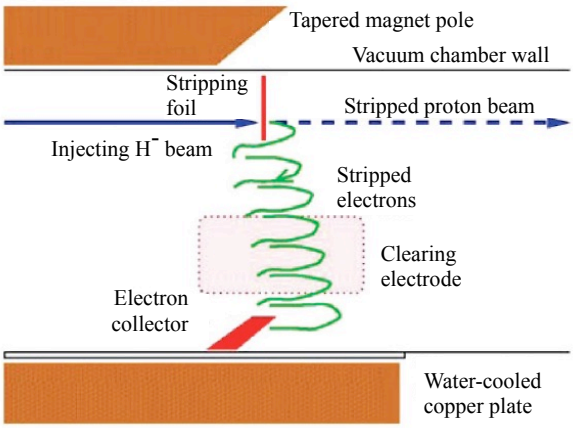


Fig. 9: The SNS stripping foil is installed inside one of the injection chicane magnets to drive the convoy electrons towards an electron catcher [13].

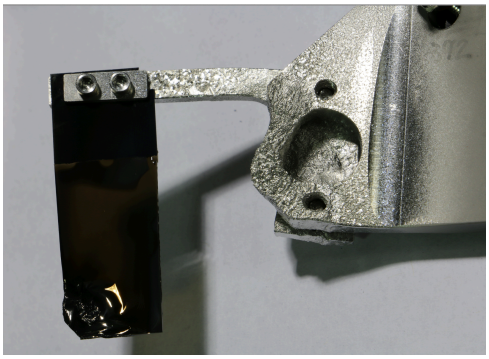


Fig. 10: Foil and bracket damaged by convoy electrons reflected back up from the catcher after 3 months of operation at 1.1–1.4 MW [11].

6 Main lessons

For hadrons, the beam density at injection can be limited by space charge or by the injector capacity. Transverse phase space painting allows the total stored intensity to be augmented but the final emittance increases linearly with the number of injected turns and high losses occur at the septum unless charge exchange injection is used. The combination of charge exchange injection and transverse phase space painting enables control of the size and distribution of the injected particles and the production of high-intensity and high-brightness beams while minimizing the foil hits for the circulating beam. Longitudinal painting, performed by modulating the energy during the injection, permits further reduction of the charge density and thus the space charge effects.

Stripping foils are the best technology available today for charge exchange injection. They need to be thick enough to maximize the stripping efficiency but, at the same time, thin enough to minimize losses and energy loss through scattering processes. The robustness and lifetime of the foils determine the main limitation to the presently achievable beam power. Other complications include need of a careful handling of the stripped electrons and H^0 excited states to avoid damage of the equipment surrounding the stripping foil.

7 Acknowledgements

A special thanks to M.A. Plum, who kindly provided the majority of the material presented in this proceedings.

References

[1] C. Bracco, Injection: hadron beams, these proceedings.

[2] G.I. Budker *et al.*, *Sov. At. Energy* **22** (1966) 441, https://doi.org/10.1007/BF01175205.

[3] J.D. Simpson, Operating results from the ANL booster, 5th IEEE Particle Accelerator Conf., San Francisco, CA, 1973, p.198, https://doi.org/10.1109/TNS.1973.4327079l.

[4] A.J. Gorka, A fast-spinning stripper mod II for the zero gradient synchrotron (ZGS) booster, 5th IEEE Particle Accelerator Conf., San Francisco, CA, 1973, p.387, https://doi.org/10.1109/TNS.1973.4327126.

[5] J. Wei *et al.*, *Phys. Rev. Accel. Beams* **3** (2000) 080101, https://doi.org/10.1103/PhysRevSTAB.3.080101.

[6] D. Raparia *et al.*, Linac energy manipulation for longitudinal phase space painting for SNS ring, XXI International Linac Conf. (Linac 2002), Gyeongju, 2002.

[7] M. Meddahi *et al.*, LHC injectors upgrade, technical design report—volume I: protons, CERN-ACC-2014-0337, (CERN, Geneva, 2014).

[8] J.L. Abelleira *et al.*, Painting schemes for CERN PS Booster H^- injection, Proc. IPAC 2015, Richmond, VA, 2015.

[9] M. Aiba *et al.*, Stripping foils for the PSB injection system, CERN-sLHC-Project-Note-0005, (CERN, Geneva, 2009).

[10] M.S. Gulley *et al.*, *Phys. Rev. A* **53** (1996) 3201, https://doi.org/10.1103/PhysRevA.53.3201.

[11] M.A. Plum, H^- charge exchange injection issues at high power, Proc. HB2016, Malmö, 2016.

[12] J. Donahue *et al.*, Measurement of H^0 excited states produced by foil stripping of 800-MeV H^- ions, Proc. PAC 93, Washington DC, 1993.

[13] L. Wang *et al.*, *Phys. Rev. Accel. Beams* **8** (2005) 094201, https://doi.org/10.1103/PhysRevSTAB.8.094201.

Proceedings of the CAS–CERN Accelerator School: Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018)

Injection and Extraction in Cyclotrons

M. Seidel
PSI, Villigen, Switzerland

Abstract
After an introduction to the related basics of cyclotrons, this article describes injection and extraction schemes in cyclotrons. Often these represent the critical aspects of cyclotron designs. For injection, we discuss internal sources, electrostatic inflectors, horizontal injection and matching. For extraction, we cover the scaling of turn separation, electrostatic septa and charge-exchange extraction. Low loss extraction of high-intensity beams is discussed in particular detail.

Keywords
Cyclotron; injection; extraction.

1 Introduction

Cyclotrons have a long history in accelerator physics and are used for a wide range of medical, industrial, and research applications [1]. The first cyclotrons were designed and built by Lawrence and Livingston back in 1931 [2]. The cyclotron represents a resonant accelerator concept with several properties that make it well suited for cost-effective and compact accelerator installations, but also for the acceleration of hadron beams with high average intensity. It is a powerful concept for the following reasons. The circulation time of the beam can be kept constant by arranging the scaling of bending radius and velocity in proportion. Thus neither frequency nor magnetic field must be varied in the process of acceleration and the beam can be accelerated continuously. A simple-shaped vertically oriented magnetic field plus a pair of electrodes for acceleration by a radiofrequency (RF) field are sufficient to construct a classical cyclotron. The acceleration device can be used repetitively, thus presenting a cost-effective solution. The most significant limitations of cyclotrons are the limited energy range and the relatively weak focusing, which imposes a limit for the bunch charge. While the acceleration process can be realized in a straightforward way in cyclotrons, clean injection and extraction into and out of this continuous accelerating scheme are challenging.

In the series of the CERN accelerator school CAS, the topic of injection/extraction for cyclotrons have been considered already several times in depth [3–5]. This paper touches on all relevant techniques, but focuses a bit more on specific issues with high-intensity beams. Most of the cited papers are available in electronic format on the internet, and related links are given in the references section.

2 Related cyclotron basics and magnetic versus electric deflection

The revolution frequency of an ion beam in a classical cyclotron is given by

$$f_\mathrm{c} = \frac{\omega_\mathrm{c}}{2\pi} = \frac{qB_z}{2\pi\gamma m_0} \approx 15.2\,\mathrm{MHz}\cdot B(\mathrm{T})\,(\,\text{for protons}\,). \quad (1)$$

Here m_0 and q are ion rest mass and charge, and B_z the vertically oriented bending field. The frequency is constant for low energies, a fact that made the construction of the first classical cyclotrons relatively easy (Fig. 1). But as soon as γ deviated by a few per cent from 1 the particles would run out of phase, and extended concepts had to be implemented, in the form of the isochronous cyclotron

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-005.151

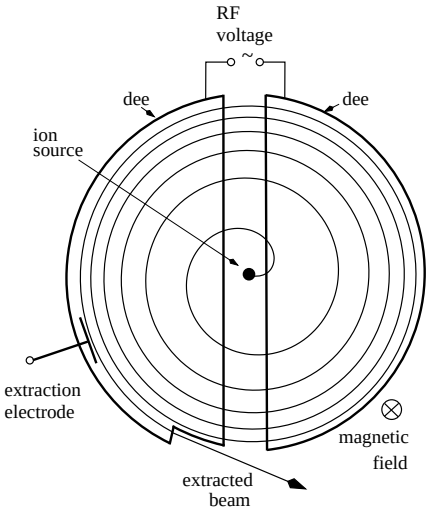


Fig. 1: Conceptual sketch of a classical cyclotron in plan view. Thanks to the simple concept, the first cyclotrons were constructed as far back as the 1930s.

and the synchro-cyclotron. In the isochronous cyclotron the average magnetic field is raised radially in proportion to γ, to keep the circulation time constant. Vertical focusing is provided by azimuthal variation of the bending field. The radial betatron frequency ν_r can be computed by expanding the equation of motion around the orbit radius R and by taking into account the radial slope of the bending field [6]. The result is

$$\nu_r \approx \sqrt{1 + \frac{R}{B}\frac{\mathrm{d}B}{\mathrm{d}R}} \approx \gamma. \tag{2}$$

The relation to γ is computed from the square-root term by taking into account the isochronous condition, i.e. $R \propto \beta$ and $B \propto \gamma$. In comparison to alternating gradient synchrotrons the focusing in cyclotrons is relatively weak since the beam energies are usually below 1 GeV. Also the vertical tune ν_z is typically a low number. This is important for injection matching and for acceleration of high-intensity beams, since the transverse space-charge forces counteract the relatively weak focusing of cyclotrons. Average beam optics parameters as beta-function and horizontal dispersion are estimated by $\beta_r \approx R/\gamma$ and $D_r \approx R/\gamma^2$. In order to prevent emittance blowup the injected beam must be matched to the appropriate optics on the first turns.

For clean extraction of high-intensity beams the turn separation is the most relevant parameter. To extract the beam by electrostatic deflection a thin electrode has to be placed in between the turns. To understand the basic scaling we consider here first the non-relativistic case with $\gamma = 1$, $B = \mathrm{const}$ and ΔE_k a constant energy gain per turn:

$$\begin{aligned} BR &= \frac{p}{q} = \frac{1}{q}\sqrt{2mE_k}, \\ \frac{\mathrm{d}R}{R} &= \frac{1}{2}\frac{\mathrm{d}E_k}{E_k}, \\ \Delta R(R) &= \frac{m}{q^2B^2}\frac{\Delta E_k}{R},\ \ \Delta R_{\mathrm{extr}} = \frac{1}{2}\,R_{\mathrm{extr}}\frac{\Delta E_k}{E_{k,\mathrm{extr}}}. \end{aligned} \tag{3}$$

Thus, during the acceleration process the turn separation is reduced as $1/R$. For a given final energy $E_{k,\mathrm{extr}}$ the turn separation scales with the extraction radius, i.e. with the size of the cyclotron.

In order to compute the turn separation for the case of the isochronous cyclotron with higher energies we have to include relativistic factors and the radial change of the magnetic field. As a starting point we use the total differential of the expression for the magnetic rigidity (6):

$$\frac{\mathrm{d}B}{B} + \frac{\mathrm{d}R}{R} = \frac{\gamma\,\mathrm{d}\gamma}{\gamma^2 - 1}. \tag{4}$$

This single equation relates the three parameters kinetic energy (through γ), orbit radius R and magnetic field B. With n_t denoting the turn number, the radius increment per turn is computed as

$$\begin{aligned} \frac{\mathrm{d}R}{\mathrm{d}n_\mathrm{t}} &= \frac{\Delta E_k}{{m_0c^2}} \frac{\gamma R}{(\gamma^2-1)\nu_r^2} \\ &= \frac{\Delta E_k}{{m_0c^2}} \frac{R}{(\gamma^2-1)\gamma}. \end{aligned} \tag{5}$$

The first line in (5) is a more general expression, indicating the option to change the radial slope of the magnetic field that affects ν_r and the turn separation. If the condition of isochronicity is applied through (2) one obtains the second line in which the slope of the field is not a free parameter. However, in the vicinity of the outer radius in a cyclotron this condition can be violated over a few turns in order to increase the turn separation locally for better extraction efficiency. For larger energy the strong dependence on γ leads to quickly diminishing turn separation. A practical limit is reached for proton cyclotrons at an energy of $0.8 \ldots 1$ GeV. For fixed final energy the turn separation also scales with the orbit radius at extraction, so it is generally easier to extract from a large cyclotron. The derivation of this relation is given in more detail in [6]. For example in the PSI Ring cyclotron with protons of 590 MeV kinetic energy and an extraction radius of 4.5 m this number is about 6 mm. The scaling with the size is illustrated in Fig. 2 for a hypothetical cyclotron that accelerates to 800 MeV with different maximum field strength at extraction, thus a varying overall size.

In compact cyclotrons with a factor 1000 less intensity, turn separation is not as critical and one can even allow overlapping turns with acceptable extraction losses. In order to realize extraction and injection schemes the beam must be deflected onto the first turn or out of the last turn. For this deflection static electric or magnetic fields are used. The challenge is to deflect only one turn and to minimize the effect on neighboured turns. Magnetic fields are used in septum magnets and the bending strength is given by the well-known magnetic rigidity, here given as a function of the relativistic γ factor:

$$B\rho = \frac{m_0 c}{q}\sqrt{\gamma^2 - 1}. \tag{6}$$

For injection and extraction elements the application of electric fields has some advantages. More complex field shapes can be realized, for example using a spiral inflector as described later. An electrostatic electrode can be very thin in cases where a neighboured turn passes closely. The corresponding electrostatic rigidity is calculated as follows:

$$E\rho = \frac{m_0 c^2}{q}\frac{\gamma^2 - 1}{\gamma}. \tag{7}$$

At low energies, e.g. at ion source energies, (7) can be approximated by $E\rho \approx 2U_\mathrm{acc}$, where U_acc is the accelerating voltage in the source. As it turns out, electrical fields are more effective at lower energy. For protons a few numerical examples for attainable bending strength are given in Table 1.

3 Internal sources

Internal ion sources are often used in compact cyclotrons. That is possible since the magnetic field covers also the central region of the cyclotron, and the beam can be accelerated from very low energy.

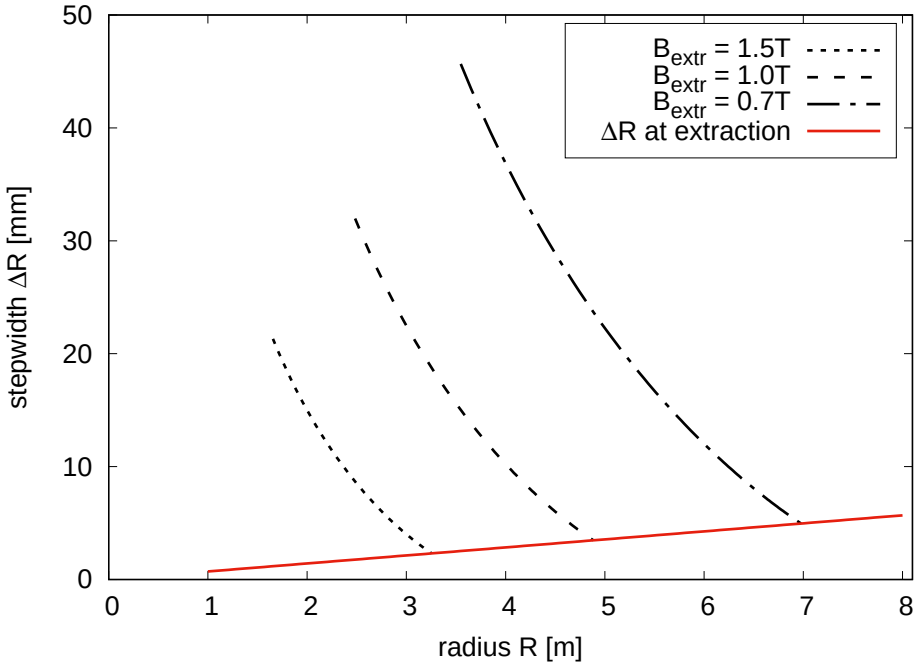


Fig. 2: During the course of acceleration, in all examples from 100 MeV to 800 MeV, the turn separation decreases as a function of radius. However, the final turn separation at the extraction radius is proportional to this radius, i.e. a larger cyclotron with lower magnetic field is advantageous for clean extraction.

Table 1: Bending radii achieved with typical magnetic and electric fields for protons at different kinetic energies

E_k	$B = 1\,\mathrm{T}$	$E = 10\,\mathrm{MV/m}$
60 keV	0.035 m	0.012 m
1 MeV	0.14 m	0.2 m
1 GeV	5.6 m	150 m

As compared with the external injection of a quality beam, an internal source delivers a diffuse cloud of ions. Typically a mixture of different charge states is created. An important aim of the central region design in a compact cyclotron is therefore the efficient capture of the desired ion species at low energies. If unwanted ions are accelerated to higher energies, e.g. beyond $\approx$ 7 MeV, these will create secondary activation. Fig. 3 shows a sketch of a cold cathode ion source and the concept of the central region in a 250 MeV cyclotron for cancer treatment (Varian).

In the 'chimney', electrons emitted from cathodes ionize injected hydrogen gas. The vertical magnetic field in the compact cyclotron is essential for this process since the electrons follow a spiral path in the field, which greatly enhances the ionization rate. The principle is similar to the function of an ion getter pump, or Penning ionization. The generated ions leave the chimney though a narrow slot. The first step of acceleration is achieved by a so-called puller electrode which is part of one of the four Dees in this cyclotron. Repetitive acceleration of the beam is performed by applying a high voltage RF potential to the Dees. The paths of the other charge states H_2^+ and H^- are also indicated in the figure. Unwanted ions are collimated at an early stage. For certain applications like fast pencil-beam scanning of cancer tumours, the stability of the beam current on a fast time-scale is important. Several parameters must be optimized for high stability and a high current for the Penning discharge seems to be particularly beneficial. Due to the injected gas the internal ion source is a limiting factor for the vacuum quality in a cyclotron. Another limitation of internal sources is the fact that heavier ions and specific charge states like H^- cannot be injected in this way. In summary internal sources present an elegant and cost-effective

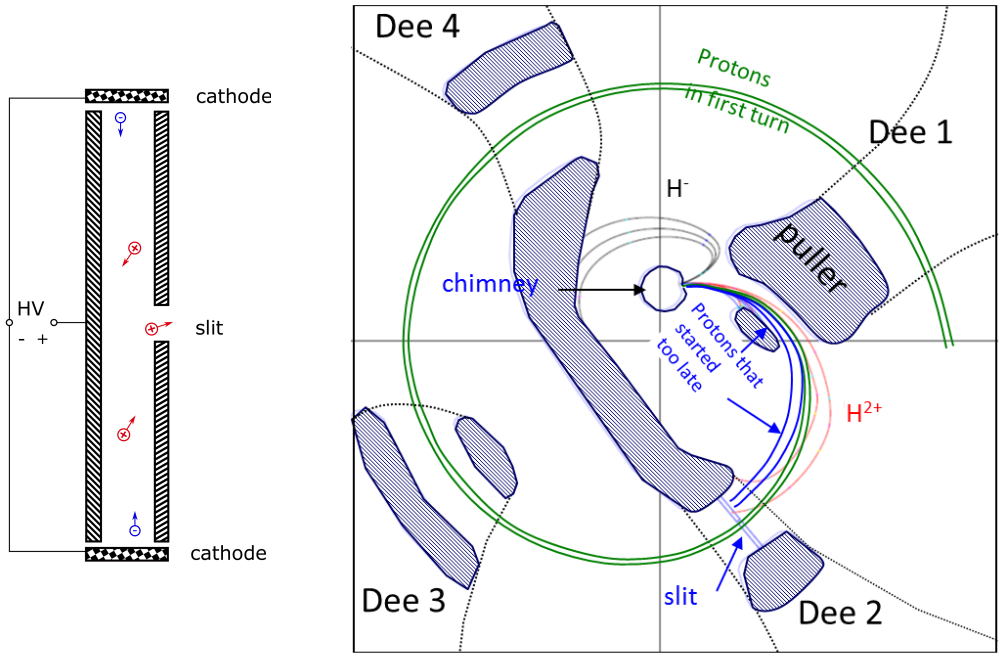


Fig. 3: Hydrogen atoms are ionized by electrons and emerge from the slit in the centre of the tube (left). The inner region concept of the compact superconducting 250 MeV Varian cyclotron is shown on the right (courtesy of M. Schippers, PSI). Collimation of unwanted charge states and deviating protons must be done at low energy.

solution for compact cyclotrons, but several limitations exist.

4 Injection

In many situations an external ion source or even a pre-accelerator must be employed. The reasons can be high intensity requiring a complex source, better beam quality, the need for specific ions or charge states. Separated-sector cyclotrons with a field-free region in the centre require a minimum beam energy, excluding the use of internal sources. Axial injection (compare Fig. 4) is often used for small cyclotrons with magnetic field in the centre. The beam enters vertically and is bent by 90 deg into the plane of the circulating orbits. The bending is realized by an electrostatic or magnetostatic inflector. The vertical magnetic field supports radial bending at the same time, turning the beam towards the direction of the first turn in the cyclotron. For horizontal injection the beam enters the cyclotron radially in the plane of the circulating orbits. It passes the centre region and is bend by magnetic elements towards the first turn orbit. The last step is often realized by an electrostatic deflector with the challenge not to affect the second and other circulating turns. Another option is stripping injection, where the beam passes a foil to remove electrons from the ion. The change in charge state brings the ion on the orbit of the first turn.

The design of a cyclotron injection requires consideration of many aspects, some of which affect the design of the entire cyclotron:

- longitudinal bunching and optimization of capture efficiency;
- optics matching including space charge for high-intensity beams;
- choice and design of the inflector element;
- central region design, radial and vertical centring.

An ion source delivers a DC beam, but for acceleration in a cyclotron the beam must be bunched. In the simplest case, similarly to the function of an internal source, the bunches are formed in the cy-

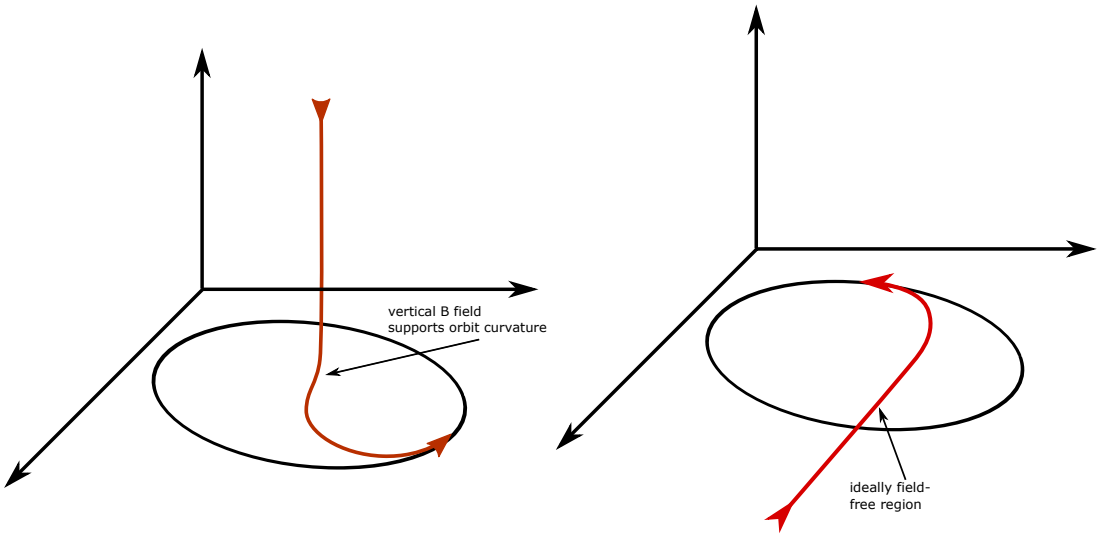


Fig. 4: Beam orbit paths for axial (left) and horizontal injection (right)

clotron itself since only ions arriving within a certain RF phase window will be accelerated. However, this scheme has low capture efficiency and one risks unwanted beam losses throughout the acceleration process.

The beam can be bunched by utilizing a buncher cavity outside the cyclotron. The buncher cavity introduces a velocity modulation resulting in the formation of bunches. Due to the acceleration in the cyclotron and the decrease of the relative velocity spread, the bunches sustain their shape. In the PSI high-intensity accelerator the beam is provided by a Cockcroft Walton pre-accelerator at an energy of 870 keV with low energy spread. In order to maximize the capture efficiency in the Injector II cyclotron one uses a third-harmonic buncher cavity at 150 MHz in addition to a fundamental-mode 50 MHz buncher. The harmonic cavity increases the linear region of the voltage modulation seen by the beam particles. The central region of the cyclotron has several adjustable collimators to remove protons with deviating phases from the beam at low energy. As an alternative to a Cockcroft Walton pre-accelerator, an RFQ pre-accelerator, providing an already-bunched beam, could be used.

Also in the transverse planes the injected beam must be matched to the envelope functions of the circulating beam. If the beam is not properly matched, the bunches will rotate in phase space, resulting in tails and potential difficulties at extraction. Complete filamentation of mismatched bunches is normally not an issue in a cyclotron, due to the short acceleration time. In comparison, the focusing strength in a cyclotron is much weaker than in alternating gradient synchrotrons. Beam envelope functions are thus just mildly modulated around the circumference. An example calculation is given in Fig. 5. In the case of high-intensity beams, space-charge forces must be taken into account. Transverse space-charge forces act against the focusing of the magnetic lattice and alter focusing frequency and envelope functions. Strong space-charge effects set an intensity limit for cyclotron, e.g. [9].

An electrostatic inflector for axial injection can be realized in different ways [3, 4]. The mirror inflector is the most simple design, consisting of a pair of charged plates. Geometry and voltage are adjusted so as to achieve the desired bending angle, normally 90 deg. As soon as the beam deviates from the vertical direction, the magnetic bending field of the cyclotron starts to deflect the beam as well, which forces the beam onto a spiral path. The magnetic force can be utilized to gradually bend the beam into the direction of the first turns orbit. The electrostatic force is applied in an optimal way if it acts perpendicular to the velocity vector of the particles at each point of the orbit. Only in this way a change of the kinetic beam energy is avoided. Such scheme is realized in the spiral inflector, e.g. Fig. 6, in which the two electrode plates are gradually bend with the changing direction of the beam. A detailed treatment

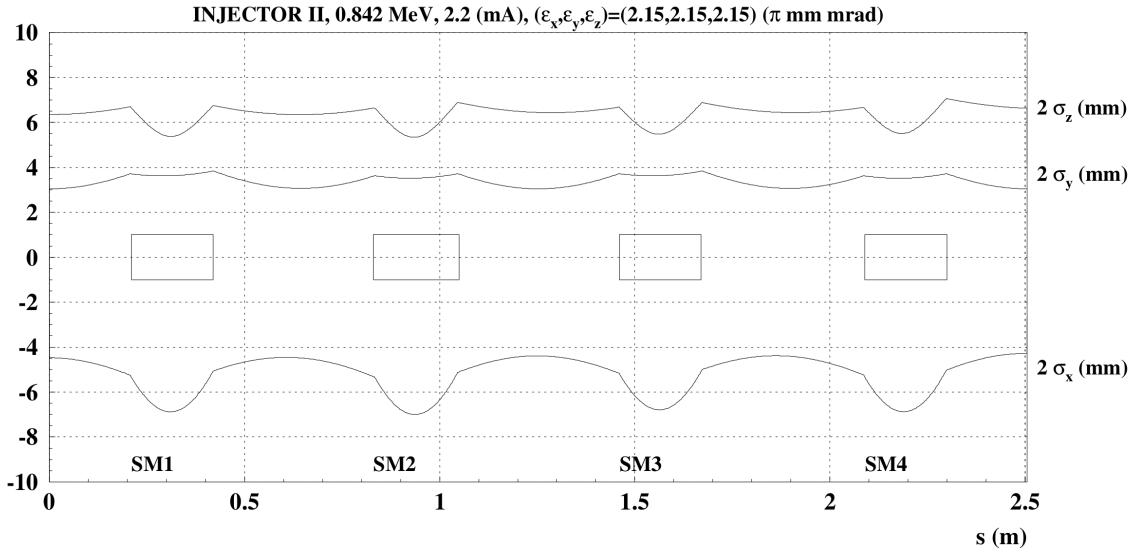


Fig. 5: Matched beam size in the PSI Injector II cyclotron as a function of azimuth at 0.84 MeV. The simulation includes space-charge forces [8].

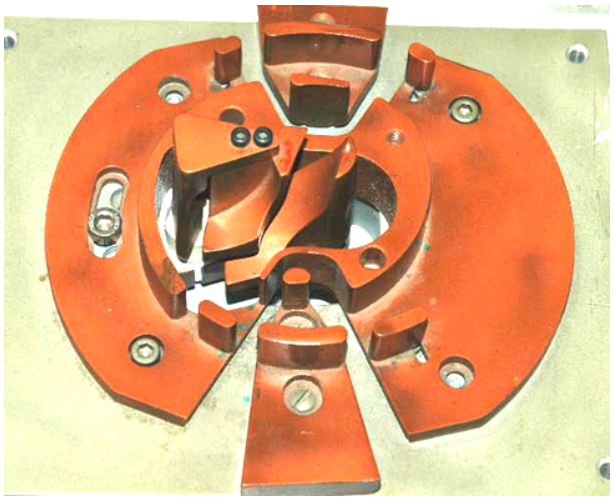


Fig. 6: Spiral inflector as used in a 30 MeV cyclotron (courtesy: IBA). The pair of curved electrodes is visible in the centre of the image.

of the beam dynamics in a spiral inflector is given in [7].

For horizontal injection in a separated-sector cyclotron the free space in the central region can be used to install magnets for bending and focusing of the beam to be injected. In order to finally deflect the beam onto the first turn orbit, often an electrostatic deflector with a thin electrode placed between the first and second turns is used. The design of this injection element can be similar as the electrostatic extraction channel in Fig. 7.

5 Extraction

For some applications, for example isotope production with low beam intensity, it is even unnecessary to extract the beam. The target can be inserted into the cyclotrons vacuum chamber for irradiation by a proton or ion beam. For compact, low-cost cyclotrons internal targets are the most effective solution. In

this case the main effort has to be spent on a reliable mechanism to exchange the targets.

For higher intensities extraction of the beam to an external beamline is required, and the methods of choice are then electrostatic deflection of the last turn, self-extraction in the fringe field of the cyclotron or charge-exchange extraction, 'stripping'. The latter of course requires acceleration of a not fully stripped ion, introducing specific issues.

Self-extraction of a beam can be achieved in the fringe field of the magnet at a radius where the normally positive slope of the magnetic field has decreased to a negative value at which the radial focusing stability is lost ($\nu_r = 0$, compare Eq.(2)). For this method we must avoid losing isochronicity too early. An extraction channel is machined in the magnet iron to provide a septum field and transverse focusing for the extracted beam. More details on this method are given in [3].

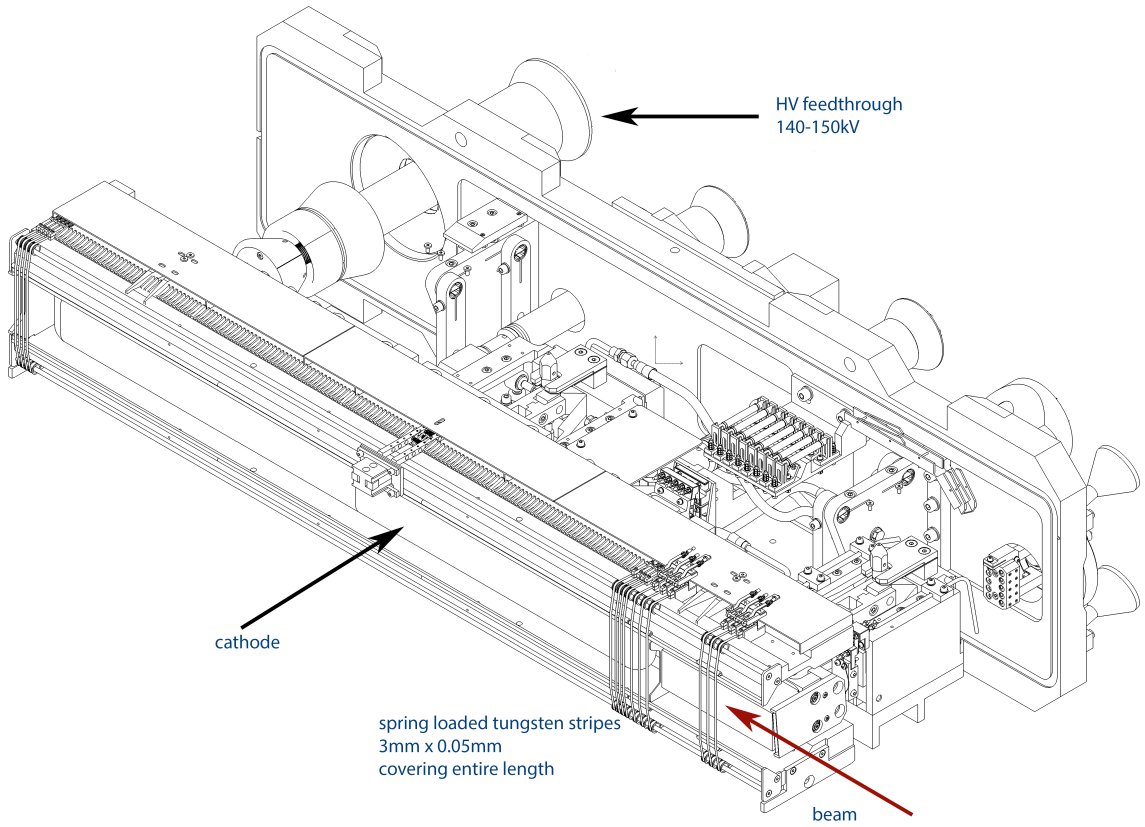


Fig. 7: The electrostatic extraction channel (EEC) of the PSI Ring cyclotron allows extraction of 2.4 mA CW current with relative losses below 2×10^{-4}.

For extraction with an electrostatic element (e.g. Fig. 7) the turns should ideally be separated to place the septum electrode in-between the turns, minimizing the rate of ions hitting the electrode and the resulting activation. The scaling of the average turn separation in cyclotrons was discussed in Section 2. Typical turn separation is in the range of a few millimetres. In summary, it diminishes quickly for high energy with γ^{-3} and it scales with the size of the cyclotron. Especially for compact superconducting cyclotrons with high fields the turn separation can be smaller than the beam size. But also for normalconducting cyclotrons this is a long-standing problem, and several methods have been developed to increase turn separation beyond the value generated by acceleration only. In several ways the concept of resonant extraction can be utilized. Coherent radial beam oscillations are excited by an harmonic bump. This is particularly effective in the vicinity of the integer resonance $\nu_r \approx 1$. As

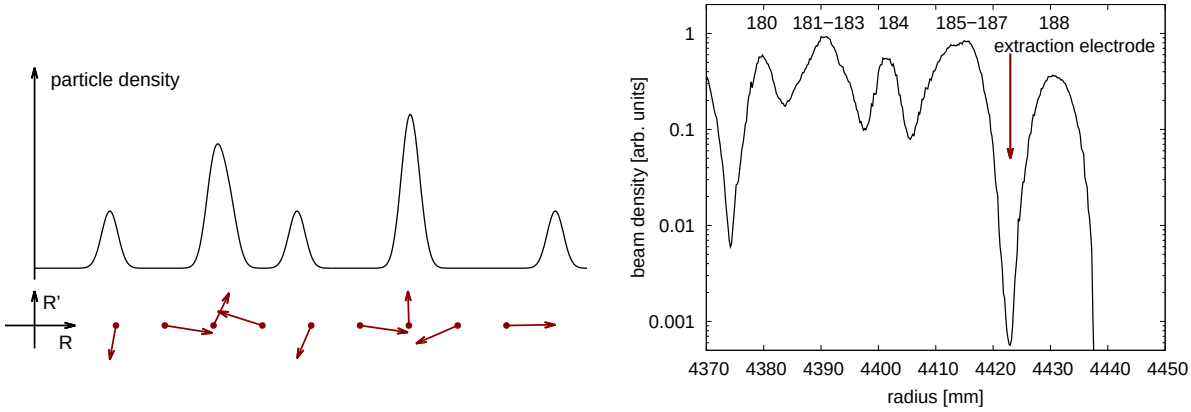


Fig. 8: Turn separation at extraction is enhanced by introducing coherent radial beam oscillations. The concept graph shows the clockwise rotating phase vector of the beam centre in the lower part. The multi-turn beam profile as a result of superimposing the shifted Gaussian beam distributions of individual turns is shown in the upper part. The right picture shows the measured beam density at extraction in the PSI Ring cyclotron on a logarithmic scale. Turn numbers and location of the electrode of the electrostatic deflector are indicated.

shown in Eq. (2), the radial tune is normally larger than 1 to keep the beam isochronous. However, close to extraction a phase slip can be tolerated for a few turns and the focusing is naturally weaker at the edge of the magnetic field. Close to the betatron resonance the amplitude of coherent oscillations is increasing quickly, thereby providing larger turn separation. In a modified version of this method, called precessional extraction, the tune drops quickly below 1 after exciting the resonance. This leads to an azimuthal shift of the maximum of the orbit oscillation from turn to turn, providing larger turn separation at the fixed azimuth of the septum. If the tune is close to a half-integer resonance, it is also possible to excite beam oscillations by introducing a gradient bump. Extraction efficiencies of 80 % can be achieved even in compact superconducting cyclotrons using these methods of resonant extraction.

In the following, we discuss the extraction process in the PSI Ring cyclotron in more detail. Coherent beam oscillations and a local shift of the radial betatron frequency are used to enhance the turn separation from the normal 6 mm gain by acceleration to a tripled value of 18 mm. Using an electrostatic septum it is possible to extract a beam of 1.4 MW average power with an efficiency of 99.98 %. The lost beam power of $\approx$ 200 W leads to a peak activation of $\approx$ 10 mSv/h in magnets of the extraction beam line, which is a tolerable level for service. The design of the electrostatic septum has been optimized in several iterations to improve performance and reliability, see Fig. 7. The electrode is realised as a series of 3 mm wide and 50 μm thin tungsten stripes. Tungsten has a high melting point and tolerates best accidental beam impact. If a stripe breaks it is pulled out of the beam plane by a spring. The effective electric length is 920 mm and the deflection angle 8.2 mrad. To reduce the number of high voltage trips, the element is covered by a metallic cage inside the cyclotrons vacuum chamber, to avoid accumulation of surface charges on the ceramic insulators. In the PSI Ring cyclotron the beam is already injected off-centre at 72 MeV. The coherent radial oscillation persists throughout the acceleration process. This approach has no significant drawbacks since the machine behaves sufficiently linear. At the coupling resonance $\nu_r \approx 2\nu_z$ the large radial oscillation could potentially be coupled to the vertical plane. In practice the resonance can be crossed quickly, avoiding vertical oscillations (Fig. 9). When the beam approaches the extraction radius the tune equals roughly $\nu_r \approx 1.75$. The phase space vector rotates by 3/4 of a circle per turn as shown in Fig. 8. At the azimuth of the septum the beam oscillation results in a certain pattern, with some turns overlapping. The three turns before the last turn, 185-187, are fully

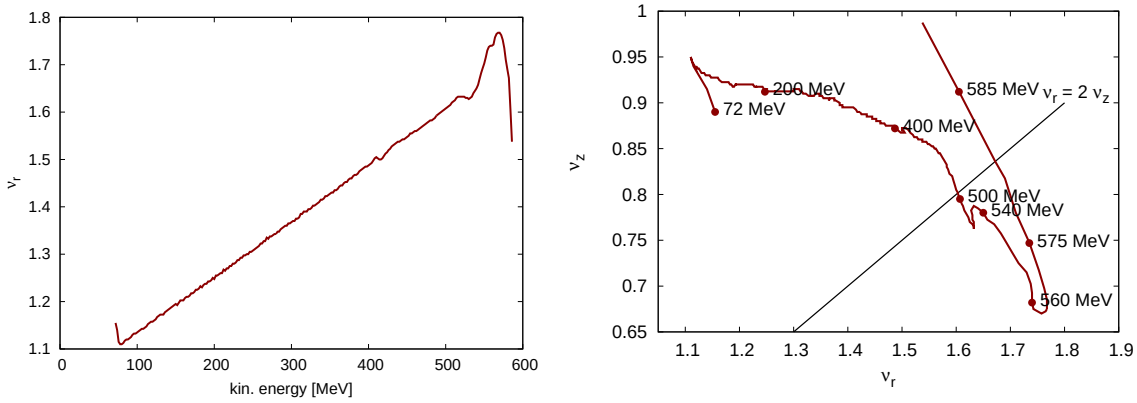


Fig. 9: The radial betatron tune in the Ring cyclotron follows closely the relativistic γ-factor (left). Over the last turns the focusing is reduced, enhancing turn separation. Coupling resonances must be crossed quickly (right) to avoid excitation of vertical oscillations.

overlapping and turn 187 is shifted inwards by the maximum oscillation amplitude. At this radius the tune is quickly reduced towards 1.5 so that the phase vector of the last turn is pointing outwards instead of downwards. As a result the normal turn separation is tripled in this scheme. In the right part of Fig. 8 a measured profile of the turns 180 to the extracted beam 188 is shown. At the location of the septum the particle density is three orders of magnitude lower than in the beam centre.

Besides the large turn separation, another ingredient of clean extraction are low beam tails. One important mechanism for tail production is longitudinal space charge, introducing energy spread, which is then converted into transverse tails. Fast acceleration is the best method to reduce these tails. Joho has shown in [10] that the attainable intensity in the PSI Ring cyclotron scales as the inverse of the cubed number of turns. Indeed, over the history of the accelerator significant improvements in intensity were achieved by raising the accelerating voltage, and the achieved intensity is in agreement with Joho's scaling law.

Stripping or charge-exchange extraction is another elegant scheme to extract a beam from a cyclotron. In this case ions are accelerated that are not fully ionized. The electrons are removed from the ions by passing them through a thin foil. The sudden change of the charge-to-mass ratio causes a change of the curvature of the ion path in a magnetic field. In this way the stripped ions are separated from the circulating beam and they are easily extracted. A prominent example for this technology is the acceleration of H^-, a proton with two bound electrons. After removing the two electrons the orbit curvature is inverted. By introducing the stripping foil at varying radius it is even possible to extract the beam at different energies, Fig. 10. While the advantages of the method are obvious, a number of difficulties must be mentioned. The second electron of H^- is bound rather weakly to the proton. In a magnetic field or by interaction with residual gas atoms, the ion is dissociated with relatively high probability. For example in the TRIUMF cyclotron the maximum magnetic field is limited to about $0.6\,\mathrm{T}$ for this reason. Another issue is the lifetime of the stripping foil, which is heated by direct interaction with the ion beam, but also by the stripped electrons. These electrons are spiralling in the magnetic field at a radius lower than the ion-bending radius by the ratio of the rest masses, which is typically just millimetres. The electron could pass the stripping foil many times until its kinetic energy is fully deposited in the foil. Thus design and positioning of the foil must be carefully optimized to obtain a good lifetime and serviceability.

Consequently H^- acceleration is only possible with low bending fields, but on the other hand superconducting magnet technology with high magnetic fields, is one of the major advancements in accelerator technology. As it has been shown, septum extraction is also difficult in compact cyclotrons. As a possible way out, the use of stripping extraction with accelerated H_2^+ ions is discussed. These ions have a higher binding energy (2.7 eV) for the electron than H^- (0.7 eV) and resulting dissociation prob-

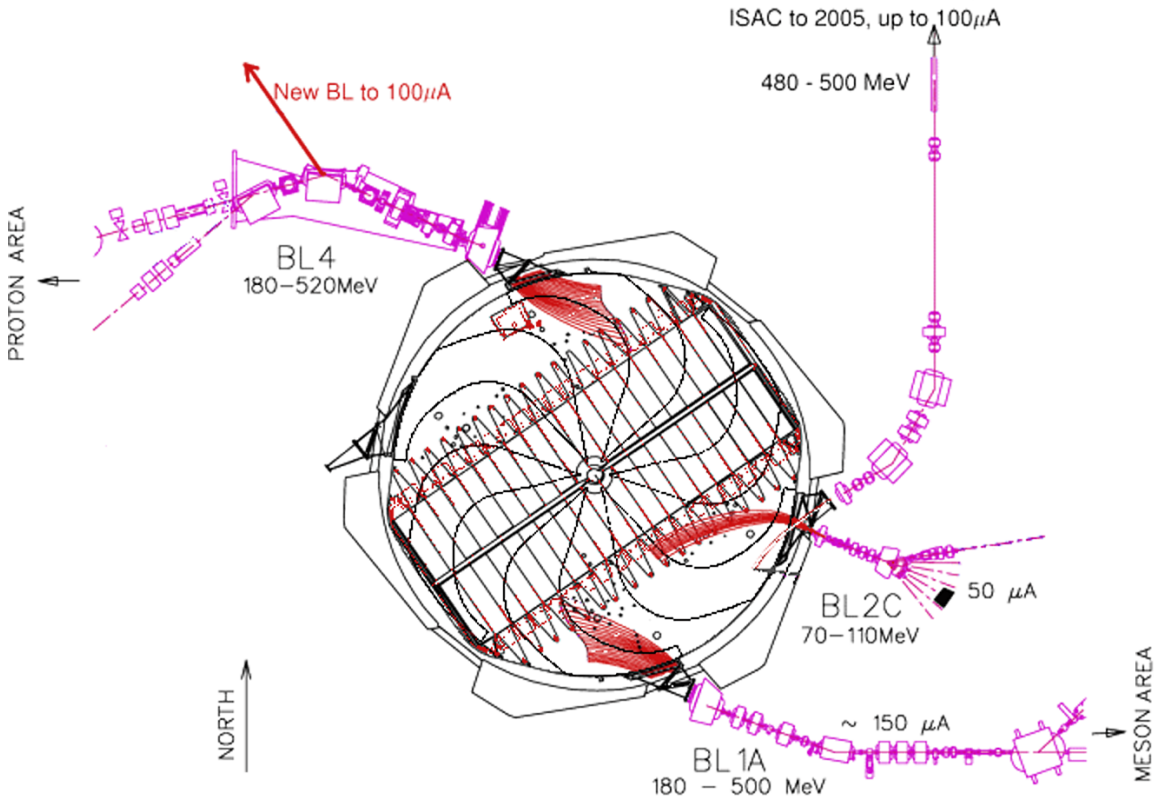


Fig. 10: Acceleration of H^- ions in the TRIUMF cyclotron and stripping extraction allows to provide multiple beams in parallel at variable energy (courtesy TRIUMF).

abilities even in high magnetic fields are low, allowing the acceleration of megawatt beams in relatively compact cyclotrons. On the downside, H_2^+ has a charge-to-mass ratio of 1/2, requiring higher bending fields than for protons. As another complication, stripping of the electron does not invert the curvature but reduces the bending radius by a factor of 2. This is sufficient to separate the stripped ions from the circulating beam, but it leads to a complicated extraction path across the cyclotron [11]. While the statement on higher binder energy is correct for the ground state of the ion, the situation is more complicated, in particular since this ion exhibits vibrational states with relatively long lifetime. Already in the source a certain fraction of the ions is produced in an excited state, from which dissociation is possible in strong fields. Thus the details of loss mechanisms for these ions are still under discussion.

6 Summary

Owing to the simplicity of the repetitive acceleration scheme with constant frequency the cyclotron presents an effective concept for ion beam acceleration. Often the challenging aspects of a design are efficient injection and clean extraction. In the simplest case an internal ion source can be used in compact cyclotrons. For higher intensity, better beam quality or certain ion beams, the beam must be generated outside the cyclotron and injected. Axial injection is used for smaller cyclotrons while horizontal injection is typically used for separated-sector cyclotrons with field-free regions in the centre. Electrostatic inflectors provide sufficient bending angles at low energy and allow realization of complex injection paths like the case of the spiral inflector. Charge-exchange or stripping processes can also be used to inject beams.

Extraction of a beam is a challenging process as well. Excessive uncontrolled beam losses lead to secondary activation of components, which should be avoided as best as possible. The most common extraction concept involves the use of an electrostatic septum, deflecting only the last turn. Turn separation is the most critical parameter and several schemes are used to increase the gap between last and second last turn beyond the value that is given by the energy gain of the beam. Self-extraction of a beam without a septum is also possible, but requires specific shaping of the magnetic field at the radius of extraction. Another elegant method is acceleration of ions that are not fully ionized and stripping of the electrons by a foil at the desired extraction radius. While this method provides more flexibility to vary the energy of the extracted beam, it has also disadvantages such as non-negligible probabilities of Lorentz dissociation before extraction. For high beam intensities the use of stronger bound H_2^+ ions is investigated.

References

[1] L.M. Onishchenko, *Phys. Part. Nuclei* **39** (2008) 950, https://doi.org/10.1134/S106377960806004X.

[2] E.O. Lawrence and N.E. Edlefsen, *Science* **72** (1930) 376.

[3] W. Kleeven, Injection and extraction for cyclotrons, CAS (2004), https://cds.cern.ch/record/1005057/files/p271.pdf.

[4] P. Mandrillon, Injection into cyclotrons, CAS (1996), https://cds.cern.ch/record/399426/files/p153.pdf.

[5] P. Heikkinen, Injection and extraction for cyclotrons, CAS (1994), http://cds.cern.ch/record/398439/files/p819.pdf.

[6] M. Seidel, Cyclotrons for high intensity beams, CAS (2013), https://cds.cern.ch/record/1513944/files/CERN-2013-001-p17.pdf.

[7] R. Baartman, W. Kleeven, *Particle Accelerators* **41** (1993) 41, https://cds.cern.ch/record/248437/files/p41.pdf.

[8] Ch. Baumgarten, *Phys. Rev. ST Accel. Beams* **14** (2011) 114201, https://journals.aps.org/prab/abstract/10.1103/PhysRevSTAB.14.114201, https://doi.org/10.1103/PhysRevSTAB.14.114002

[9] R. Baartman, Space charge limit in separate turn cyclotrons, Proc. 20th Int. Conf. Cyclotrons and Their Applications, Vancouver, 2013, http://accelconf.web.cern.ch/AccelConf/CYCLOTRONS2013/papers/we2pb01.pdf

[10] W. Joho, High intensity problems in cyclotrons, Proc. 5th Int. Conf. on Cyclotrons and Their Applications, Caen, 1981, p. 337, http://accelconf.web.cern.ch/AccelConf/c81/papers/ei-03.pdf

[11] C. Aberle *et al.*, Whitepaper on the Daedalus Program, arXiv:1307.2949v1[physics.acc-ph] (2013), https://www.researchgate.net/profile/Santo_Gammino/publication/248394961_Whitepaper_on_the_DAEdALUS_Program

Proceedings of the CAS–CERN Accelerator School: Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018)

Timing, Synchronization & Longitudinal Aspects

H. Damerau
CERN, Geneva, Switzerland

Abstract

The transfer of beams from one accelerators to another requires a set of parameters correctly adjusted, as well as a sequence of events to trigger all elements relevant to the transfer exactly at the right moment with respect to the beam. The course introduces the basic building blocks of synchronization systems and beam synchronous timing generation. The parameters for the transfer are then derived starting from longitudinal beam dynamics. Furthermore, The steps for the transfer of lepton and hadron beams between circular accelerators are summarized.

Keywords

Longitudinal beam dynamics, beam synchronous timing, bunch-to-bucket transfer, beam synchronization, energy matching, longitudinal matching

1 Introduction

Synchronization is required every time two or more events must take place at the same time and at the same location. This is as true for people who would like to get together for a meeting, where nonetheless being late by few minutes is usually considered acceptable, as for particles beams which meet another beam or injection and ejection elements that have to be triggered for the beam transfer. The precision needed depends on the system considered. While two synchronized events may be millions of years apart in astronomy, particle beams travelling at a significant fraction of the speed of light require time precisions from nanoseconds down to few femtoseconds.

Table 1 gives the order of magnitude for the typical time scales, mostly defined by the length of the particle bunches. The technology for the synchronisation system must be chosen accordingly.

Table 1: Time scales of common accelerator types, mainly defined by their bunch length.

Proton bunches in low energy synchrotrons (few GeV)	$>$ 10 ns
Proton bunches in medium energy synchrotrons	100 ns...1 ns
Bunch length in hadron colliders ($>$ 100 GeV)	10 ns...100 ps
Electron storage rings	10 ps...1 ps
Plasma wakefield accelerators	1 ps...100 fs
Self Amplified Spontaneous Emission (SASE) FELs	1 ps...10 fs
Pump-probe FELs	$<$ 100 fs
Jitter achievable with low-level RF electronics	10 ps...1 ps

While standard low-level RF electronics hardware is fully sufficient for hadron and circular electron accelerators, free-electron lasers with bunch lengths in the pico-second regime and below require optical synchronization techniques [1].

To illustrate the underlying principles and for the sake of simplicity this course focuses on standard techniques of synchronization and longitudinal transfer of beams, with particular attention to the transfer of bunched beams between two circular accelerators. The implementation is described only to the detail required to understand the concepts to synchronize beams, derive beam synchronous timings, achieve energy, as well as longitudinal matching.

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-005.163

Due to their low mass, leptons are extremely close to the speed of light in most accelerators. Already 99.99995 % of it are reached even in a small size electron synchrotron with an energy of only 500 MeV [2]. The oscillations of particles in the longitudinal and transverse directions are moreover damped by synchrotron radiation [3], forcing the particles to the longitudinal position enforced by the RF fields. In electron accelerators beam transfers are therefore fully deterministic in most cases. All signals for the RF systems and the triggers involved in the transfer are derived from a single master clock.

The fixed frequency approach is not possible in hadron accelerators where beam synchronous signals are derived from the circulating beam. The beam-based generation of RF signals is required to avoid longitudinal beam oscillations in the absence of synchrotron radiation. The exact moment of a beam transfer in absolute time hence depends on the beam itself and changes from acceleration cycle to cycle. Additionally, the RF frequency sweeps to track the increasing revolution frequency during acceleration.

Following the introduction of basic techniques and building blocks for beam transfer systems, the transfer between two circular accelerators is discussed. The longitudinal beam dynamics aspects are introduced in view of the transfer of bunched beams. Thereafter, the relevance of periodicity and bucket numbering is explained. The generation of trigger events for the beam transfer completes the requirements for electron accelerators. In hadron accelerators frequencies change and the target for a beam transfer can move in time. Additional difficulties are caused by the absence of synchrotron radiation damping. Strategies to match the beam energy and the longitudinal parameters are described to transfer hadrons with virtually no degradation of the longitudinal beam quality.

2 Techniques and building blocks

The basic techniques and building blocks of synchronization systems for beam transfer, in particular for large scale accelerator facilities, are introduced in what follows. Signals, either in the form of time marker pulses, RF signals, or as general timing infrastructure must be exchanged between accelerators involved in a beam transfer. The distances for these transmission can range up to several kilometers.

2.1 Signals with noise

The quality of any RF signal is degraded due to noise which leads to an unwanted variation of the signal phase or amplitude with time, the so-called jitter. In particular a phase variation, the phase jitter, introduces an uncertainty to the definition of phase and time information contained in the signal.

2.1.1 Coherent and incoherent signals

The difference between a coherent signal and noise is illustrated in Fig. 1. It shows the same signal

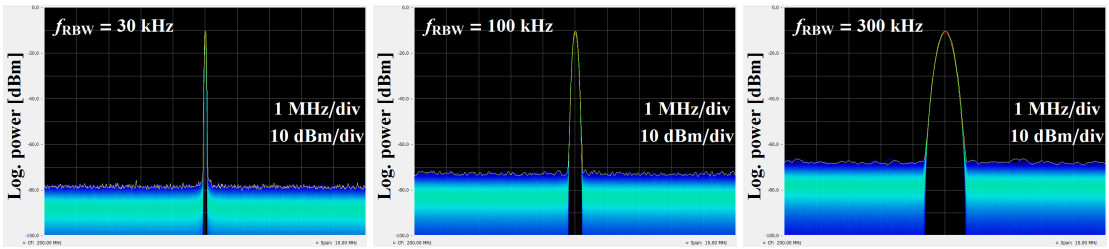


Fig. 1: Spectrum of an RF signal at 200 MHz measured with a resolution bandwidth, f_{RBW} of 30 kHz (left), 100 kHz (middle) and 300 kHz (right).

at 200 MHz measured with a spectrum analyser at three different resolution bandwidths. While the peak of the spectrum, representing the coherent part of the signal, is always at the same amplitude, the

noise level, as well as the width of the peak change proportionally to the observation bandwidth called resolution bandwidth, f_{RBW}, of the spectrum analyser.

Since the incoherent noise has a quasi-constant power density, i.e., power per bandwidth, the total power passing through the observation filter of the spectrum is proportional to its bandwidth. In case of the measurement shown in Fig. 1, the noise level is mainly dominated by the front-end amplifier and the mixer of the spectrum analyzer.

In the ideal case the thermal noise power density of a passive object at the temperature T is given by [4, 5]

$$\frac{P}{\Delta f} = kT\,G(f) \quad \text{with} \quad G(f) = \frac{hf}{kT} \Big/ \left(e^{hf/kT} - 1\right) .$$

The Boltzmann and Planck constants are k and h. In the relevant frequency and temperature range of electric signals on conductors, the form factor $G(f)$ can be approximated to one, since $hf/kT \ll 1$. At room temperature, $T \simeq 296$ K, the noise power density hence becomes $P/\Delta f \simeq kT = -174\,\mathrm{dBm/Hz}$.

2.1.2 Analysis of phase noise

The power spectrum of oscillator at the centre frequency, f_c, with phase jitter is sketched in Fig. 2. All

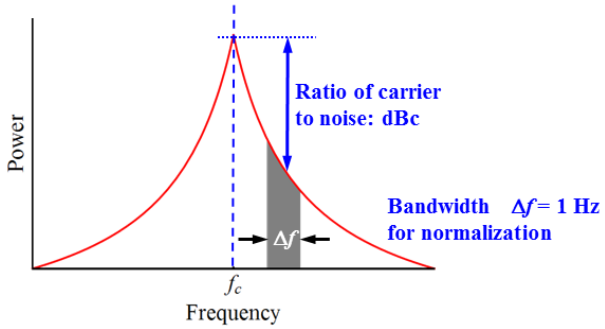


Fig. 2: Power spectrum of an oscillator signal at the frequency, f_c, with side-bands caused by noise.

power levels are defined as ratios with respect to the amplitude of the signal at the centre frequency, the carrier, in units of dBc. The noise power density is defined as the ratio of the power density and carrier power [6],

$$\mathcal{L}(f) = \frac{\text{Power density}}{\text{Carrier power}} \left[\frac{\mathrm{dBc}}{\mathrm{Hz}}\right] = \frac{1}{2} S_\phi(f)\,. \tag{1}$$

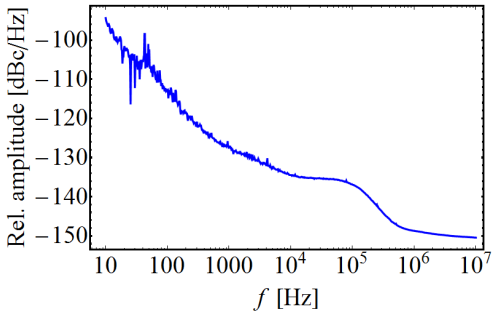
Integrating the noise power density Eq. (1) and using $\Delta t = \Delta\phi/(2\pi f_c)$ gives the RMS jitter of the signal

$$\Delta t_{\mathrm{rms}} = \frac{1}{2\pi f_{\mathrm{c}}} \sqrt{\int_{f_1}^{f_2} S_\phi(f)\, df} \tag{2}$$

in the frequency range from f_1 to f_2. Seen that the noise power of most signals decreases with the frequency offset from the central frequency, the integration limits are conveniently given as offset frequencies from the carrier.

Fig. 3 (left) shows the phase noise density of a laboratory RF generator (Agilent/Keysight E8663B) at offset frequencies from 10 Hz to 10 MHz from the carrier at 1 GHz, together with the jitter per decade (Fig. 3, right). Due to the form of Eq. (2) the individual jitters per decade, $t_{\mathrm{rms},n}$, can be added as a geometric sum such that the total jitter becomes

$$\Delta t_{\mathrm{rms}} = \sqrt{\Delta t_{\mathrm{rms},1}^2 + \Delta t_{\mathrm{rms},2}^2 + \ldots + t_{\mathrm{rms},n}^2}\,.$$



Frequency range	Δt_{rms} [fs]
10 Hz. . .100 Hz	12.4
100 Hz. . .1 kHz	5.5
1 kHz. . .10 kHz	5.5
10 kHz. . .100 kHz	11.1
100 kHz. . .1 MHz	13.0
1 MHz. . .10 MHz	21.3
Total	31.0

Fig. 3: Phase noise density versus frequency offset (left) and RMS jitter per decade of frequency offset (right).

2.2 Transmission of RF signals over long distances

In large scale accelerator facilities reference signals must often be transmitted over distances of several kilometres, from one accelerator to another. Without taking any precautions, the thermal drift of such long coaxial cables or optical fibres may become significant with respect to the bunch length. The thermal coefficient of delay (TCD), defined as

$$\mathrm{TCD} = \frac{\Delta\tau}{\tau} \cdot \frac{1}{\Delta T} = \frac{\Delta\phi}{\phi} \cdot \frac{1}{\Delta T}\,,$$

as measured for a standard RG223 coaxial cable is plotted in Fig. 4. A 2 km long cable of this type

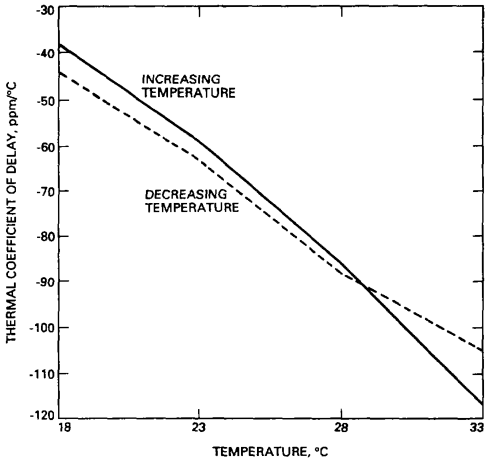


Fig. 4: Thermal coefficient of delay for an RG223 coaxial cabled measured with increasing and decreasing temperature [7].

has about 10 μs delay. Changing the temperature by only 1 K at room temperature changes its delay by about 0.5 ns. This may be acceptable for the RF frequency as low as 10 MHz like in the Proton Synchrotron (PS) at CERN, but is certainly not in the case of the Large Hadron Collider (LHC), where the same delay would lead to a phase error of 73 ° at an RF frequency of about 400 MHz.

The thermal coefficient of delay for optical fibres is typically 10 to 100 times smaller compared to coaxial cables, but still the remaining drift may not be acceptable. The phase drift of a optical fibre with a length of about 6.3 km measured at a frequency of 200 MHz together with the temperature is shown in Fig. 5. Without compensation, the delay of the fibre changes by almost 1 ns which is of the same order as the length of the proton bunch and well beyond the acceptable range.

The stability in time of the signal transmission can be significantly improved by active compensation. Part of a signal transmitted from point A to point B (Fig. 6, left) is sent back to the point A and

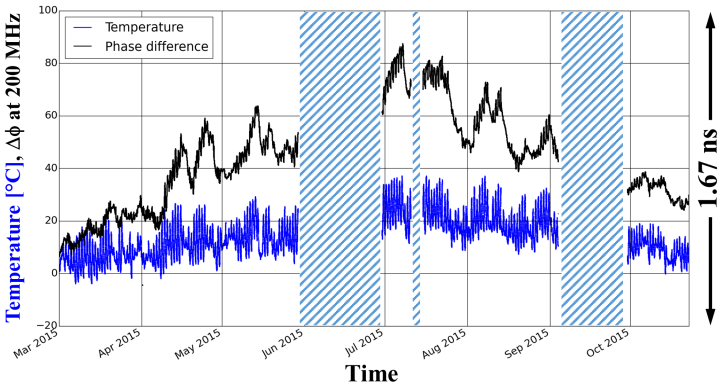


Fig. 5: Drift of a 6.3 km long optical fibre, from the longitudinal beam control system of the CERN SPS to the AWAKE experiment and back [8]. The large fraction of the fibres is installed underground and the temperature may therefore vary along the fibres.

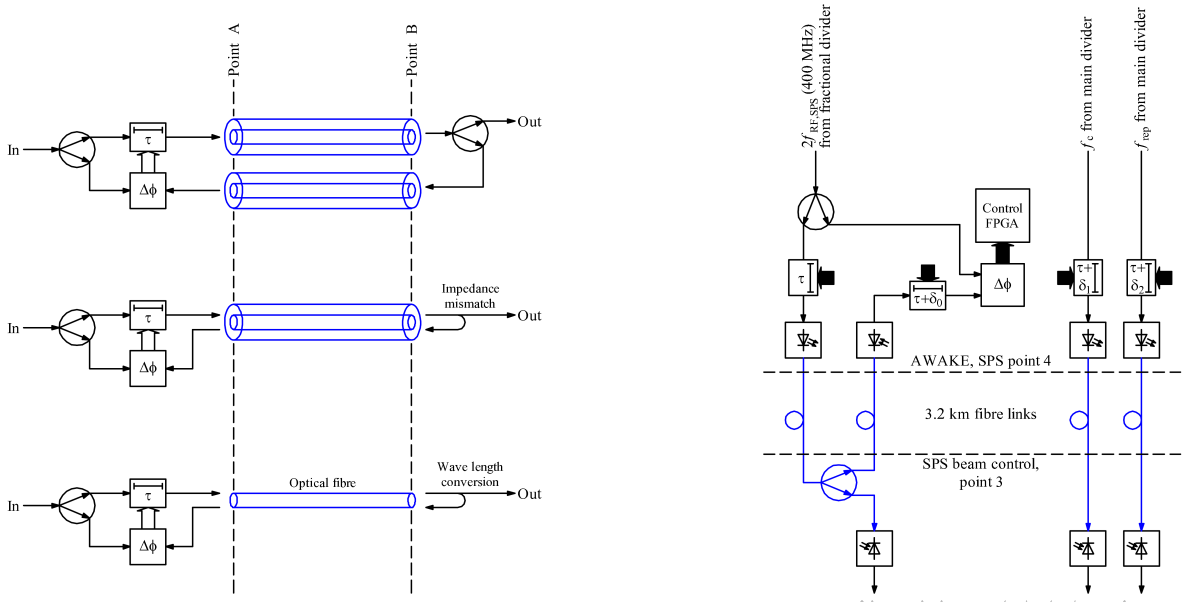


Fig. 6: Principle configurations for active link delay stabilization (left) and set-up to compensate the delay of a link for the AWAKE experiment at CERN between (right). The length of the optical fibres is about 3.2 km [9].

compared there in phase to the source signal. In the most simple case the return signal is split at the destination point and part of it is transmitted back. Such a scheme assumes, however, that the behaviour of both transmissions is identical. This may be difficult to achieve, especially for coaxial cables.

Using the same transmission medium to return part of the signal removes that constraint (Fig. 6, left middle). For a coaxial cable an intended impedance mismatch at the end of the line (point B) causes a reflection travelling back to point A. A directional coupler (not shown) extracts the reflected signal before comparing it in phase with the source signal. The directivity of the directional coupler may limit the achievable performance.

Next to their inherently smaller thermal coefficient of delay, optical fibres also allow to transmit light at several wavelengths simultaneously (Fig. 6, left bottom). The return path can therefore use the same fibre, but at a different wavelength. This removes the need to split forward and reflected waves. To further improve the stability, parts of the link compensation hardware can me implemented optically, e.g., using an optical phase discriminator.

Fig. 6 (right) shows an overview diagram of the optical link for the synchronization of proton bunches from the SPS to the AWAKE experiment at CERN. Next to the compensated transmission of an RF signal at 400.8 MHz, two more signals are sent. Both of them are only required to start counters at a specific period of the RF signal and no particular precision is needed. Their delay simply tracks the one of the 400.8 MHz link to keep all three signals approximately in phase at the arrival in the SPS. Although the configuration with separate optical fibres for forward and return paths stabilizes the delay to the order of 1 ps, the configuration according to Fig. 6 (left bottom) will further reduce the drift.

To obtain even higher stability, the phase measurement and the compensation delay can be implemented as optical components. Ultimately, the phase comparison is performed at a high harmonic of ultra-short laser pulses (5 THz) or directly at the frequency of the light (200 . . . 1000 THz) [1, 10, 11].

2.3 Synchronous frequency dividers

The most common technique to generate a lower frequency from a higher one are frequency dividers. A frequency divider by n, the division ratio, is an electronic counter which counts up to n with every period of the signals at its input. It then generates a pulse at its output and resets itself. Frequency counters exist in many variants, counting up or down, with and without preload, etc. However, independent of their implementation, the counting value of two independent dividers is not necessarily the same.

Fig. 7 illustrates the indeterministic behaviour of two frequency dividers. The circuit on the left

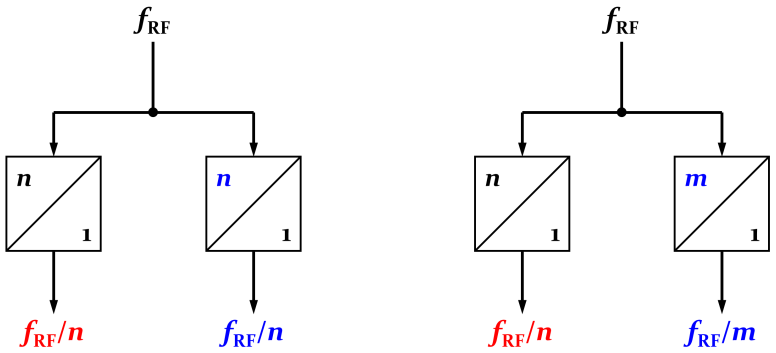


Fig. 7: Configuration of two frequency dividers counting the same input signal. The undefined initial counting state of the dividers results in indeterministic behaviour of the set-up. Both output signals f_{RF}/n (left) may not be identical. The output signals of the circuit dividing by n and m (right) change their relative output phase according to the initial phase.

divides the same RF signal twice by n. However, keeping in mind that the initial condition is not defined when powering the circuit on, the output phase of both dividers can differ by multiples of $2\pi/n$. The circuit on the right (Fig. 7) can even have up to $n \cdot m$ different output phases. Multiple dividers without reset are a common pitfall in synchronization systems. They work correctly until one of the divider misses one or multiple counts with respect the other divider and the relative of the output signals changes.

Two options to force a deterministic behaviour are sketched in Fig. 8. The output of one the two dividers (master, left) is used to reset the second one. This assures that the slave divider starts counting at a well-defined counting state of the master divider. The output signals of both will remain at a fixed phase relationship until one of the dividers misses one clock pulse or counts, e.g., an additional glitch. If a programmable phase shift in units of input clock cycles between both output signals is required (e.g., Sec. 5.2), the circuit shown in Fig. 8 can be employed. The counter value of the first divider is added to an offset by a modulo n counter. This counter wraps around when reaching the programmed division ratio.

In circular accelerators all beam synchronous signals periodic with the revolution period of the

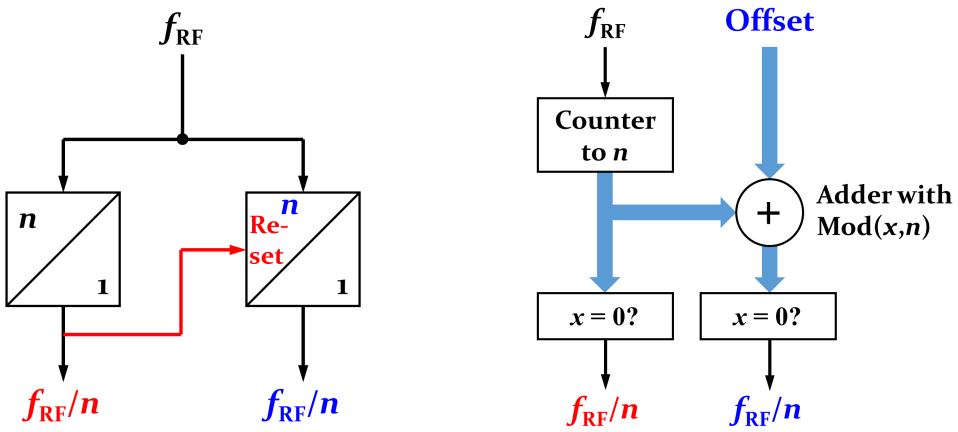


Fig. 8: Configurations to force an unambiguous phase relationship at the output of two dividers. One of the two divider becomes the phase reference (left) and its output resets the second, slave divider. A modulo n adder with a programmable offset (right) can be used to generate a phase offset between both outputs.

circulating beam and dividers are usually reset with a master revolution frequency signal. For the beam transfer between to circular accelerators a fiducial or common frequency is chosen with respect to which the whole transfer process is periodic (Sec. 4.1).

2.4 Phase detectors

Measuring and correcting frequency and phase differences between RF signals is important to synchronize beams in different accelerators in preparation of the transfer. The phases of these signals represent the relative azimuthal positions and revolution frequencies of beams in circular accelerators.

Frequency, $\omega = 2\pi f$, and phase, ϕ are actually equivalent, since

$$\omega = \frac{d\phi}{dt} \quad \Leftrightarrow \quad \phi = \int \omega \, dt \, . \tag{3}$$

Phase is a relative quantity and phase detectors, also called discriminators, always measure the phase difference between two signals. The Eqs. (3) can also be applied to phase and frequency differences $\Delta\omega = d(\Delta\phi)/dt$. Two special cases are very common in synchronization systems. Firstly, the phase difference of two signals at the same frequency should be detected, e.g., to determine phase between bunches of a beam and the accelerating RF signal. In this case the phase difference is constant with time. Secondly, the frequencies of two signals, e.g., circulating beams in two rings of a collider, are constant, but slightly different. Following Eq. (3) the phase then changes linearly. The latter example is plotted in Fig. 9. However, even an ideal phase discriminator cannot distinguish between a phase difference $\Delta\phi$ and $\Delta\phi + n \cdot 2\pi$. The measured phase difference will hence not just accumulate to infinity, but will wrap around within the range of the phase discriminator. The result is, for the case of two signals with a constant frequency difference, a sawtooth-like phase (Fig. 9).

The most simple phase detector is a frequency mixer, a device which multiplies two signals. This can be easily shown for two sinusoidal at given frequencies ω_n and phases ϕ_n. The product

$$\sin(\omega_1 t + \phi_1)\sin(\omega_2 t + \phi_2) = \frac{1}{2}\{\cos\left[(\omega_1 - \omega_2)t + (\phi_1 - \phi_2)\right] - \cos\left[(\omega_1 + \omega_2)t + (\phi_1 + \phi_2)\right]\} \tag{4}$$

can be expressed as a sum of two components at the frequencies $\omega_1 - \omega_2$ and $\omega_1 + \omega_2$. The first component contains the wanted phase difference $\phi_1 - \phi_2$. To extract it, the output signal of the mixer is low-pass filtered to remove the higher frequency component. For the two signals at the same frequency, the phase difference is constant and the low frequency component in Eq. (4) hence becomes just $1/2\cos(\phi_1 - \phi_2)$.

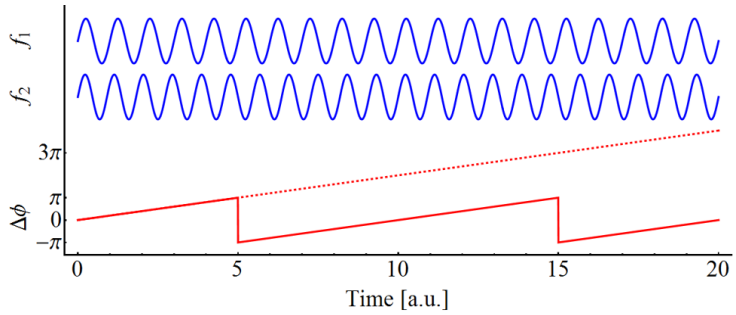


Fig. 9: Phase difference of two signals with slightly different frequency. The phase difference is a linear function with time (red dotted line), but wraps around when measured with a phase detector.

Arbitrarily subtracting a constant phase of $\pi/2$, keeping in mind that the phase detection is a relative measurement, results in a phase dependency proportional to $\sin(\phi_1 - \phi_2)$, which can be approximated to $\Delta\phi = \phi_1 - \phi_2$ for $\Delta\phi \ll 1$. In the given range of phase differences, the frequency mixer will hence indicate the instantaneous phase difference between both inputs (Fig. 10). Table 1 gives a list of

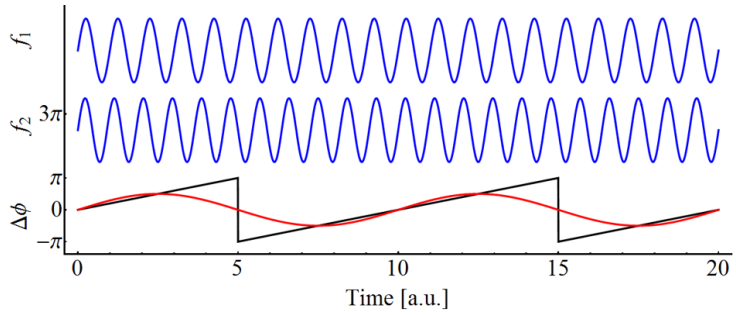


Fig. 10: Phase difference of two signals with slightly different frequency as in Fig. 9. A mixing phase detector only resolves the phase within a small range.

common phase detectors including their basic properties [12, 13]. Phase detectors with full 2π coverage

Table 2: Common phase detector types with range and behaviour.

Type	Range	Behaviour	Reference
Mixer, analogue 4 quadrant multiplier	π	Sinusoidal: $s_{\text{out}} \propto \cos\phi$	[14, 15]
Exclusive OR gate	π	Linear: $s_{\text{out}} \propto \phi - 3\pi/2$, or $s_{\text{out}} \propto -\phi + \pi$	[16]
Sample and hold	π	Sinusoidal: $s_{\text{out}} \propto \sin\phi$	[17]
Flip-flop phase detector	π	Linear: $s_{\text{out}} \propto \phi - \pi$	[12]
Tri-state double flip-flop	2π	Linear: $s_{\text{out}} \propto \phi$	[18]
COordinate Rotation DIgital Computer CORDIC	2π	Linear: $s_{\text{out}} \propto \phi$	[19]
Sagnac loop with optical modulator	$< \pi$	Sinusoidal: $s_{\text{out}} \propto \sin\phi$ (clipped)	[20]

are preferred whenever. They avoid the $\pm\pi$ ambiguity which may, e.g., cause a loop to lock onto an unstable phase.

In terms of absolute precision in time, a phase measurement at a frequency as high as possible is preferable since any error in time is inversely proportional to the frequency, $\Delta\tau = \Delta\phi/(2\pi f)$. However, increasing in frequency may introduce ambiguity. A train of bunches, a so-called batch, spaced by $1/f_{\text{RF}}$ with $f_{\text{RF}} = h f_{\text{rev}}$ can have the same phase every period of the RF frequency. Measuring the phase of the same batch at the revolution frequency defines one unique position of the complete batch per turn.

2.5 Phase-locked loops

A phase-locked loop (PLL) allows to keep two RF or beam signals at a well defined phase relationship. It is also a standard technique to multiply the frequency of a signal, like an inverse frequency divider. The basic building blocks of the PLL are phase discriminator, loop filter and voltage controlled oscillator (VCO) as sketched in Fig. 11. A divider may be added in case the two frequencies to be locked

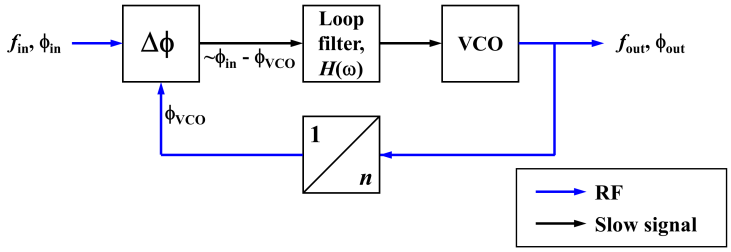


Fig. 11: Typical PLL set-up to synchronize RF signals or to multiply them in frequency.

together are not identical, but in an integer ratio. The input signal, characterised by its frequency, f_{in}, and phase, ϕ_{in}, is compared with the signal from the VCO whose frequency can be controlled closely around nf_{in}. The phase difference, a low bandwidth signal proportional to $\phi_{\mathrm{in}} - \phi_{\mathrm{VCO}}$, is passed through a loop filter which drives the VCO. This closes the loop.

The frequency of the VCO, f_{out}, is hence locked to f_{in} according to $f_{\mathrm{out}} = nf_{\mathrm{in}}$. At the same time the phase relationship between ϕ_{in} and ϕ_{out}/n is fixed. It is worth noting that the divider does not need to be synchronized as described in Sec. 2.3. The phase comparison is performed with the output signal of the divider with respect to which the counting state is unambiguously defined. The derivation of the transfer function of the PLL is detailed in [12, 21].

Next to its application as a frequency multiplier, the PLL is also used to regenerate signals from a circulating particle beam. In the context of synchronisation systems, this so-called beam phase loop (Sec. 5.1) locks an RF signal to the beam, which allows to derive beam synchronous signals.

3 Longitudinal beam dynamics

The energy gain per turn, ΔE_{turn}, of a particle with the charge, q, in a synchrotron is given by the effective voltage, $V(\phi) = V \sin\phi$, in the accelerating RF cavity at moment of the passage of the particle such that

$$\Delta E_{\mathrm{turn}} = qV \sin\phi \,. \tag{5}$$

To keep the particle on the central orbit, this energy gain must match the energy gain per turn due to the changing magnetic field, $B(t)$,

$$\Delta E_{\mathrm{turn}} = 2\pi q\rho R \frac{d}{dt} B(t) = 2\pi q\rho R\dot{B} \,. \tag{6}$$

The average radius of the accelerator is described by R and its magnetic bending radius by ρ. The combination of Eqs (5) and (6) defines the stable phase

$$\sin\phi_0 = \frac{qV}{2\pi q\rho R\dot{B}} \,.$$

For a particle with a phase offset, $\Delta\phi$, with respect to the synchronous particle the energy offset can be written as

$$\frac{d}{dt}\Delta E = qV\frac{\omega_{\mathrm{rev}}}{2\pi}\left[\sin(\phi_0 + \Delta\phi) - \sin\phi_0\right] \,. \tag{7}$$

The relationship between a frequency and momentum offset is called phase slip factor and defined as[1]

$$\eta = \frac{\Delta\omega/\omega}{\Delta p/p} = \frac{1}{\gamma^2} - \alpha_c = \frac{1}{\gamma^2} - \frac{1}{\gamma_{\mathrm{tr}}^2}\,, \tag{8}$$

with the momentum compaction factor $\alpha_c = (\Delta R/R)/(\Delta p/p) = 1/\gamma_{\mathrm{tr}}^2$.

Replacing $\Delta p/p = 1/\beta^2 \Delta E/E$ and using [22] $d(\Delta\phi)/dt = -h\Delta\omega$ results in [23]

$$\frac{d}{dt}\Delta\phi = -h\omega_{\mathrm{rev}}\frac{\eta}{E\beta^2}\Delta E = -\frac{h\eta}{pR}\Delta E\,. \tag{9}$$

The ratio between RF and revolution frequency is called harmonic number, $h = \omega/\omega_{\mathrm{rev}}$.

Inserting Eq. (7) in the derivative of Eq. (9) results in the equation of motion of a particle in phase

$$\frac{d^2}{dt^2}\Delta\phi + \frac{h\eta\omega_{\mathrm{rev}}qV}{2\pi pR}\left[\sin(\phi_0 + \Delta\phi) - \sin\phi_0\right] = 0\,. \tag{10}$$

For small deviations from the synchronous particle the non-linear terms can be approximated according to

$$\sin(\phi_0 + \Delta\phi) - \sin\phi_0 \simeq \cos\phi_0\,\Delta\phi$$

and Eq. (10) becomes

$$\frac{d^2}{dt^2}\Delta\phi + \omega_s^2\Delta\phi = 0 \quad \text{with} \quad \omega_s^2 = \frac{h\eta\omega_{\mathrm{rev}}qV}{2\pi pR}\cos\phi_0\,. \tag{11}$$

This is the well known equation of a harmonic oscillator and the frequency of the longitudinal oscillations, ω_s, is called synchrotron frequency.

As will be shown in Sec. 6.5, Eqs. (7) and (9) also define the aspect ratio of the bunch distribution important for the longitudinal matching.

4 Bunch-to-bucket transfer between circular accelerators

For the transfer of bunched beams, the particle distribution from the sending accelerator should be placed in time and energy into the centre of a bucket, the stable region in time-energy phase space (Fig. 12). This requires alignment of the energy of both accelerators, the so-called energy matching, as well as

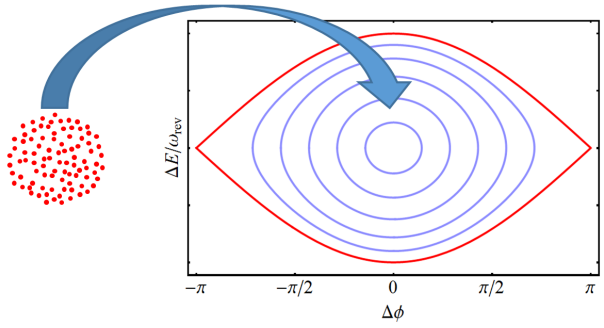


Fig. 12: The bunch-to-bucket transfer places the particle distribution from the sending accelerator into ideally matched RF buckets in the receiving one.

the phase alignment of the RF buckets. A single bunch or a train of bunches can be injected into any bucket, as long as the receiving accelerator has no beam yet and the bunch train fits in terms of length. However, for multiple transfers the bucket number must be carefully controlled to avoid over-injecting

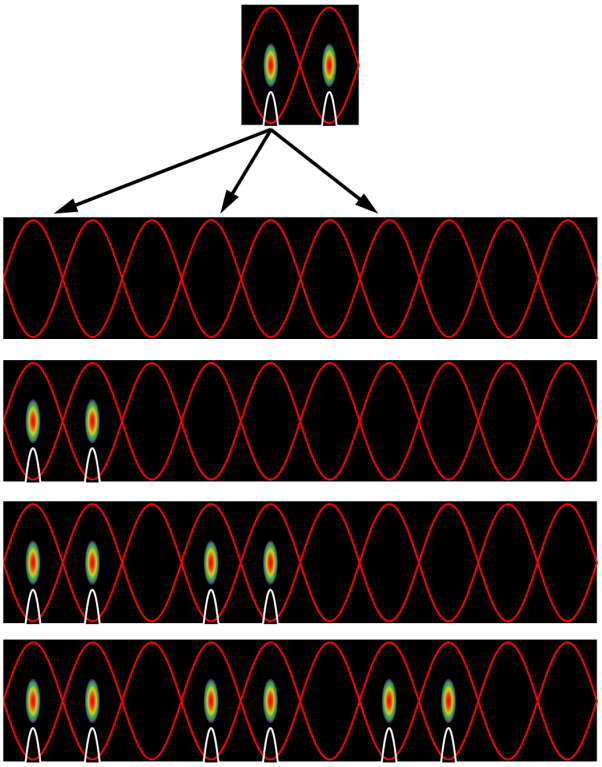


Fig. 13: Longitudinal stacking by bunch-to-bucket transfer of pairs of bunches.

unintentionally into a bucket which is already occupied by a bunch. In Fig. 13 the three injections of two bunches are identical with the exception of the bucket number in the receiving accelerator. The beam structure of filled and empty buckets in an accelerator is referred to as filling pattern.

For the following considerations a synchronized bunch-to-bucket transfer between two circular accelerators is assumed.

4.1 Fundamental periodicity

Before the beam transfer between two circular accelerators, their revolution periods, $T_{\mathrm{rev},1}$ and $T_{\mathrm{rev},2}$, must be synchronized. Though rather common, an integer ratio of revolution periods is generally not required. The ratio can also be fractional such that

$$T_{\mathrm{rev},2} = \frac{n_1}{n_2} T_{\mathrm{rev},1} \quad \leftrightarrow \quad n_2 T_{\mathrm{rev},2} = n_1 T_{\mathrm{rev},1} \equiv T_{\mathrm{common}} = T_{\mathrm{fiducial}}$$

The time period with which the whole system of both beams repeats is often referred to as super-period, common or fiducial period. Hence a beam transfer can take place under the same conditions every n_1 turns of the accelerator with a revolution period of $T_{\mathrm{rev},1}$ which is identical to n_2 turns of a beam at $T_{\mathrm{rev},2}$. Integer revolution period ratios simplify the transfer which can then take place at any turn of the larger of the two accelerators.

Table 3 summarizes the revolution frequency ratios for some accelerator facilities. An example for two electron synchrotrons with a non-integer revolution frequency ratio are booster and storage ring at BESSY II as illustrated in Fig. 14. All signals are directly derived from a reference generator at close to $f_{\mathrm{RF}} = 500$ MHz, the main RF frequency of both booster and storage ring. Following a frequency division by 16, the resulting signal at 31.3 MHz is then synchronously divided by $10, 25$ and 50 to obtain the revolution frequency of booster and storage ring, as well as the frequency of the super-period.

[1] The phase slip factor is sometimes defined with the opposite sign, e.g. [24].

Table 3: Revolution frequency ratios of selected accelerator facilities.

BESSY Booster	BESSY SR	$2/5$	Fixed frequency
SLS Booster	SLS SR	$15/16$	Fixed frequency
J-PARC RCS	J-PARC MR	$2/9$	Profit from ratio for bucket selection
GSI SIS12/18	GSI ESR	$2/1$	
CERN PS Booster	CERN PS	$1/4$	
CERN PS	CERN SPS	$1/11$	
CERN PS	CERN AD	$3/1$	Particle type and energy change at transfer
CERN SPS	LHC	$7/27$	$f_{\mathrm{common}} = 1/T_{\mathrm{common}}$ as low 1.6 kHz

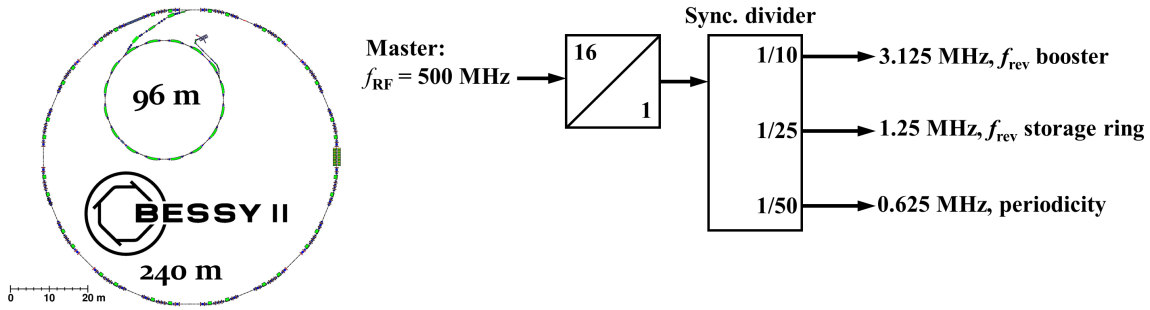


Fig. 14: Generation of beam synchronous frequencies at the BESSY facility (courtesy of BESSY).

The periodicity is illustrated in Fig. 15 for the example of the BESSY II facility with a revolution frequency ratio of $2/5$. With every half turn of the bunch in the booster synchrotron the bunch in the storage ring advances by $1/2 \cdot 2/5 = 1/10$th of a turn. As expected, the azimuthal position of the bunches is identical every 5 turns of the booster, equivalent to every 2 turns in the storage ring.

4.2 Beam synchronous triggers

As mentioned above several devices need to be triggered shortly before, during or after the beam transfer, in particular fast kicker magnets and fast beam instrumentation. These beam synchronous triggers can be generated by chains of counters counting beam synchronous clocks. Such a cascade of counters is sketched in Fig. 16. The transfer process is launched by the master start which does not need be to synchronous to anything. It is just a time marker to indicate the upcoming transfer. The first counter in the chain counts the super-period. The value to which is counts is not relevant, but the fact that the trigger $\#1$ is generated based on the super-period clock, i.e., the fundamental periodicity of the transfer. Since all clocks are derived from a common refrence, the super-period clock is synchronous with the revolution and RF clocks. One can hence simply chain a counter on the revolution frequency $\#2$ and a further one on the RF frequency. This allows to generate an unambiguous trigger $\#3$ anywhere in time with a granularity of the revolution or RF period and corresponds to the bucket number. This scheme allows to select a specific turn and bucket number.

For a beam transfer between two accelerators, such triggers must be generated on the sending and on the receiving side both of which may, in large facilities, be several kilometres apart from each other. This is easily achieved by adding parallel branches to the timing cascade (Fig. 17). As the counters in sending and receiving accelerator are started with a common trigger synchronous to the super-period clock and, since all clocks (super-period and revolution frequency clocks) have a fixed phase relationship, the delay between the triggers $\#2$ and $\#3$ in Fig. 17 is reproducible. Again counters using the RF clock allow the adjustment of bucket numbers.

Thanks to the practically constant revolution frequencies of electrons, the basic ingredients to generate all beam synchronous triggers are complete for the transfer of beams between circular lep-

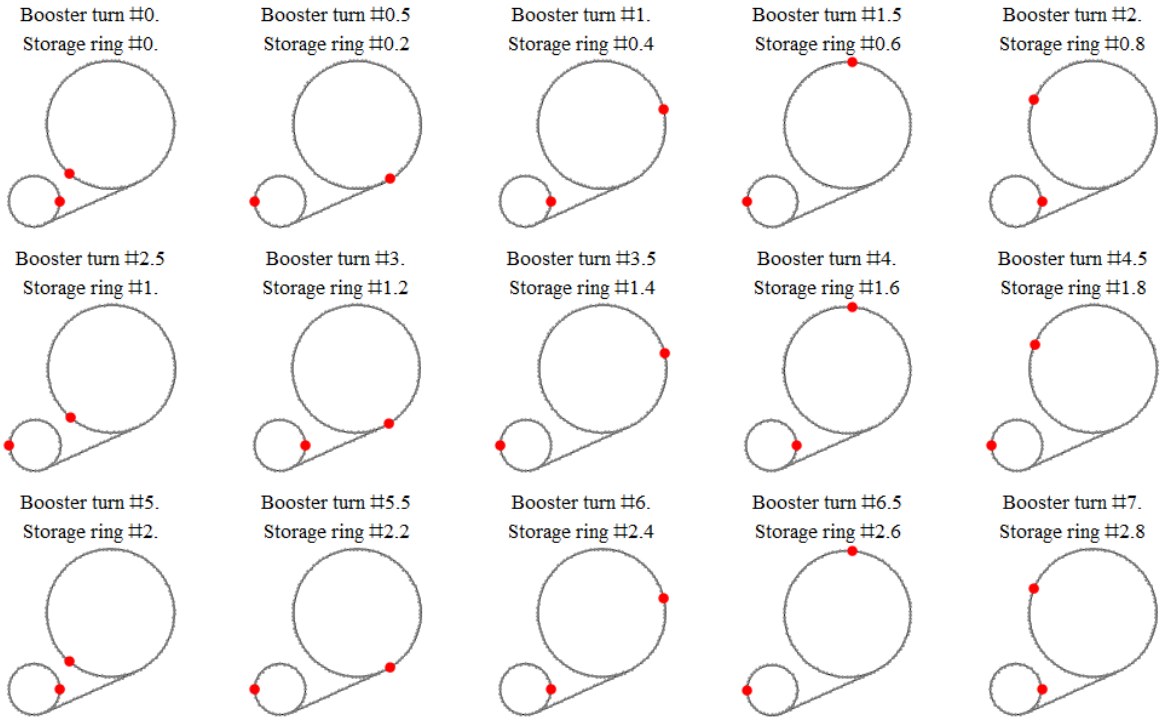


Fig. 15: Bird's eye view of the facility Fig. 14, with a revolution frequency ratio of $2/5$. Snapshots are taken every half turn of the booster synchrotron. Note that in BESSY the beams in both accelerators circulate clockwise.

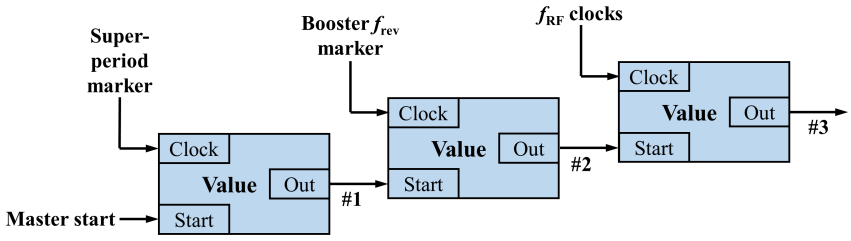


Fig. 16: Generation of beam synchronous triggers.

ton accelerators. Additionally, due to the strong synchrotron radiation damping at electron energies in the GeV regime, the precision in time is usually not very demanding. As long as the bunch is injected into the correct bucket it will be forced into the centre of the bucket by synchrotron radiation.

All signals can be derived from a common, fixed-frequency master oscillator which makes the entire transfer process fully deterministic in lepton accelerators. Starting from an event triggering the transfer, all beam synchronous signals for kicker magnets or fast diagnostics to monitor the transfer, all timing triggers are derived by counting the RF master clock or a one of its sub-multiples. Everything is predictable from the beginning, which significantly simplifies the synchronisation and timing system for the beam transfer compared to the one of hadron accelerators.

5 Transfer between hadron accelerators

Due to the slower velocity of hadrons, the revolution frequency changes during acceleration by a considerable amount. Especially in low energy hadron synchrotrons it may sweep by several octaves. In the Extra Low ENergy Antiproton (ELENA) ring at CERN it changes by almost one order of magnitude, whereas the relative particle velocity in the LHC only increases by a few 10^{-6} during acceleration.

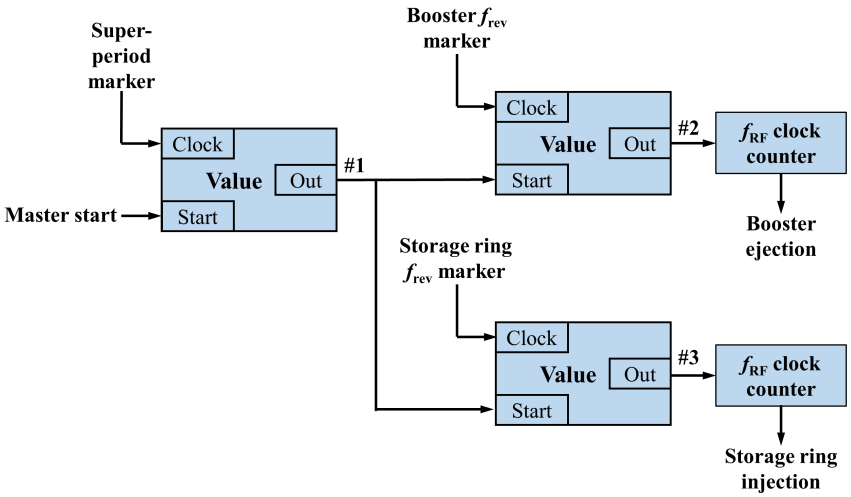


Fig. 17: Generation of multiple beam synchronous triggers.

With the revolution frequency sweep, also all counters generating beam synchronous timings for kickers and beam instrumentation must be supplied with sweeping clocks. The delay of cables and electronics, τ, leads to a frequency dependent phase according to $\Delta\phi = 2\pi h f_{\mathrm{rev}} \tau$ for any signal at the hth harmonic of the revolution frequency. The exact revolution frequency is moreover derived from the beam, via the beam phase loop.

The absence of synchrotron radiation damping also poses more stringent requirements on the transfer. It is not sufficient to place the bunch into the correct bucket and let it move to the exact energy and phase of the bucket in the receiving accelerator. The injected hadron bunch must be placed into the centre of the bucket in terms of phase and energy instead (Fig. 12). Additionally, the aspect ratio of the injected bunch, the ratio between bunch length and energy spread, must correspond to the aspect ratio of the bucket in the receiving accelerator. This can be achieved by longitudinal matching (Sec. 6.5).

The harmonic number during the acceleration may be changed with so-called RF manipulations. In the presence of two RF harmonics simultaneously a beating pattern of the effective RF voltage depending on the azimuth position of the bunches is generated. This splits the behaviour for different bunches, thus requiring careful counting of bucket numbers.

5.1 Beam phase loop

Due to the absence of synchrotron radiation damping, the bunch in a hadron accelerator does not match itself to the bucket. Any error in phase or energy will cause longitudinal oscillations which, due to the non-linearity of these synchrotron oscillations, lead to uncontrolled growth of the spread of the particle distribution.

In hadron synchrotrons the beam itself is therefore used as a reference for the phase of the RF voltage which keeps it bunched and accelerates it. This is an application of the PLL introduced in Sec. 2.5. Figure 18 shows a simplified diagram of a PLL in the configuration of a beam phase loop. The beam phase is compared with the phase from a probe of the voltage in the accelerating cavity. This phase difference gives the longitudinal position of the particle bunch with respect to the RF bucket. As the reference phase of the bucket intentionally moves during acceleration, the synchronous phase, ϕ_s, must be subtracted from the measured phase difference between beam and acceleration RF signals. This bunch-to-bucket phase error then pilots, via a loop filter, the frequency correction of a precision VCO generating the signal at $f_{\mathrm{RF}} = h f_{\mathrm{rev}}$ driving the power amplifier. The configuration of the PLL is hence the same as sketched in Fig. 11, but with the beam replacing the input signal.

To minimize the errors to be corrected by the loop, the VCO is implemented as an oscillator

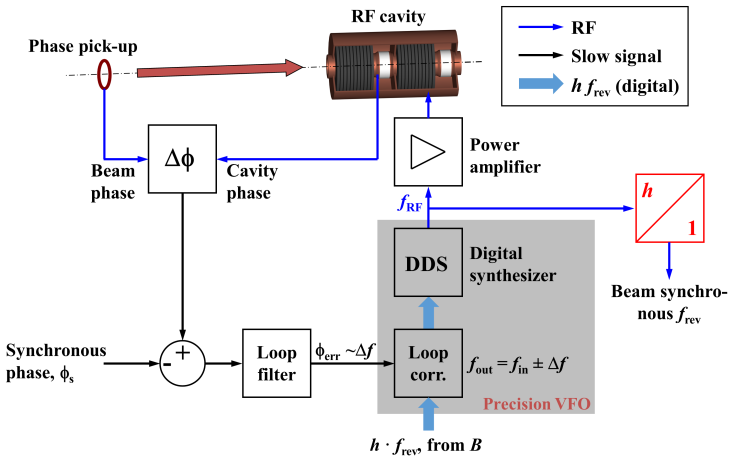


Fig. 18: Typical overview diagram of a beam phase loop.

which can only be tuned in a small range around the expected revolution frequency of the beam. This expected revolution frequency can be derived from the knowledge of the magnetic bending field defining the energy of the accelerator.

Compared to the standard electronic phase-locked loop (Sec. 2.5) the beam behaviour now contributes to the transfer function of the beam phase loop. It can be included by calculating the response of the beam in phase to a change of frequency driven by the RF system.

As shown in Sec. 3 the energy offset change for small deviations in $\Delta\phi$ with respect to the synchronous phase can be approximated, and the equations of motion become

$$\frac{d}{dt}\Delta E = qV\frac{\omega_{\mathrm{rev}}}{2\pi}\cos\phi_0\,\Delta\phi \tag{12}$$

$$\frac{d}{dt}\Delta\phi = -\frac{h\eta}{pR}\Delta E + \delta\omega_{\mathrm{RF}}\,. \tag{13}$$

A small frequency modulation by the RF system contributes as a perturbation term, $\delta\omega_{\mathrm{RF}}$, and is added in Eq. (13) to the phase change from the energy offset [25].

To extract the longitudinal beam transfer function in frequency domain, the equations can then be solved by an ansatz $\Delta\phi(t) = \Delta\phi e^{i\omega t}$ and $\Delta E(t) = \Delta E e^{i\omega t}$. The equations of motion then simplify to

$$i\omega\Delta E = qV\frac{\omega_{\mathrm{rev}}}{2\pi}\cos\phi_0\,\Delta\phi$$

$$i\omega\Delta\phi = -\frac{h\eta}{pR}\Delta E + \delta\omega_{\mathrm{RF}}$$

in frequency domain, which can be solved to

$$\Delta\phi = \frac{i\omega}{\omega_s^2 - \omega^2}\delta\omega_{\mathrm{RF}}\,. \tag{14}$$

This is the response of the beam in phase to a perturbation of the driving RF frequency. Replacing $s = i\omega$, the beam transfer function hence becomes

$$H(s) = \frac{\Delta\phi}{\delta\omega_{\mathrm{RF}}} = \frac{s}{\omega_s^2 + s^2}\,. \tag{15}$$

Dividing the RF frequency from the precision VCO (Fig. 18, divider marked in red) by the harmonic number, h, results in a revolution frequency signal with fixed relationship with respect to the beam and maintained by the beam phase loop. It can be used to clock counters generating beam synchronous triggers, and it represents the longitudinal beam position for a synchronization loop discussed below.

5.2 Beam synchronous revolution frequency

Deriving a beam synchronous revolution frequency by division (Fig. 18) from an RF signal locked to the beam introduces the ambiguity of the phase depending on the counting state of the divider (Sec. 2.3). The most common approach to avoid this ambiguity is counting RF cycles correctly right from the beginning. With the first injection the beam is already transferred with a reproducible delay with respect to the output of the revolution frequency divider.

Moving the beam in time with respect to the beam synchronous revolution frequency can be easily achieved by simply inserting a delay into the branch to the sending accelerator (Fig. 19, left). Since

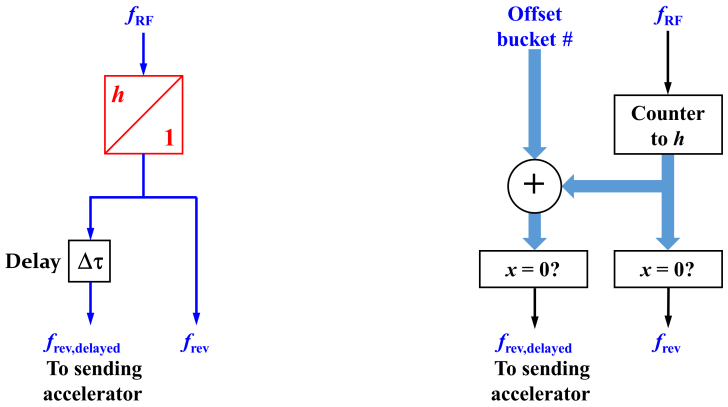


Fig. 19: Set-ups to achieve a reproducible phase relationship between the beam synchronous revolution frequency and the injected beam. Technically the delay (left) is mostly implemented with counters allowing to comfortably adjust the arrival time of the incoming beam in units of RF buckets.

the delayed revolution frequency, $f_{\mathrm{rev,delayed}}$, to the upstream accelerator and the internal revolution frequency, f_{rev}, are derived from the same source, after the frequency division, both have an unambiguous phase relationship by definition.

Adjustable delays are technically difficult to realize, in particular for large accelerators, where one turn can take several tens of microseconds. The delayed revolution frequency can be generated by an offset counter (Fig. 19, right) more easily, as introduced in Sec. 2.3. The offset is programmed in units of input clock cycles of the counter, which is usually the bucket number when counting f_{RF}.

For multiple injections, the bucket number offset is simply changed in between transfers, such that multiple injections with a programmable filling pattern can be be achieved. The beam synchronous revolution frequency will always have a reproducible delay or phase with respect to the circulating beam. It therefore indicates the longitudinal position of the beam.

5.3 Bucket numbering

Keeping track of the position of occupied and empty RF buckets is essential to prepare an intended filling pattern in the case of multiple injections. A newly injected bunch should not be placed in time into an already occupied bucket, as the circulating bunch would be removed by the injection kicker and replaced by the incoming one.

Measuring whether a bucket is empty or occupied by a bunch must not be necessary in the framework of the beam transfer process. Using beam signals do derive triggers for the beam transfer is not appropriate since a bunch may be missing [26] or may have an intensity or profile which would now allow a correct detection.

Bunches are always injected with a well defined delay with respect to a beam synchronous revolution frequency. The beam synchronous revolution frequency itself has a reproducible delay with respect

to the circulating beam since every injection is performed using it as a reference for the longitudinal beam position. Hence, knowing the sequence of injected bunches, allows to keep track of the filling pattern in a synchrotron.

In an accelerator with a single harmonic RF system, and neglecting any intensity effects, all RF buckets are equal. This changes when applying two or more RF harmonics simultaneously. The greatest common divider of all involved RF harmonics defines the frequency with which the sum RF voltage is periodic. In the extreme case, when the harmonic numbers have no common divider other than one, the total RF voltage is only periodic with the revolution frequency. Consequently, the size and form of all RF buckets becomes different.

The longitudinal phase space for such a case is shown in Fig. 20, where the RF voltage is the same at two harmonics, $h = 9$ and $h = 10$ [27]. This bucket configuration is present in the middle of a

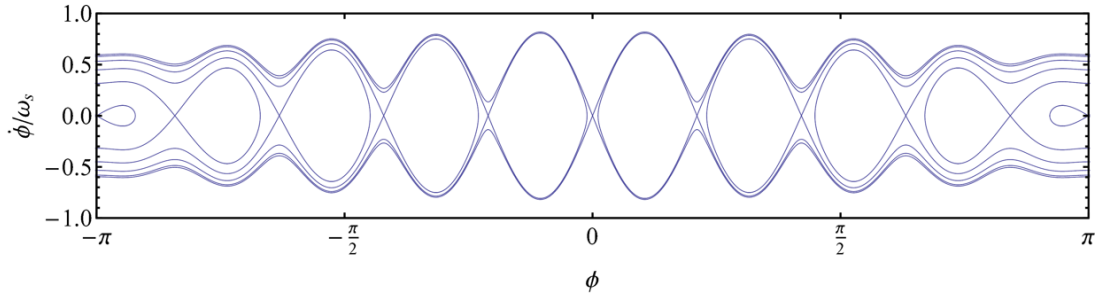


Fig. 20: Normalized longitudinal phase space for a double harmonic RF system with equal voltage at both harmonics, $h = 9$ and $h = 10$.

so-called batch compression process in the PS at CERN. The spacing between bunches can be reduced by incrementally moving to a larger RF harmonic. This happens at the cost of introducing an additional RF bucket at an azimuth which should not be occupied by particles (Fig. 20, at $\phi = \pm\pi$).

Due to the periodicity of the RF voltage with the revolution frequency, there are nine possibilities of injecting eight bunches into the nine buckets (Fig. 21, left). Only one of them (right) keeps the empty bucket at the location where the additional bucket is introduced during the $h = 9$ to $h = 10$ handover. This example at the beam transfer between PS Booster (PSB) and PS illustrates the importance of bucket numbering.

In accelerators performing RF manipulations like the PS at CERN, the number of bunches may change during the acceleration cycle as the one shown in Fig. 21. Depending on the type of manipulation, even the position of the first bunch with respect to the beam synchronous revolution frequency changes. Therefore, prior to ejection, the bucket numbering must be rearranged to realign, e.g., the position of the first bunch with respect a phase reference given by the receiving synchrotron.

Renumbering the buckets for extraction practically means to shift signal phases such that the RF synchronization process will move the beam to the intended, reproducible phase with respect to the external reference marker from the receiving accelerator. As will be shown in Sec. 6.3, a reference revolution frequency, marking the expected position of the beam after synchronization, is compared in phase to the internal beam synchronous revolution frequency. Acting on the phase of either of these two signals moves the relative delay of the marker from the receiving accelerator and the extracted beam.

For the example of the transfer between PS and SPS at CERN, where the SPS as a receiving synchrotron defines the azimuth position of the bunches, it turns out to be technically more easy to move the phase of the reference for the synchronization. The simplified set-up is sketched in Fig. 22. A synchronous frequency divider (Sec. 2.3) generates a signal at the revolution frequency of the PS to serve as a reference for the RF synchronization. It is compared in phase with the beam synchronous revolution frequency. The reset for the divider is generated from the revolution frequency marker from the SPS,

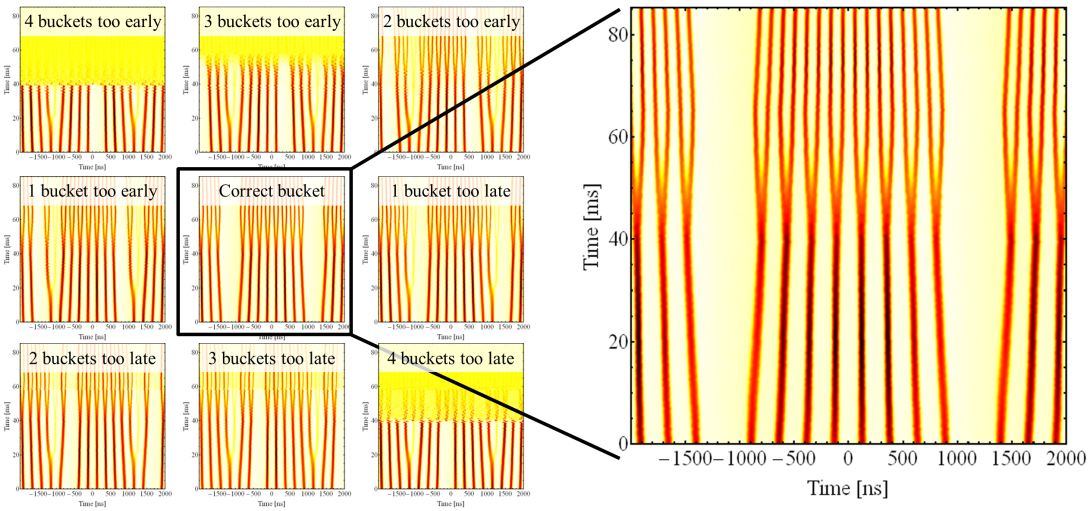


Fig. 21: Measured mountain range plots when injecting eight bunches into $h = 9$ in the PS and moving them from harmonic $h = 9$ to 10 (first 40 ms) and further via $h = 20$ to 21. The RF voltages at $h = 9$ and $h = 10$ are equal at 20 ms after the start of the manipulation and he longitudinal phase space corresponds to the one illustrated in Fig. 20. The azimuth of the incoming bunches is moved with respect to the RF buckets by multiples of $h = 9$ buckets, $n \cdot 2\pi/9$, $n = 0 \ldots 8$.

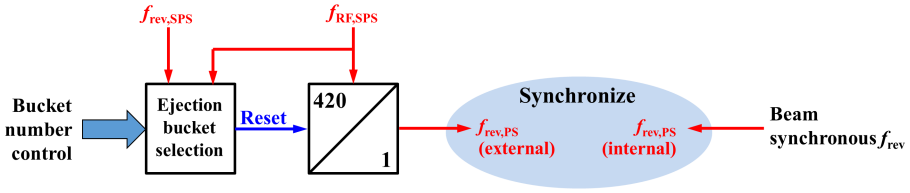


Fig. 22: Simplified diagram of the ejection bucket numbering system between PS and SPS. The reference revolution frequency for the RF synchronization is shifted in phase with respect to the revolution frequency marker from the SPS.

but delayed by a programmable number of RF clocks, again from the SPS. This shifts the reference revolution frequency in programmable steps of SPS buckets, and with it the phase position of the beam in the PS with respect to the revolution frequency marker from the SPS. This mechanism allows to adjust the phase position of the outgoing bunches of the extracted such that the position of the bucket with, e.g., the first bunch is aligned with respect to the expected position, independent of RF manipulation or injection bucket numbering in the PS.

6 Synchronization and transfer

The bunch-to-bucket transfer of beam between circular hadron accelerators usually follows a sequence of energy matching, RF synchronization and further actions directly related to the transfer. These steps are summarized in Tab. 4 in chronological order [28]. The first two steps prepare the bunch-to-bucket transfer. Thereafter the transfer can take place with every cycle of the fundamental periodicity (Sec. 4.1). The remaining steps are all related to the transfer itself, trigger elements required for the extraction and injection bumps, the septa. The last step is the generation of triggers for the ejection and injection elements which guide the beam to the transfer line trajectory at the exit of the sending accelerator and back to the trajectory for the circulating beam in the receiving accelerator.

Table 4: Chronological order of steps relevant to synchronization and timing at beam transfer.

1.	Set bending fields in both accelerators to the same magnetic rigidity Prepare transfer line
2.	RF synchronization of sending or receiving accelerator
	→ **Ready for transfer**
3.	Start counting clock of fundamental periodicity Trigger bump and septum elements
4.	Start counting f_{rev} clock (sending and receiving accelerator idependently) Start counting bucket clock
5.	Fine delays started Ejection and injection kicker triggers
	→ **Transfer**

6.1 Matching of bending strength

Matching the bending field in both accelerators to the same magnetic rigidity is a first precondition to prepare the beam transfer [29].

To keep charged particles on a circular trajectory, the centripetal force, F_Z, must be compensated by the Lorentz force, F_L. This equality can written such that the particle properties momentum, p, and charge, q, are separated from the accelerator parameter bending radius, ρ and strength, B:

$$F_Z = F_L \quad \rightarrow \quad \frac{p}{q} = \rho B \,. \tag{16}$$

As long as the energy, charge state and type of the particle remain unchanged during the transfer, the magnetic rigidity must be the made identical in sending and receiving accelerator,

$$\rho_1 B_1 = \rho_2 B_2 \,. \tag{17}$$

Exceptions are transfers with particle type change in the line between both accelerators. Anti-protons are produced by conversion on a target between the accelerator for the primary beam and the one collecting the anti-protons from the debris produced at the target. At Fermilab a proton beam at an incident momentum of $p = 120$ GeV/c was producing anti-protons at 8 GeV/c. The Anti-proton Decelerator (AD) at CERN receives anti-protons at a momentum of $p = 3.6$ GeV/c produced by a primary proton beam at 26 GeV/c from the PS. A different magnetic rigidity is also needed when the charge state changes during beam transfer. In the ion injector chain at CERN charge state $^{208}\mathrm{Pb}^{54+}$ in the Low Energy Ion Ring (LEIR) and PS is stripped to $^{208}\mathrm{Pb}^{82+}$ in the transfer line between PS and SPS. In addition to the change of charge state, the stripping process also causes an average energy loss of the beam. The magnetic rigidities are programmed according to Eq.(16).

6.2 Matching RF frequencies and bunch distance

Besides the bending fields, also the RF frequencies in both accelerators must have a distinct ratio. Assuming that the beam velocity is unchanged during the transfer and using $f_{\mathrm{rev}} = f_{\mathrm{RF}}/h = \beta c/(2\pi R)$, the condition for the ratios of RF frequencies, harmonic numbers and average radii becomes

$$\beta c = 2\pi R_1 f_{\mathrm{rev},1} = 2\pi R_2 f_{\mathrm{rev},2} \quad \rightarrow \quad R_1 \frac{f_{\mathrm{RF},1}}{h_1} = R_2 \frac{f_{\mathrm{RF},2}}{h_2} \,. \tag{18}$$

A common choice, in particular for most electron accelerators, is to operate sending and receiving accelerator at the same RF frequency, $f_{\mathrm{RF},1} = f_{\mathrm{RF},2}$ (Fig. 23, left) at transfer. The harmonic number simply becomes proportional to the ring circumference, $2\pi R$. Again, Eq. (18) is only valid for fixed velocity

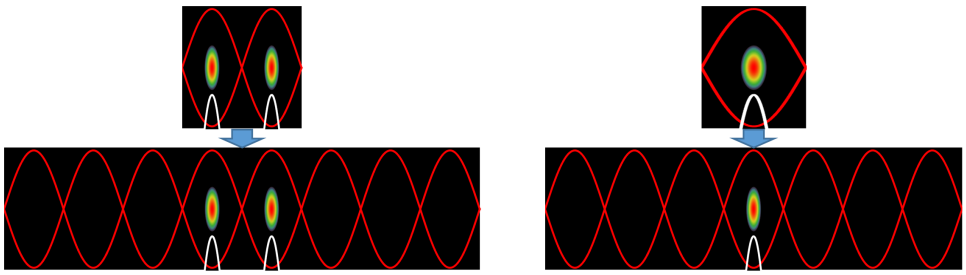


Fig. 23: Bunch-to-bucket transfer with equal (left) and unequal (right) RF frequencies.

during the transfer, which is not the case, e.g., for the production of anti-protons at a target in the transfer line. The velocity factors β_1 and β_2 must be added in the denominators of Eq. (18).

Next to the RF frequencies also the distance between bunches in the sending accelerator must match the distance of the buckets in the receiving one to transfer multiple bunches with with a single transfer. An integer ratio of RF frequencies is common in cascades of hadron synchrotrons. At the transfer of protons for the LHC between PS and SPS, the RF frequency in the SPS, $f_{\mathrm{RF,SPS}} = 200.3$ MHz, is five times the principal RF frequency in the PS of about 40.1 MHz, while at transfer to the LHC, the RF frequency in the LHC is twice larger than the one in the SPS. By definition, the PS can hence fill every 5th bucket in the SPS, and the bunches from the SPS can only occupy every second bucket of the LHC. The accelerator chain of PS and SPS consequently transfers one bunch into every 10th bucket of the collider.

Of course the bunch distance is only relevant for the transfer of more than one bunch. This gives more flexibility at the transfer from the PS Booster (PSB) to the PS at CERN. While the PS is four times larger in circumference than the PSB, suggesting that $h_{\mathrm{PS}} = 4h_{\mathrm{PSB}}$, many more combinations of harmonic numbers are possible thanks to the transfer of single bunches, e.g., $h_{\mathrm{PS}}/h_{\mathrm{PSB}} = 7$ or 9.

A further exception is the transfer using multi-harmonic RF systems in one of the accelerators. In the PSB, two bunches at $h_{\mathrm{PSB}} = 2$ with a natural distance for an injection into the PS at $h_{\mathrm{PS}} = 8$, have been pushed further apart to inject them into $h = 7$ buckets in the PS [30].

6.3 RF synchronization

Even with the magnetic rigidity set according to Eq. (16), the beam derived revolution frequencies eventually will not be at their theoretical ratio due to imperfections of bending field strength and length of beam orbit. These errors lead to a slippage of the bunches with respect to the positions of the buckets, which is illustrated in Fig. 24. The bucket positions in the sending accelerator, projected into the receiving one (Fig. 24, grey dots), move in time with respect to the real bucket positions (Fig. 24, purple dots). This phase slippage is caused by the revolution frequencies not being exact integer or rational multiples.

Before starting the RF synchronization to lock the revolution frequencies of two accelerators [31], it must be decided if the sending synchrotron should be synchronized to the receiving one or vice versa.

For a bunch-to-bucket transfer with the sending accelerator as master for the beam transfer, revolution frequency and bucket positions in the receiving adapt to the incoming beam. This is a common choice when the receiving accelerator has no beam before the transfer. An example is the injection of anti-protons into the AD ring at CERN which are generated by an incident proton beam from the PS.

However, for multiple transfers there is little interest in synchronising the receiving accelerator, as every synchronisation process requires to move all bunches to the revolution frequency of the newly incoming beam, perturbing the accumulated bunches more than necessary. Therefore the receiving accelerator is commonly chosen as master whenever it already has beam before the transfer and in particular for the case of multiple transfers. Only the newly injected beam needs to be synchronized and may suf-

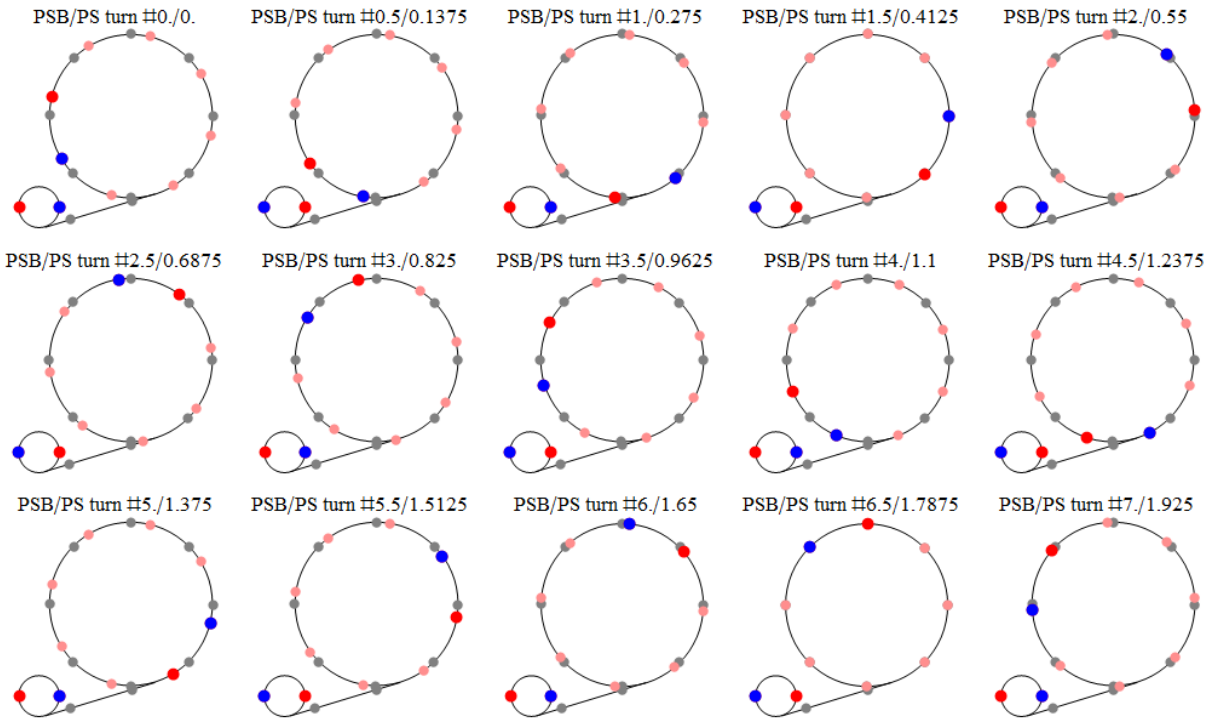


Fig. 24: Bird's eye view of the PSB ($h_{\mathrm{PSB}} = 2$) to PS ($h_{\mathrm{PS}} = 8$) transfer before the RF synchronization. The circulating bunches are indicated by blue and red dots. Empty bucket positions are marked as grey dots in the sending accelerator and purple in the receiving one.

fer once from the synchronization process. Multiple re-synchronizations of the accumulated beam are avoided. It is the most common choice and implemented in, e.g., the injector chain of the LHC. Without loss of generality, the receiving accelerator is considered as the master of the beam transfer in what follows.

Even though the relative phase of beams moves corresponding to revolution frequency errors (Fig. 24) [32], this slippage is usually too slow to just wait until the intended phase relationship is reached for a possible transfer. Firstly, the beam to be synchronized is therefore, at constant bending field, $dB/dt = 0$, slightly accelerated or decelerated to an off-momentum orbit according to Eqs. (8) and (26)

$$\frac{df}{f} = \frac{\gamma_{\mathrm{tr}}^2 - \gamma^2}{\gamma^2 \gamma_{\mathrm{tr}}^2} \frac{dp}{p} ,$$

which establishes a defined frequency difference between the accelerators. Secondly, the azimuth error is measured to close a synchronization loop when the beam is at the intended azimuthal position.

As introduced in Sec. 2.4, a constant frequency error between to RF signals translates into a linearly changing phase difference (Fig. 25), as measured at the phase detector comparing the reference azimuth and the actual beam azimuth. Once the reference phase position is reached, the synchronization loop closes by acting on the RF frequency of the synchronized accelerator. Azimuth, revolution and RF frequencies are then locked with respect to the external reference.

An example of the phase measurement between reference and actual beam position during the synchronization of the bunches in the SPS prior to their transfer to the LHC [33] is illustrated in Fig. 26, together with the synchronization steps and their effect on the phase of the beam. Following the arrival at the flat-top, the revolution frequency of the SPS is moved to $f_{\mathrm{RF,SPS}} = 27/7 f_{\mathrm{RF,SPS}}$. Both beams in SPS and LHC are then circulating according to the fundamental periodicity Eq. (4.1) such that $T_{\mathrm{common}} = 27 T_{\mathrm{rev,SPS}} = 7 T_{\mathrm{rev,LHC}}$. The azimuth error of the whole beam in the SPS is measured and a frequency

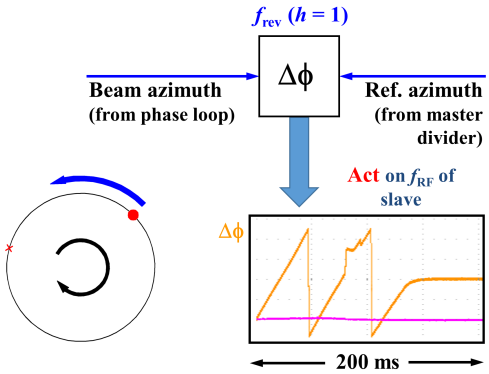


Fig. 25: Synchronization loop to move the beam to an intended azimuthal position.

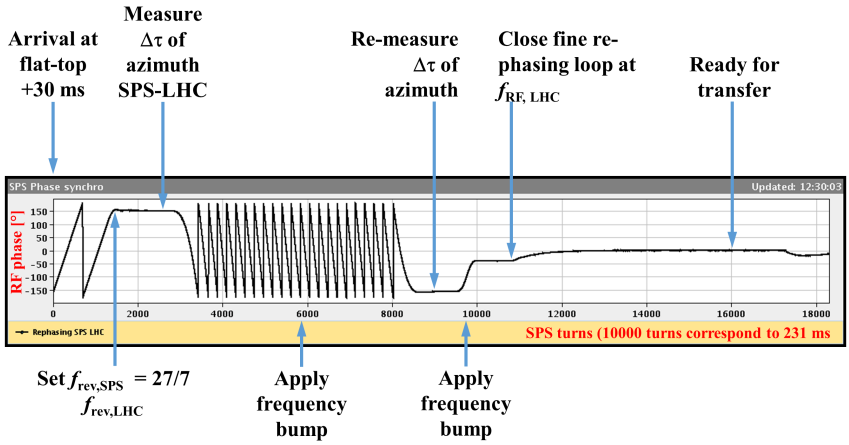


Fig. 26: Phase detector at twice the RF frequency in the SPS during the synchronization before the transfer to the LHC.

bump is generated to move the actual to the reference azimuth. This sequence of measurement and frequency bump repeats until the beam has been moved to the reference azimuth within an accuracy of much better than the azimuth $2\pi/h$ of one RF bucket.

To then precisely align the bunch centres with respect to the position of the bucket centers in the LHC, a synchronisation loop compares the reference RF at about 400.8 MHz from the LHC with twice the RF frequency (about 200.4 MHz) in the SPS and locks them together.

Following this synchronization process, which takes about half a second, the beam in the SPS circulates synchronously and at the correct azimuth position with respect to the beam in the LHC. The transfer process can be launched with every period of the fundamental periodicity.

Putting all relevant parts of the RF synchronization system together, the synchronization chain between outgoing beam and the reference marker signals at which the beam is actually expected, is illustrated in Fig. 27. The beam to be extracted is locked at a reproducible phase relationship with

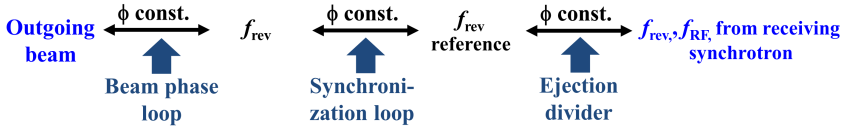


Fig. 27: Synchronization chain from the reference signal of the receiving accelerator to the extracted beam of the sending one.

respect to the beam synchronous revolution frequency, f_{rev} by the beam phase loop (Sec. 5.1). This beam synchronous revolution frequency is itself kept at a constant phase with respect to the external reference revolution frequency by one or several synchronization loops. Finally, the external revolution frequency for the sending synchrotron is generated from the shifted reference RF and revolution frequency signals of the receiving synchrotron, again at a fixed phase relationship. In summary, the beam in the sending accelerator and the reference marker from the receiving one are in a reproducible phase relationship. By shifting its reference signals the receiving synchrotron controls when in time the beam must be delivered by the sending synchrotron. This concept allows complex filling patterns generated under full control of the receiving accelerator. For the sending synchrotron all beam transfers are simply the same.

Once the RF synchronization completed the case of two proton accelerators is exactly identical to the case of two electron accelerators prior to the transfer, with an integer or rational revolution frequency ratio (Fig. 15). The beams are synchronized and can be extracted with every marker of the fundamental periodicity.

6.4 Energy matching

In an ideal synchrotron the beam circulates exactly at the expected revolution frequency, $\Delta f_{\mathrm{rev}} = 0$, on the central orbit, $\Delta R = 0$ and it has the reference momentum, $\Delta p = 0$. A real accelerator has errors on all of these parameters.

6.4.1 General orbit equation and differential relations

According to Eq. (16) a particle on the central orbit (index zero) has a momentum given by

$$p_0 = qB\rho_0\,. \tag{19}$$

For a particle with a momentum offset, the orbit length is defined via the transverse optics by the momentum compaction factor at constant magnetic field [34]

$$\alpha = \frac{1}{\gamma_{\mathrm{tr}}^2} = \frac{\partial L/L}{\partial p/p} = \frac{\partial R/R}{\partial p/p} = \frac{p}{R}\frac{\partial R}{\partial p}\,, \tag{20}$$

where $L = 2\pi R$ is the circumference. Integration of Eq. (20) results in

$$\ln R = \alpha \ln p + c_1 \quad \Leftrightarrow \quad R^{1/\alpha} = p\, e^{c_1/\alpha} = c_2 p \quad \text{with} \quad c_2 = e^{c_1/\alpha}\,. \tag{21}$$

The variables c_1 and c_2 are arbitrary integration constants.

The latter constant c_2 can be derived by keeping in mind that Eq. (21) must again be fulfilled by the reference particle with a momentum, p_0, given by Eq. (19), hence

$$R^{1/\alpha} = c_2 p = c_2 qB\rho_0 \Leftrightarrow \quad c_2 = \frac{R_0^{1/\alpha}}{qB\rho_0}\,.$$

The circumference for the reference particle is $2\pi R_0$. This leads to the generalization of Eq. (19) for a particle at any average radius [34, 35]

$$p = qB\rho_0 \left(\frac{R}{R_0}\right)^{1/\alpha}\,. \tag{22}$$

Based on this general equation, a set of differential relations can be derived for particles with a small derivation from the reference parameter as summarized in the table below [34].

Variables	Relation	
B, p, R	$\frac{dp}{p} = \gamma_{\mathrm{tr}}^2 \frac{dR}{R} + \frac{dB}{B}$	(23)
f, p, R	$\frac{dp}{p} = \gamma^2 \frac{df}{f} + \gamma^2 \frac{dR}{R}$	(24)
B, f, p	$\frac{dB}{B} = \gamma_{\mathrm{tr}}^2 \frac{df}{f} + \frac{\gamma^2 - \gamma_{\mathrm{tr}}^2}{\gamma^2} \frac{dp}{p}$	(25)
B, f, R	$\frac{dB}{B} = \gamma^2 \frac{df}{f} + (\gamma^2 - \gamma_{\mathrm{tr}}^2) \frac{dR}{R}$	(26)

The index for the reference particle has been dropped since, e.g., $R = R_0 + \Delta R \simeq R_0$.

6.4.2 Application to energy matching

Each of the differential Eqs. (23) to (26) depends on three variables. It means that the choice of two parameters from the set B, p, R and f automatically constrains all others in both accelerators. By fixing, e.g., the bending field in a synchrotron, $dB = 0$, the revolution frequency with the radial position offset of the beam are linked directly by the relation

$$\gamma^2 \frac{df}{f} = (\gamma_{\mathrm{tr}}^2 - \gamma^2) \frac{dR}{R} \,.$$

Without changing the magnetic field the beam can thus not circulate at the central orbit and at a given revolution frequency simultaneously. The bending field must be changed in this case to match it.

The energy matching of an incoming beam with respect to the expected energy in the receiving synchrotron can be verified by observing its revolution frequency during a de-bunching in combination with a measurement of radial position offset [36, 37]. Fig. 28 shows mountain range plots during the de-bunching during the first 1.8 ms after injection of a proton bunch into the PS at CERN. For this

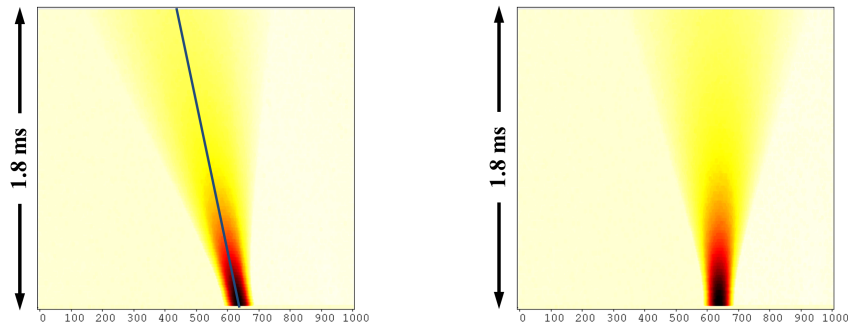


Fig. 28: De-bunching with an offset in bending field (left) and well matched (right).

transfer the revolution frequency in the PS and the radial position offset are kept constant. Hence the bending field and the momentum are defined according to Eqs. (23) to (26). The RF is switched off, but the triggers for the acquisitions of the beam signals are still derived from a synthesizer generating the reference revolution frequency for the transfer. In the left image the bunch centre advances with respect to the trigger corresponding to a too large revolution frequency. At the same time a radial position offset of the circulating beam after injection is measured. The energy of the incoming beam is therefore slightly too high. The beam circulates too far outside with a too high revolution frequency. After correction of

both, the bending field at injection and the momentum of the incoming bunch, the beam de-bunches exactly at the expected revolution frequency and at the central orbit, $\Delta R = 0$.

6.5 Longitudinal Matching

As shown above, the phase and the absolute energy of an incoming beam must match the position of the RF bucket, as well as the expected energy of the receiving accelerator. However, also the aspect ratio in terms of bunch length and energy spread of the injected bunch must be matched to the aspect ratio of the RF bucket to avoid longitudinal quadrupole oscillations and uncontrolled emittance growth.

The equations of motion (7) and (9) have the form of Hamilton equations for which the Hamiltonian can be written as [38]

$$H\left(\Delta\phi, \frac{\Delta E}{\omega_{\mathrm{rev}}}\right) = -\frac{1}{2}\frac{h\eta\omega_{\mathrm{rev}}}{pR}\left(\frac{\Delta E}{\omega_{\mathrm{rev}}}\right)^2 + \frac{qV}{2\pi}\left[\cos(\phi_0 + \Delta\phi) - \cos\phi_0 + \Delta\phi\sin\phi_0\right] . \tag{27}$$

In the special case of small phase deviations $\Delta\phi \ll 1$, the potential term can again be approximated

$$\begin{aligned}\cos(\phi_0 + \Delta\phi) &= \cos\phi_0\cos\Delta\phi - \sin\phi_0\sin\Delta\phi \\ &\simeq \cos\phi_0\left(1 - \frac{1}{2}\Delta\phi^2\right) - \sin\phi_0\Delta\phi\end{aligned}$$

such that the Hamiltonian simplifies to

$$H\left(\Delta\phi, \frac{\Delta E}{\omega_{\mathrm{rev}}}\right) \simeq -\frac{1}{2}\frac{h\eta\omega_{\mathrm{rev}}}{pR}\left(\frac{\Delta E}{\omega_{\mathrm{rev}}}\right)^2 - \frac{1}{2}\frac{qV}{2\pi}\Delta\phi^2\cos\phi_0 . \tag{28}$$

This approximated form of the Hamiltonian contains only quadratic terms of $\Delta\phi$ and ΔE. Consequently, it describes elliptical trajectories in the longitudinal E-ϕ-phase space.

The aspect ratio of the elliptical trajectories is evaluated by considering that, the Hamiltonian at the points of maximum phase or maximum energy deviation

$$\begin{aligned}H\left(\Delta\phi = 0, \frac{\Delta E}{\omega_{\mathrm{rev}}}\right) &= -\frac{1}{2}\frac{h\eta\omega_{\mathrm{rev}}}{pR}\left(\frac{\Delta E}{\omega_{\mathrm{rev}}}\right)^2 \\ H\left(\Delta\phi, \frac{\Delta E}{\omega_{\mathrm{rev}}} = 0\right) &= -\frac{1}{2}\frac{qV}{2\pi}\Delta\phi^2\cos\phi_0\end{aligned}$$

must be identical for a given trajectory such that

$$H\left(\Delta\phi = 0, \frac{\Delta E}{\omega_{\mathrm{rev}}}\right) = H\left(\Delta\phi, \frac{\Delta E}{\omega_{\mathrm{rev}}} = 0\right) .$$

Replacing the phase $\Delta\phi$ by $\Delta\tau = 2\pi f_{\mathrm{RF}}\Delta\tau = h\omega_{\mathrm{rev}}\Delta\tau$ as a frequency independent variable and using

$$pR = \frac{E\beta^2}{\omega_{\mathrm{rev}}} ,$$

Eq. (6.5) can be written as

$$-\frac{1}{2}\frac{h\eta\omega_{\mathrm{rev}}^2}{E\beta^2}\left(\frac{\Delta E}{\omega_{\mathrm{rev}}}\right)^2 = -\frac{1}{2}\frac{qV}{2\pi}h^2\omega_{\mathrm{rev}}^2\Delta\tau^2\cos\phi_0 .$$

The aspect ratio $\Delta E/\Delta\tau$ of the elliptical trajectory finally becomes

$$\left(\frac{\Delta E}{\Delta\tau}\right)^2 = \frac{qV}{2\pi}E\beta^2 h\omega_{\mathrm{rev}}^2\frac{\cos\phi_0}{\eta} . \tag{29}$$

For a longitudinally matched transfer between to synchrotrons the parameters must be chosen such that the aspect ratios Eq. (29) are the same in both, sending and receiving, accelerators. This assures that the particle distribution in the longitudinal phase space can remain unchanged before and after the beam transfer. It results in the general longitudinal matching condition

$$q_1 V_1 E_1 \beta_1^2 h_1 \omega_{\mathrm{rev},1}^2 \frac{\cos\phi_{0,1}}{\eta_1} = q_2 V_2 E_2 \beta_2^2 h_2 \omega_{\mathrm{rev},2}^2 \frac{\cos\phi_{0,2}}{\eta_2} \,.$$

It it worth noting that matching is only perfect for small bunches in large buckets due to the approximation of the Hamiltonian.

In most cases the charge, velocity and energy of the particles do not change during the transfer. Most transfers take moreover place constant energy, with a synchronous phase, $\phi_{0,1}$ and $\phi_{0,2}$ of 0 or π, which simplifies the voltage ratio for a matched transfer to

$$\frac{V_1}{V_2} = \left(\frac{R_1}{R_2}\right)^2 \left|\frac{\eta_1}{\eta_2}\right| \frac{h_2}{h_1} \,. \tag{30}$$

The transfer between two synchrotrons where the receiving one operates at twice the RF frequency as the sending accelerator, $f_{\mathrm{RF},2} = 2 f_{\mathrm{RF},1}$, is shown as an example for longitudinal matching. Figure 29 illustrates the combination of the longitudinal phase space in sending and receiving synchrotron. Even

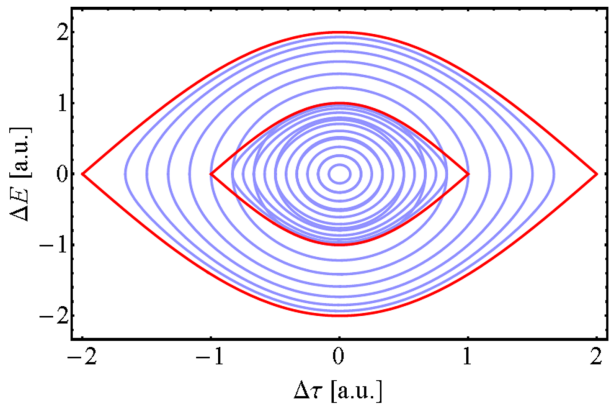


Fig. 29: Overlay of the longitudinal phase space in sending and receiving synchrotron.

though the bucket length and heights are very different, the inner trajectories of both buckets match very well. As long as the longitudinal emittance stays within that area, the transfer is longitudinally well matched.

6.5.1 Effect of voltage mismatch

Longitudinal mismatch at transfer leads to quadrupolar bunch oscillations, which is observed as a breathing of the bunch length, and to longitudinal emittance growth. The particle distributions obtained with single-particle tracking shown in Fig. 30 demonstrate this effect. While the matched distribution stays macroscopically unchanged (left), it rotates in the longitudinal phase space in the unmatched case (middle and right). Due to the non-linearity of the synchrontron frequency the bunch profile is massively perturbed. The filamentation finally leads to uncontrolled growth of the longitudinal emittance. While this emittance growth at transfer should be avoided for high-brightness beams, it may become a tool to deliberately reduce the longitudinal density for high-intensity beams. This technique is, e.g., used at injection of the anti-proton production beam into the PS at CERN [39].

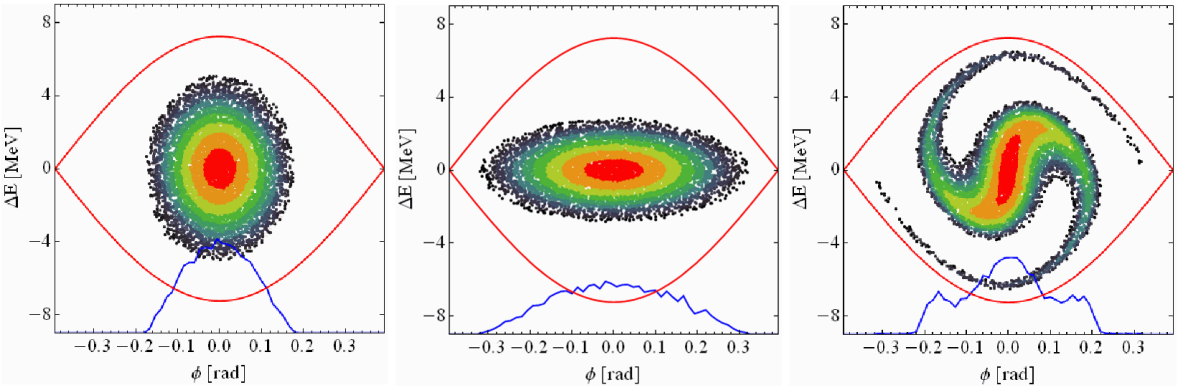


Fig. 30: Bunch distribution of a matched bunch after about 1.75 periods of the synchrotron frequency (left). Initial distribution of a longitudinally mismatched bunch (middle), and again after 1.75 periods of the synchrotron period.

6.5.2 Effect of bucket phase and energy mismatch

As discussed above also the bucket phase and energy of the RF bucket can be mismatched with respect to an incoming bunch. This can be seen in the simulated mountain range plots Fig. 31, where the RF bucket is displaced in phase (left) and energy (right). While the longitudinal phase space distribution

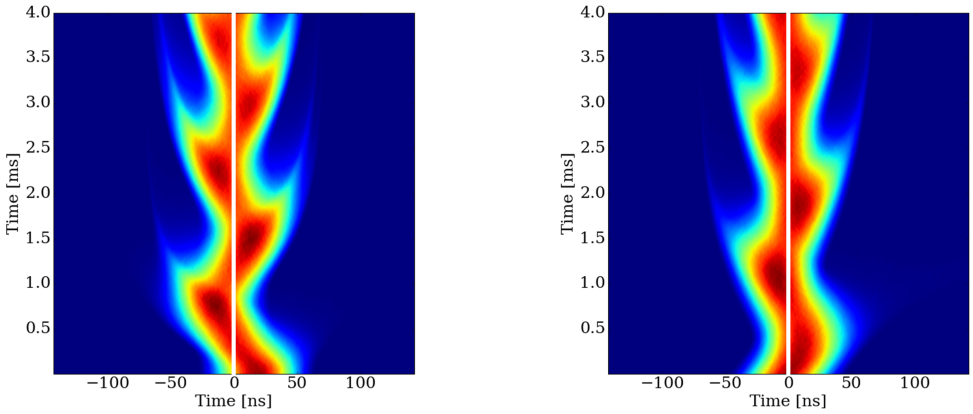


Fig. 31: Mountain range plot of bunch captured in a displaced RF bucket at injection. The initial displacement can be either in phase (left) or energy (right).

can only be accessed by tomographic techniques, the mountain range plots can be easily measured in a real accelerator. The phase error established at injection in the left plot is set to $45°$ and the energy error set for the simulation shown on the right plot also corresponds to the same oscillation amplitude. Qualitatively both images look very similar. However, the initial phase of the oscillation is different. With a pure phase error, the phase displacement with respect to the bucket center can be observed, which is not the case for an energy error. The mismatch (Fig. 31, left) can be corrected by changing the phase of the incoming bunch with respect to the RF bucket, while the energy error (Fig. 31, right) can only be removed by an energy matching (Sec. 6.4.2).

In electron accelerators, phase and energy mismatch can actually be used as a feature for special longitudinal injection schemes. For the electron option of the Future Circular Collider (FCC-ee), an off-momentum and off-phase injection is proposed [40]. Due to the strong synchrotron radiation the injected beam jumps the septum blade after one turn (Fig. 32). Thereafter is spirals toward the bucket centre at the reference energy. Similar schemes have been used for off-momentum injection and accumulation of beams into the Large Electron Positron collider (LEP) at CERN [41]. A first bunch was injected at an

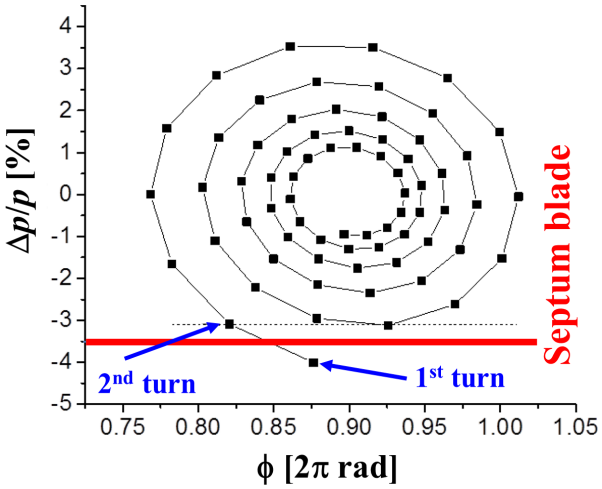


Fig. 32: Proposed synchrotron injection into FCC-ee. After the first turn, the injection bunch jumps the injection septum blade and then spirals towards the bucket centre.

energy slightly below the energy of the accumulated beam. After half a period of the synchrotron oscillation, the bunch moved to an energy above the reference energy, freeing the longitudinal phase space below for a second injection. This so-called double batch injection scheme again obviously only works for lepton bunches which match themselves to the buckets thanks to synchrotron radiation damping.

7 Acknowledgement

The author would like to thank Masamitsu Aiba, Maria-Elena Angoletta, Kalantari Babak, Diego Barrientos, Philippe Baudrenghien, Chandra Bhat, Thomas Bohl, Craig Drennan, Roland Garoby, Christopher Gough, Steven Hancock, Stephan Hunziker, Andreas Jankowiak, Boris Keil, Karim Lihem, John Molendijk, Holger Schlarb, Fumihiko Tamura, Frank Tecker, Daniel Valuch for discussions and material provided in preparation of this course.

References

[1] H. Schlarb, *Timing and Synchronization*. In CERN Accelerator School: Advanced Accelerator Physics, Trondheim, Norway, 2015, available at the URL `cas.web.cern.ch/sites/cas.web.cern.ch/files/lectures/trondheim-2013/schlarb.pptx`.

[2] Universität Bonn, *The Bonn University Electron Synchrotron*. In: CERN Symposium on High Energy Accelerators and Pion Physics, Geneva, Switzerland, 1956, pp. 482–483, `https://cds.cern.ch/record/1241624`.

[3] M. Sands, *The Physics of Electron Storage Rings: An Introduction*. SLAC-R-121, University of California, Santa Cruz, California, USA, 1970 and revised, 1979, available at the URL `http://www-public.slac.stanford.edu/scidoc/docMeta.aspx?slacPubNumber=slac-R-121`.

[4] H. Nyquist, *Thermal Agitation of Electric Charge in Conductors*. Physical Review, Vol. 32, 1928, pp. 110–113.

[5] A. Van Der Ziel, *Thermal Noise at High Frequencies*. Journal of Applied Physics, Vol. 21, 1950, pp. 399–401.

[6] IEEE, *IEEE Standard Definitions of Physical Quantities for Fundamental Frequency and Time Metrology-Random Instabilities*. IEEE Std 1139-1999, The Institute of Electrical and Electronics Engineers, Inc., New York, New York, USA, 1999.

[7] G. Lutes et al., *Thermal Coefficient of Delay for Various Coaxial and Fiber-Optic Cables*. The Telecommunications and Data Acquisition Progress Report, TDA PR 42-99, Jet Propulsion Lab., California Inst. of Tech., Pasadena, California, USA, 1989, pp. 43-59, available at the URL https://ntrs.nasa.gov/search.jsp?R=19900010123.

[8] J. Molendijk at al., *AWAKE - Status of the RF System*. AWAKE collaboration meeting, Lisbon, Portugal, 2016, available at the URL https://indico.cern.ch/event/484032/contributions/2002124/attachments/1240311/1825548/AwakeRF_Status_Lisbon.pdf.

[9] D. Barrientos et al., *RF Synchronization and Distribution for AWAKE at CERN*. IPAC2016, Busan, Korea, 2016, pp. 3743–3746, available at the URL http://accelconf.web.cern.ch/AccelConf/ipac2016/papers/thpmy039.pdf.

[10] A. Gallo, *Timing and Synchonization*. In: CERN Accelerator School: Advanced Accelerator Physics, Warsaw, Poland, 2015, available at the URL cas.web.cern.ch/sites/cas.web.cern.ch/files/lectures/warsaw-2015/gallo.pdf.

[11] F. Loehl, *Timing and Synchronization*. In CERN Accelerator School: Advanced Accelerator Physics, Chios, Greece, 2011, available at the URL cas.web.cern.ch/sites/cas.web.cern.ch/files/lectures/chios-2011/loehl.pdf.

[12] U. L. Rohde, *Microwave and Wireless Synthesizers: Theory and Design*. Wiley, New York, New York, USA, 1997.

[13] R. Garoby, *Low level RF building blocks*. In: CERN Accelerator School: RF Engineering for Particle Accelerators, Oxford, United Kingdom, 1991, available at the URL http://cds.cern.ch/record/225609.

[14] Hewlett Packard Company, *Model 10514A, B Mixers*. Operating and Service Manual, 1967.

[15] Analog Devices, *AD834, 500 MHz Four-Quadrant Multiplier*. Data sheet, 2012.

[16] W. T. Greer, *Digital phase-locked loops move into analog territory*. Electronic Design, 1982, pp. 95–100.

[17] D. Abramovitch, *Phase-Locked Loops: A Control Centric Tutorial*. American Control Conference 2002, Anchorage, Alaska, USA, 2002.

[18] Motorola, Inc. *MC12040, Phase-Frequency Detector*. Datasheet, 1997.

[19] J. Volder, *The CORDIC Computing Technique*. Western Joint Computer Conference, Boston, Massachusetts, USA, 1959, pp. 257-261.

[20] J. Kim et al., *Balanced optical–microwave phase detectors for optoelectronic phase-locked loops*. Optics Letters, Opt. Lett. 31, 3659-3661 (2006).

[21] U. L. Rohde, J. C. Whitaker, H. Zahnd, *Communications receivers: Principles and Design*. McGraw-Hill Education, New York, New York, USA, 2017.

[22] W. Pirkl, *Longitudinal Beam Dynamics*. In: CERN Accelerator School : 5th Advanced Accelerator Physics Course, Rhodes, Greece, 1995, available at the URL https://cds.cern.ch/record/302486.

[23] G. Dôme, *Theory of RF Acceleration and RF Noise*. In: CERN Accelerator School: Antiprotons for Colliding-beam Facilities, CERN, Geneva, Switzerland, 1984, available at the URL http://cds.cern.ch/record/863008.

[24] B. W. Montague, *Single-Particle Dynamics - RF Acceleration*. In: 1st International School of Particle Accelerators "Ettore Majorana" - Theoretical aspects of the behaviour of beams in accelerators and storage rings, CERN 77-13, Erice, Italy, 1977, available at the URL https://cds.cern.ch/record/118362.

[25] D. Boussard, *Une présentation élémentaire du système Beam Control du PS*. MPS/SR/Note/73-10, CERN, Geneva, Switzerland, 1973.

[26] D. A. G. Neet, *Timing for the ISR, Matching and Synchronization with the PS*. ISR-CO/67-28,

CERN, Geneva, Switzerland, 1967, available at the URL http://cds.cern.ch/record/296980.

[27] H. Damerau et al., *RF Manipulations for Higher Brightness LHC-Type Beams*. IPAC2013, Shanghai, China, 2013, pp. 2600-2602, available at the URL http://accelconf.web.cern.ch/AccelConf/IPAC2013/papers/wepea044.pdf.

[28] R. Garoby, *Timing Aspects of Bunch Transfer Between Circular Machines, State of the Art in the PS Complex*. CERN/PS/RF/Note 84-6, CERN, Geneva, Switzerland, 1984, available at the URL https://cds.cern.ch/record/2255149.

[29] R. Lauckner, *Matching the Energies of the CPS and SPS Machines*. SPS/AMS/Note/88-1, CERN, Geneva, Switzerland, 1988, available at the URL http://cds.cern.ch/record/1065731.

[30] S. Hancock et al., *Single-batch Filling of the CERN PS for the LHC-type Beams*. IPAC2010, Kyoto, Japan, 2010, pp. 699-701, available at the URL http://accelconf.web.cern.ch/AccelConf/IPAC10/papers/mopd014.pdf.

[31] A. V. Tollestrup, *R.F. Synchronization During Transfer in the Cascade Synchrotron*. Report CTSL-21, Synchrotron Laboratory, California Institute of Technology, Pasadena, California, USA, 1961, available at the URL https://authors.library.caltech.edu/73330/.

[32] , W. Schnell, *Synchronization of ISR and CPS at Slightly Unequal Frequencies*. ISR-RF/66-18, CERN, Geneva, Switzerland, 1966.

[33] T. Linnecar et al., *Hardware and Initial Beam Commissioning of the LHC RF Systems*. LHC Project Report 1172, CERN, Geneva, Switzerland, 2008, available at the URL http://cds.cern.ch/record/1176380.

[34] C. Bovet et al., *A selection of formulae and data useful for the design of A.G. synchrontrons*. CERN-MPS-SI-INT-DL-70-4, CERN, Geneva, Switzerland, 1970, available at the URL http://cds.cern.ch/record/104153.

[35] S. Hancock, *Transition revisited*. CERN/PS/90-53 (OP), CERN, Geneva, Switzerland, 1990.

[36] S. Hancock et al., *A Straightforward Procedure to achieve Energy Matching Between PSB and PS*. CERN-AB-Note-2008-042, CERN, Geneva, Switzerland, 2008, available at the URL http://cds.cern.ch/record/1125475.

[37] S. Hancock, *Energy Matching Between LEIR and PS*. CERN-ACC-NOTE-2015-0019, CERN, Geneva, Switzerland, 2015, available at the URL http://cds.cern.ch/record/2038693.

[38] F. Tecker, *Longitudinal Beam Dynamics - Recap*. These proceedings.

[39] H. Damerau et al., *Controlled Longitudinal Emittance Blow-up in the CERN PS*. PAC07, Albuquerque, New Mexico, USA, 2007, pp. 4186–4188, available at the URL http://accelconf.web.cern.ch/AccelConf/p07/PAPERS/FRPMN070.PDF.

[40] M. Aiba, *Top-up injection for FCC-ee*. CERN-ACC-2015-065, CERN, Geneva, Switzerland, 2015, available at the URL https://cds.cern.ch/record/2031423.

[41] R. Bailey, *Is Injection Up To Scratch for All Filling Scenarios?*. 6th LEP Performance Workshop, Chamonix, France, 1996, pp. 41–44, available at the URL http://sl-div.web.cern.ch/sl-div/publications/chamx96/papers/rb2_01.pdf.

Proceedings of the CAS–CERN Accelerator School: Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018)

Magnets and Special Magnets

C. Muehle
GSI Helmholtz Centre for Heavy Ion Research, Darmstadt, Germany

Abstract
Magnets are key components of all charged particle accelerators, their beam lines, and their experiments. They are used to guide, focus, and correct charged particle beams. Depending on the varying requirements of different accelerators, the simple underlying physics leads to a wide variety of types and designs.

Keywords
Accelerator magnets; dipoles; quadrupoles; sextupoles; steering magnets.

1 Introduction

This paper is intended as a brief introduction to the normal conducting accelerator magnets. Their basic physics along with the underlying principles of magnet design and technical realization are discussed and presented here. The aim is to provide a general understanding on the basis of first principles and very simple calculus.

The second chapter gives a brief introduction into the basic physics. The third chapter describes magnet types, their influence on the charged particles, as well as technical issues such as polarity and design types. The fourth chapter concentrates on coils. Methods for the estimation of excitation currents are derived, conductor materials are compared, and cooling as well as coil shapes are explained. Chapter five is dedicated to yokes. Yoke materials are compared, and saturation effects and their corresponding countermeasures are explained. Special magnets are the topic of chapter 6. Magnets that are special with respect to design features or materials, as well as magnets which cannot be described by multipoles, are illustrated here. Examples are directly integrated. Chapter 7 describes examples of standard accelerator magnets. This is elucidated with the magnets used for the cancer therapy accelerator Heidelberg Ion-beam Therapy Center (HIT), in Heidelberg.

2 Basic physics

2.1 The Lorentz force

Charged particles experience forces when subjected to electric or magnetic fields. The force exerted by an electromagnetic field on them is given by the Lorentz force[1] as in Eq. (1)

$$\vec{F} = q\left[\vec{E} + \left(\vec{v} \times \vec{B}\right)\right], \tag{1}$$

where, q is the charge of the particle, $\vec{E}$ is the electric field, $\vec{v}$ is the velocity of the charged particle, and $\vec{B}$ is the magnetic flux density.

[1] The first derivation of the Lorentz force is commonly attributed to Oliver Heaviside in 1889, although other historians suggest an earlier origin in an 1865 paper by James Clerk Maxwell. Hendrik Antoon Lorentz derived it in 1895, a few years after Heaviside. (Wikipedia)

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-005.193

2.2 Maxwell's equations

Maxwell's equations[2] are a collection of laws governing electromagnetism. Writing them in the differential and integral forms using SI convention yields the equations in Table 1.

Table1: Maxwell equations

Law	Differential form	Integral form
Gauss's law for electrostatics	$\nabla \cdot D = \rho$	$\oint_{\partial v} D \cdot \mathrm{d}A = Q$
Gauss's law for magnetism	$\nabla \cdot \mathrm{B} = 0$	$\oint_{\partial \mathrm{v}} \mathrm{B} \cdot \mathrm{dA} = 0$
Faraday's law of induction	$\nabla \times \mathrm{E} = -\frac{\partial \mathrm{B}}{\partial \mathrm{t}}$	$\oint_{\partial \mathrm{A}} \mathrm{E} \cdot \mathrm{ds} = -\frac{\mathrm{d}}{\mathrm{dt}} \int_{\mathrm{A}} \mathrm{B} \cdot \mathrm{dA}$
Ampere's circuital law	$\nabla \cdot H = j + \frac{\partial D}{\partial t}$	$\oint_{\partial A} H \cdot \mathrm{d}s = \int_{A} j \cdot \mathrm{d}A + \frac{\mathrm{d}}{\mathrm{d}t} \int_{A} D \cdot \mathrm{d}A$

The differential formulation serves to provide a basic understanding between the physical quantities and their nature while the integral form comes in handy when actually performing calculations to quantify the terms. In the context of this report, only Gauss's law for magnetism and Ampere's circuital law are of interest. The first forbids the existence of magnetic monopoles. The latter will be used for the calculation of excitation currents.

2.3 Magnetic material properties

The defining equation of a magnetic material is given by the relation,

$$B = \mu H \,, \tag{2}$$

with $\mu = \mu_0 \mu_r$.

The structure and physical properties of magnetic materials are defined by the concepts of isotropy and homogeneity [1].

(i) When μ is a constant number, the material is a *simple magnetic material.*

(ii) When μ varies according to the position in a material as $\mu(r)$, it is an *inhomogeneous magnetic material.*

(iii) When μ becomes a function of H, as in $\mu(H)$, B and H are no longer proportional, thus defining a *non-linear magnetic material.*

(iv) When the relation between B and H becomes highly non-linear such that μ cannot be clearly defined, it is a *ferromagnetic material.*

(v) When μ is a tensor $\lfloor \mu \rfloor$, it yields $B = \lfloor \mu \rfloor . H$. This means that B and H have different directions making it an *anisotropic material*. Under specific rotations, $\lfloor \mu \rfloor$ transforms as a second rank tensor usually called the permeability tensor given by,

[2] The four modern Maxwell's equations can be found individually throughout his 1861 paper, derived theoretically using a molecular vortex model of Michel Faraday's "lines of force" and in conjunction with the experimental result of Weber and Kohlrausch. But it wasn't until 1884 that Oliver Heaviside, concurrently with similar work by Josiah Willard Gibbs and Heinrich Hertz, grouped the twenty equations together into a set of only four, via vector notation. This group of four equations was known variously as the Hertz–Heaviside equations and the Maxwell–Hertz equations, but are now universally known as Maxwell's equations. Heaviside's equations, which are taught in textbooks and universities as Maxwell's equations are not exactly the same as the ones due to Maxwell, and, in fact, the latter are more easily forced into the mould of quantum physics. (Wikipedia)

$$\mu = \begin{bmatrix} \mu_{xx} & \mu_{xy} & \mu_{xz} \\ \mu_{yx} & \mu_{yy} & \mu_{yz} \\ \mu_{zx} & \mu_{zy} & \mu_{zz} \end{bmatrix}. \quad (3)$$

Different combinations of properties and their influence on the permeability tensor are compiled in Table 2.

Table 2: Overview of material types

Material	Tensor terms
Isotropic and homogeneous	$\mu_{ij} = 0, \mu_{ii} = \mu_r \neq 0$
Isotropic and inhomogeneous	$\mu_{ij} = 0, \mu_{ii} = \mu_r(x, y, z)$
Anisotropic and homogeneous	$\mu_{ii} = constant$
Anisotropic and inhomogeneous	$\mu_{ij} = constant, \mu_{ii} = \mu_r \neq 0$
Isotropic, homogeneous, nonlinear (massive iron)	$\mu_{ij} = 0, \mu_{ii} = \mu_r(\vec{H}(x, y, z))$
Anisotropic, homogeneous, nonlinear (laminated iron)	$\mu_{ij} = \mu_{ij}(\vec{H}(x, y, z))$

3 Magnet types

3.1 Magnets in the accelerator

Magnets in accelerators can be basically subdivided into three groups – namely bending magnets, focussing magnets, and non-linear magnets.

Bending magnets are dipoles. With a constant magnetic field across the aperture, they deflect charged particle beams along a curved path.

Focussing magnets are quadrupoles. The magnetic field increases linearly starting from zero on the axis (from the centre of the magnet). It can therefore be expressed by a gradient.

Non-linear magnets exhibit a non-linear increase in the magnetic field starting from zero on the axis (from the centre of the magnet). The most common ones are sextupoles and octupoles. These magnets are used as corrector magnets.

These magnets are sketched in Fig. 1.

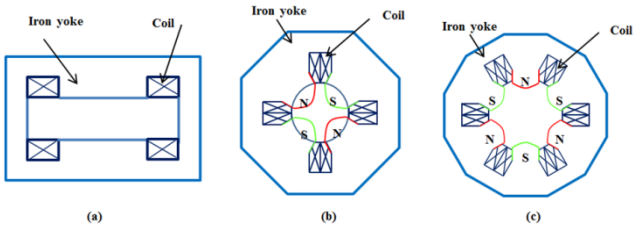


Fig. 1: (a) H-type dipole (b) quadrupole (c) sextupole

3.2 Optics analogies

The influence of the magnets on the beam can be easily shown by taking a look at their respective optical analogies, as seen in Fig. 2.

The optical analogy of a dipole is the prism. The incident beam is deflected by a certain angle.

Spherical lenses are analogous to quadrupoles. The focussing or defocussing is independent of the distance to the axis.

Aspherical lenses have the same behaviour as non-linear magnets. The strength of the focussing or defocussing is dependent on the distance to the axis.

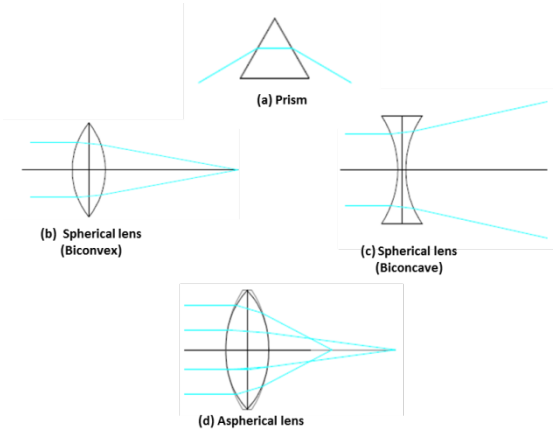


Fig. 2: (a) Illustration of a deflecting dipole (b) illustration of a focussing quadrupole (c) illustration of a defocussing quadrupole (d) illustration of non-linear magnets with dependence on the focal length.

3.3 Magnetic flux lines

The topology of the magnetic field can be visualized by plotting the equipotential flux lines, as seen in Fig. 3. Higher order multipoles follow the same pattern.

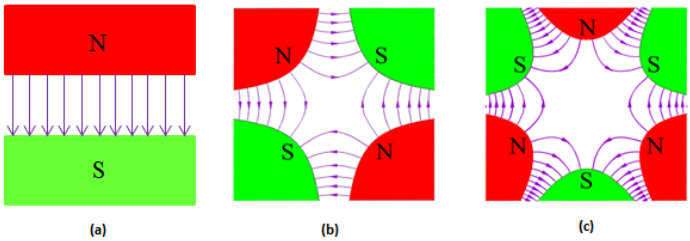


Fig. 3: (a) Dipole – parallel and equidistant lines, (b) quadrupole – radially symmetric lines with alternating directions (four poles), and (c) sextupole – radially symmetric lines with alternating directions (six poles).

3.4 Definition of conventions

The direction of the magnetic field ($\vec{B}$) in an electromagnet is determined by the direction of the current ($\vec{I}$) in its coils. To understand the orientations of $\vec{B}$ and the flux ($\vec{\Phi}$) due to $\vec{I}$, we use the right-hand rule as a tool. This rule provides the relation between these vectors in the positive direction (counter clockwise). By convention, it is taken that the positive current flows from the positive terminal of the power supply to the negative terminal. This is defined as the conventional current direction, which incidentally is the same for a positively charged ion. In the similar way, the polarity of the magnet is defined by the magnetic flux flowing from the north pole to the south pole of a magnet.

If the fingers envelope the coil with the finger tips pointing in the direction of the current, the thumb indicates the direction of the magnetic field as seen in Fig. 4.



Fig. 4: Right-hand rule to find the direction of the magnetic field in a coil

Applying the above discussions to our magnets, we can find the direction of *B* using the first part of the right-hand rule as follows:

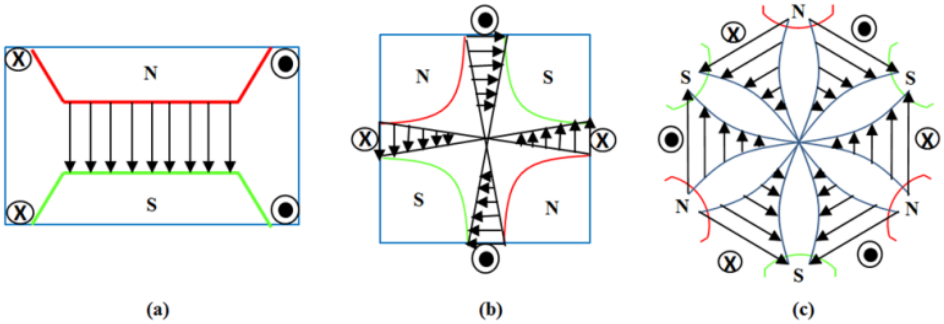


Fig. 5: Magnetic field direction in different magnets in dependence of the current direction in the coils: (a) dipole, (b) quadrupole, and (c) sextupole. '⊗' indicates positive current into the paper while '⊙' indicates positive current out of the paper.

3.5 Forces on the charged particle beam

As the force on a charged particle in a magnetic field is defined by the cross product, as in Eq. (1), the deflection can easily be found using the second part of the right-hand rule. To find the direction of the force exerted on the charged particle, point the thumb along the direction of movement of the charge. With the pointer finger along the direction of the magnetic field, the middle finger indicates the direction of force experienced by the particle.

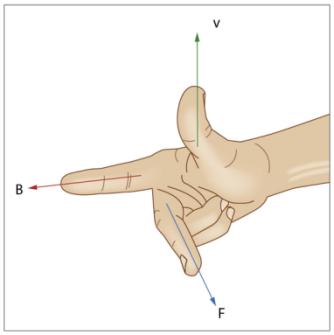


Fig. 6: Right-hand rule to find the force exterted on a charged particle

Due to the constant field in a dipole, a constant force can be found in the aperture. Figure 7, shows the force exerted by the field on a positive ion flying into the plane. It is bent to the left.

The force in a quadrupole follows the characteristic behaviour of its magnetic field (linear dependence). As the direction of the magnetic field alternates, the same is observed with the force too. In the example shown in Fig. 7, the positive ion beam is bent in the vertical plane towards the axis and away in the horizontal plane. This effect focusses the beam in the vertical plane and defocusses it in the horizontal plane. The naming convention is based on the effect in the horizontal plane. In this case

we have a defocussing quadrupole. A net focussing in both planes can only be achieved by using more than one quadrupole. A formation of two or three quadrupoles is called doublet or triplet respectively.

In a sextupole however, the magnetic field increases quadratically with the distance from the axis. The same is valid for the modulus of the force. But now we have a sixfold symmetry. Hence, we have three distinguished planes namely the median plane (along the horizontal axis) and two other planes at 120° azimuth and 240° azimuth respectively. The force now has the same direction on both sides of the median axis. Naming can be done now only with respect to the complete machine. If bending is towards the centre of a circular machine, the sextupole is called a focusing one.

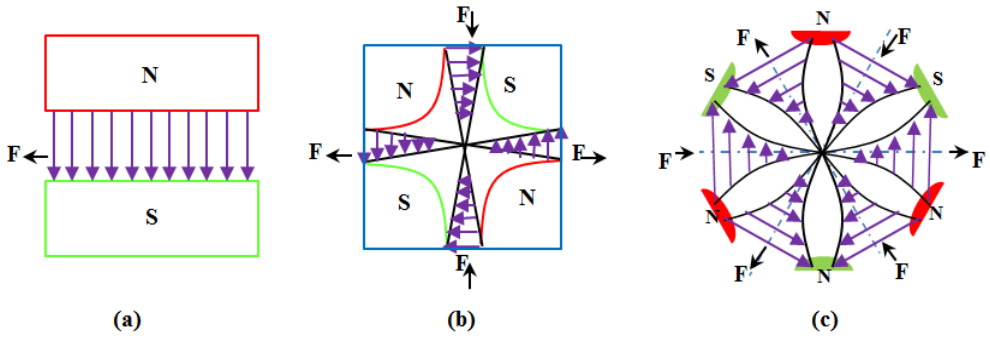


Fig. 7: Force direction in different magnets for a positive beam flying into the plane. (a) Dipole, (b) defocussing quadrupole, and (c) sextupole.

3.6 Design types

Magnets can be distinguished according to the effects dominating their field quality. The field distribution can either be current dominated or iron dominated.

Current dominated magnets are classified according to their coil shape. Typical representatives of current dominated dipoles are the $\cos n\theta$ magnets and intersecting ellipses (see Fig. 8). As our discussion is limited to the iron-dominated magnets alone, we do not discuss the current-dominated type magnets further.

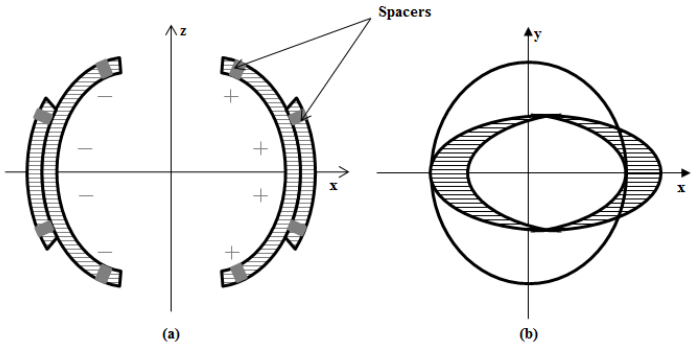


Fig. 8: Current dominated magnets: (a) $\cos n\theta$ dipole (b) dipole formed by crossed ellipses

Iron-dominated magnets are classified according to their yoke shapes. This is separately discussed further in sub sections 3.6.1 and 3.6.2 for dipoles and quadrupoles respectively.

3.6.1 Dipole types

Iron-dominated dipole magnets are classified according to their yoke shapes as the window-frame magnet and C-type magnet, or according to the shape of the aperture including the coil windows as H-type magnet (see Fig. 9). The coils are indicated in red. According to their yoke shape, the magnets exhibit different advantages and disadvantages.

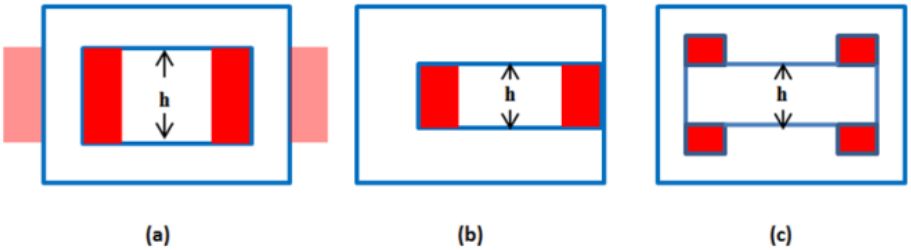


Fig. 9: Yoke shapes of iron dominated dipoles: (a) window-frame or O-type magnet, (b) C-type magnet, (c) H-type magnet.

The window-frame type has the most symmetric design. It can either be equipped with complex bedstead coils or cylindrical coils creating a higher leakage flux. Cylindrical coils are typically only used for short magnets, e.g., correction dipoles (so called steering magnets).

The C-type magnet saves space on one side of the yoke at the cost of an enlargement with an increased thickness above and below the poles in comparison to the window-frame magnet. As can be seen in Fig. 9, the same overall size gives a much smaller gap height for the C-type magnet in comparison to the window-frame magnet.

The H-type magnet has again the advantage of a symmetric design. The coils are simple to manufacture as they are flat. They have a so-called racetrack design. The comparison is compiled in Table 3.

Table 3: Comparison between iron-dominated dipole shapes

Magnet	Advantages	Disadvantages
Window frame type (O-type)	Symmetrical design	Bedstead coils or cylindrical coils with high flux leakage
C-type	Saves space on one side of the yoke, simple coil assembly	Asymmetrical, bedstead coil, heavy yoke
H-type	Symmetrical design, simple racetrack coils	Bad field quality (compared to window frame)

The field quality of the designs can be compared by using the orthogonal analogue model [2]. At first, 'flow' lines from plus to minus currents are drawn. The field lines are orthogonal to the flow lines and end perpendicularly at the iron boundaries (see Fig. 10).

The window-frame type shows a perfect field quality if the coil window is vertically completely filled with the coil. In real cases this is limited due to the insulation of the coil.

In an H-type magnet the field quality decreases when horizontally moving from the centre of the magnet. This can be mitigated either by adding adequate shims to the pole corners or by enlarging the width of the pole.

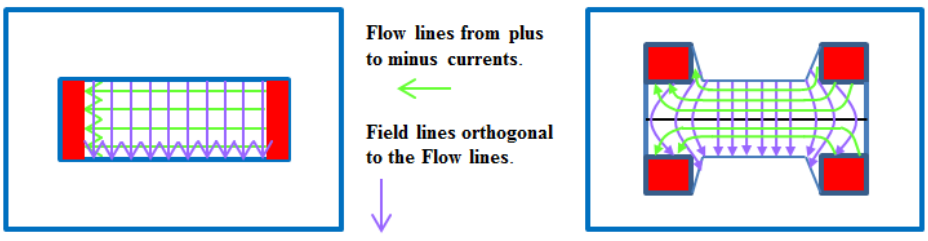


Fig. 10: Comparison of window-frame type and H-type dipoles with respect to their field quality

3.6.2 Quadrupole types

Iron-dominated quadrupoles exhibit a hyperbolic pole shape as seen in Fig. 11.

The Collins or figure of eight quadrupole is limited in width at the cost of additional height required for the yoke design. The upper and lower half of the yoke is magnetically decoupled.

The standard quadrupoles differ in their pole sides. The standard type I design has parallel sides. Therefore, the pole base has the same width as the pole itself. The standard type II design exhibits the largest possible increase in the pole base. The pole sides have an angle of 90° between each other. The type 1 has more space for simpler coils but tends to saturate inside the pole. Real quadrupoles are often hybrid designs of the above mentioned types.

The Panofsky quadrupole is a current dominated magnet shown here as an exception. The window-frame yoke collects the stray field and increases the field strength in the aperture. By using the orthogonal analogue method, one can convince oneself that this dipole-like looking magnet is a quadrupole.

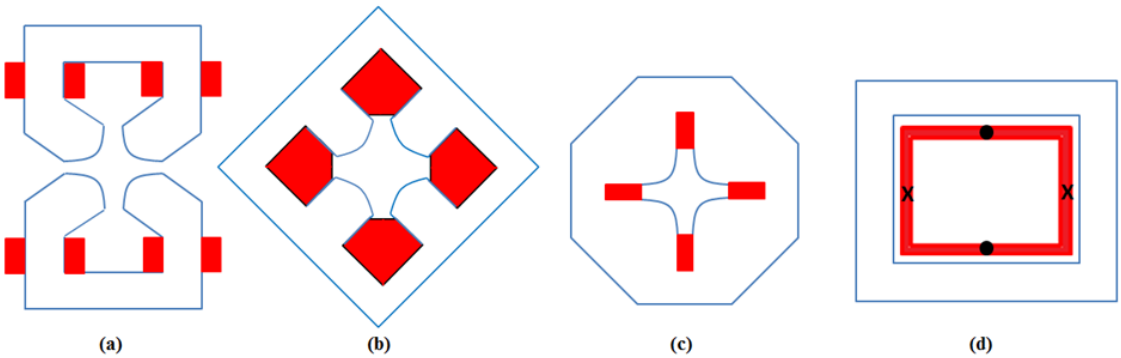


Fig. 11: Yoke shapes of quadrupole types: (a) Collins or figure of eight quadrupole, (b) standard quadrupole type I, (c) standard quadrupole type II, (d) Panofsky quadrupole.

4 Coils

4.1 Ampère's law

Ampère's law enables us to calculate the necessary excitation current in a magnet coil. The law states that (Eq. (4)),

$$\oint_{\partial A} H \cdot \mathrm{d}s = \int_A j \cdot \mathrm{d}A + \frac{\mathrm{d}}{\mathrm{d}t}\int_A D \cdot \mathrm{d}A\,.$$

Neglecting the time-dependent part, the closed loop integral over the magnetic field equals the enclosed current.

4.2 Dipole excitation

We are now going to execute the integration in Eq. (4) for a dipole. Figure 12 illustrates the following steps.

Starting at the centre of the magnet, we integrate at first the vertical air path in the gap. The path is parallel to the field direction and μ_r equals one.

$$\int \frac{\vec{B}}{\mu_0\mu_r}.\mathrm{d}\vec{l} = \frac{Bh}{\mu_0}.$$

Secondly, we do the integration through the iron yoke. The expression

$$\int \frac{\vec{B}}{\mu_0 \mu_r} . \overrightarrow{\mathrm{d}l} = \frac{Bh}{\mu_0 \mu_r} ,$$

gets small as μ_r is large.

Finally, we integrate the horizontal path. As $\vec{B}$ and $\vec{l}$ are perpendicular to each other we get

$$\int \frac{\vec{B}}{\mu_0 \mu_\mathrm{r}} . \overrightarrow{\mathrm{d}l} = 0 .$$

Summing up and taking into account that the enclosed current equals number of turns in one coil times the current in it, we get

$$NI_{\mathrm{dipole}} \approx \frac{Bh}{\mu_0} .$$

The expression NI is often called *ampere turns*.

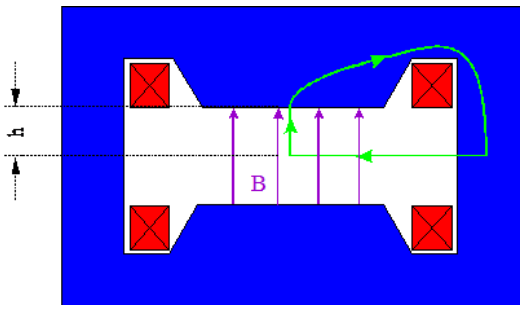


Fig. 12: Calculation of the excitation current in a dipole with the help of Ampere's law. The used integration path is shown in green.

4.3 Quadrupole excitation

The very same scheme can be used for quadrupoles, too. Again we start at the centre of the aperture. Exploiting the symmetry, we calculate for ¼th of the magnet as in Fig. 13. The path in the aperture is parallel to the field direction and μ_r equals one. We integrate a linearly increasing field represented by its gradient B' and get,

$$\int \frac{\vec{B}}{\mu_0 \mu_\mathrm{r}} . \overrightarrow{\mathrm{d}l} = \int_0^{r0} \frac{B'l}{\mu_0 \mu_\mathrm{r}} \mathrm{d}l = \frac{B' r_0^2}{2\mu_0} .$$

The integral through the iron yoke is again small as μ_r is large.

The integral along the horizontal path is again zero as $\vec{B}$ and $\vec{l}$ are perpendicular to each other.

Summing up we get,

$$NI_{\mathrm{quadrupole}} \approx \frac{B' 2 r_0^2}{2\mu_0}$$

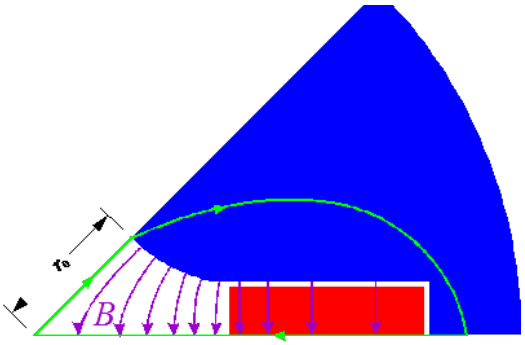


Fig. 13: Calculation of the excitation current in a quadrupole with the help of Amperes law. The used integration path is shown in green.

The very same procedure can be used for sextupoles too. This yields the ampere turns to be:

$$NI_{\text{sextupole}} \approx \frac{B'' r_0^3}{6\mu_0}.$$

4.4 Conductor materials

The power consumption of a magnet is given by the conductivity of its conductor material. Therefore, high conductivity copper and pure aluminium are used. Their different properties make them potential candidates for typical applications (see Table 4).

At a first glance, copper is the better choice due to its 1.6 times higher conductivity and smaller oxidation in comparison to aluminium. But, the weight price of copper is three to four times more than that of aluminium. Additionally, copper has three times the density of aluminium.

However, the co-efficient of thermal expansion of both materials lie in a comparable range among other physical properties between the two. The elasticity modulus of copper is about 1.7 times larger. Also, the deformation of the conductor cross-section while using small bending radii is larger for copper. This effect is called key stoning, as the square cross-section is deformed to a trapezoidal like shape (as seen in Fig. 14) similar to the keystone in the roman arch.

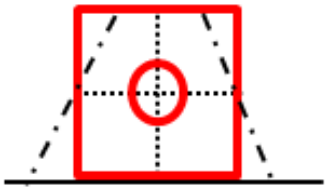


Fig. 14: Keystone effect in a copper conductor with cooling duct showing the deformation of a square conductor into a trapezoidal shape.

For the same amount of ampere turns we can conclude the following: Copper allows us to build smaller but heavier coils. Due to the higher density, the transparency for particles is smaller than for aluminium. The investment costs are higher. However, the operating costs are smaller as the conductivity is larger. Therefore, an application with a long annual operational period is favourable. This is typical for accelerators and their corresponding beam lines.

On the contrary, detector magnets have typically shorter annual operational periods. While using aluminium in this case, it is an advantage as this leads to lower investment costs. In addition, it can be used for detecting particles that have passed through the coil as the lower material density of aluminium enables this.

In any case, mixing of copper and aluminium coils inside one cooling circuit must be avoided. Otherwise, the aluminium coil would act as a sacrificial anode in perspective of a long-term operation.

Table 4: Comparison of the properties of aluminium and copper

Property	**Aluminium (pure, > 99.5%)**	**Copper (OFHC- oxygen free high conductivity)**
Conductivity	36 S m/mm^2	58 S m/mm^2
Oxidation	In air: AlO layer is formed. Dissolves in mixed copper/ aluminium cooling circuits.	Small
Price (large quantities)	2.7−3.4 EUR/kg	8−16 EUR/kg
Specific weight	2.70 g/cm^3	8.96 g/cm^3
Linear expansion coefficient	23 ×10^{-6} K^{-1}	17 × 10^{-6} K^{-1}
Elasticity modulus	72,000 N/mm^2	123,000 N/mm^2
Key stoning effect	Smaller	Higher
Conclusions	Larger coils are needed	Smaller coils are needed
(for same N × I)	Lighter	Heavier
	Higher transparency[3] for particles	Reduced transparency for particles
	Lower investment costs	Higher investment costs
	Higher operating costs	Lower operating costs
	⇨ Better suited for detector magnets	⇨ Better suited for accelerator magnets

4.5 Coil cooling

The power produced in the coils during operation has to be dissipated to avoid overheating. Typical media for cooling are air and water.

Air cooling is limited to low-power applications as the heat capacity is smaller than that for water. In the case of free flow, the cooling is based on convection. If this is not sufficient, forced flow air cooling can be provided by fans.

The most common method is direct water cooling with hollow conductors. The water is in direct contact with the conductor and provides efficient cooling. For low-current (high turn) applications, the conductor cross-section becomes too small to incorporate an internal cooling channel. In this case, separate water cooled radiators are used. An example with cooling discs [4] is shown in Fig. 15.

[3] The stopping of particles depends on the number of protons in the nucleus of the metal. For detector magnets, a low stopping power of materials (= high transparency) is of interest.

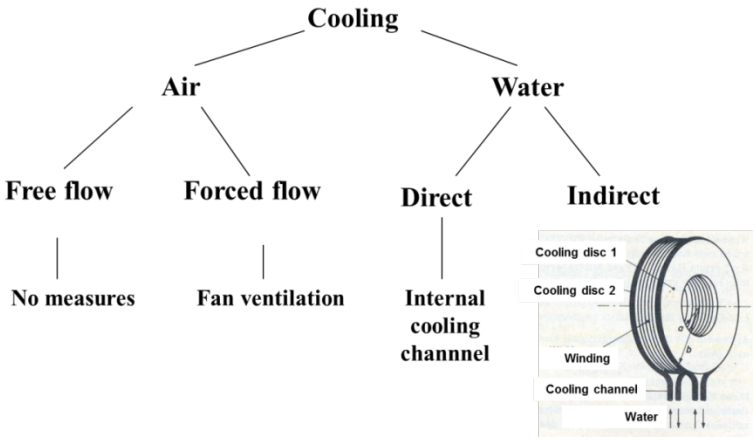


Fig .15: Cooling methods for normal conducting magnets

4.6 Conductor profiles

Wide variety of hollow copper conductors are commercially available (refer to Fig. 14). The most common shapes are the quadratic or rectangular conductors with a round cooling channel. These conductors allow good packing factors in the coil cross-section. Bending radii smaller than three to five times the cross-section should be avoided because of the keystone effect and rupture problems.

Round conductors are used in the case of small coils with tight bends in several directions. Examples are quadrupoles and steering magnets.



Fig. 16: Examples of hollow copper conductors [3]

4.7 Coil shapes

The coil shape is determined by the (yoke) design type. Besides supplying the needed ampere turns, it has to fit to the yoke and must not interfere with the beam aperture.

When recalled from section 3.6.1, H-type dipole magnets have simple flat coils. They are called cylindrical or racetrack coils as in Fig. 17 (c).

For window-frame magnets, either cylindrical or bedstead coils can be used. In case of cylindrical coil, it is wound around the yoke. This increases the stray fields around the magnet. Additionally, the coil circumference is increased to more than double the magnet length per coil. Therefore, cylindrical coils are only used in short magnets, like steering magnets. While using bedstead coils however, additional care must be taken such that the coil ends for long window-frame

magnets be cranked, to keep the beam aperture free. In this case they resemble Fig. 17 (a) or (b) and are named saddle or bedstead coils.

To sum up, the most commonly used coil for window-frame dipoles is the bedstead type. Racetrack coil is used only in cases where the overall magnet length has to be limited.

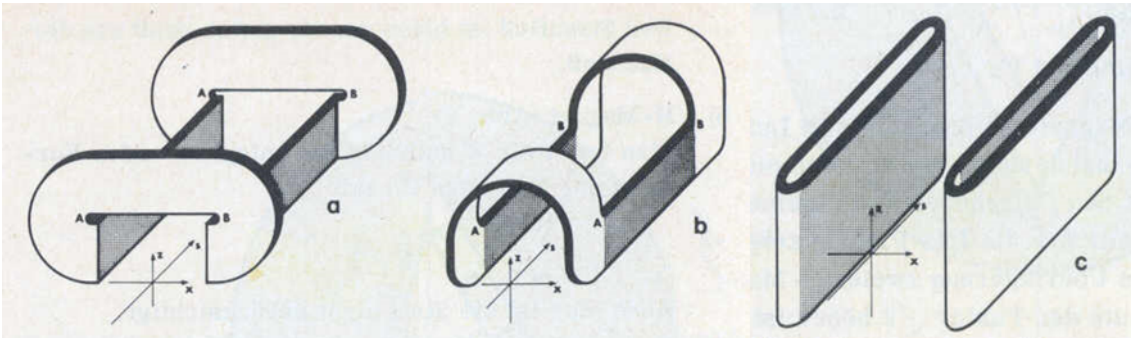


Fig. 17: Coil types for dipole magnets [4]: (a) saddle or bedstead coil with horizontal bend, (b) saddle or bedstead coil with a vertical bend. For (a) and (b) only the upper part is shown. (c) Cylindrical or racetrack coil. A pair is used in (short) window-frame magnets. One cylindrical coil per pole is used in H-type magnets.

A similar classification can be done for quadrupole magnets too. The coils for figure of eight quadrupoles and standard quadrupoles of type I have cylindrical coils. The coils for standard quadrupoles of type II need to have cranked ends. They are similar to the bedstead coils in Fig. 17 (a).

4.8 Coil potting

To form a rigid coil body and to maintain the insulation properties over long periods, a process known as coil potting is to be performed. Initially, the conductors of standard coils are wrapped with glass fibre tapes with half overlap during winding. The final coil body is then wrapped again with the same for ground insulation. Next, it is vacuum impregnated with an epoxy resin. Depending on the requirements, dedicated precision moulds for single coils or large common moulds for several coils are used. Finally, the coil body is often wrapped with a mould release film before potting as this eases the removal of the excess epoxy resin after the potting process.

5 Yoke

5.1 Yoke materials

All yoke materials (ferromagnetic in nature) exhibit a nonlinear relationship between the magnetic flux density and the magnetic field strength (H) given by the equation,

$$B = \mu(H)H,$$

with $\mu = \mu_0\mu_r$.

This behaviour is depicted in the B-H curves (see Fig. 18). A high relative permeability μ_r is always desirable as this guarantees the flux guiding function of the yoke.

Most commonly used is the so called soft iron which has high purity in iron content (low carbon). Saturation starts at about 1.5 T. Higher saturation flux densities can be reached by adding up to 50% of cobalt.

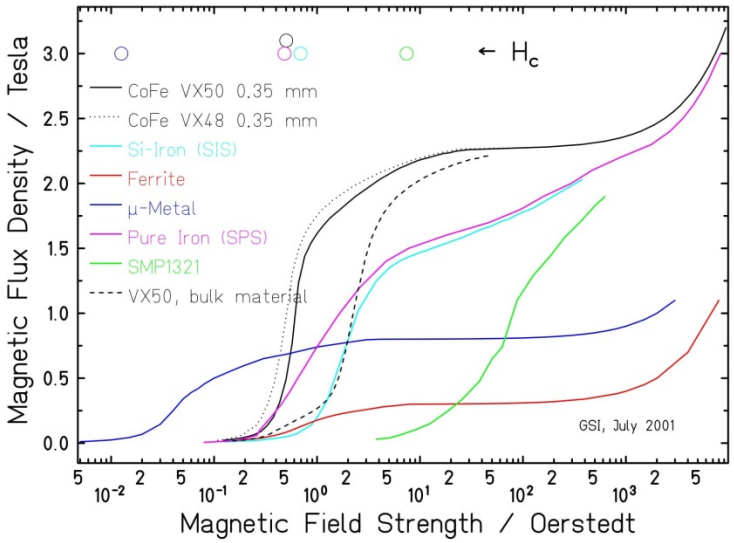


Fig. 18: Compilation of *B-H* curves for typical yoke materials. For comparison, ***μ***-metal is shown too. It has very high permeability at low fields and is therefore used for screening at low fields.

In non-DC applications, the conductivity of the yoke material must be taken into account. The time varying fields generate eddy currents in the yoke. These currents can be limited by increasing the electrical resistance of the yoke. The basic measure is to use a laminated yoke instead of a massive one. In this case, the yokes are stacked from thin plates as in a transformer. Also, the resistance of iron can be increased by adding up to ~4% silicon to it. If this is not sufficient, sintered iron or ferrites can even be used. Unfortunately, with all these measures the permeability is reduced.

The final choice of the yoke material is therefore determined by the required flux density and the ramp rate, as well as the space requirements and material costs (as seen in Fig. 19).

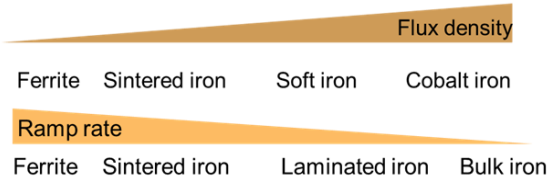


Fig. 19: Suitability of different yoke materials with respect to flux density and ramp rate

5.2 Saturation effects

As described above, the saturation for soft iron starts at about 1.5 T. This means, that μ gets smaller for flux density levels above 1.5 T. The flux density level as well as the field quality is influenced by this effect (summarized in Fig. 20).

In comparison to the equations derived in section 4.2 and 4.3, the field level gets smaller as the assumption of a large μ is not valid anymore. This effect is named *ampere turn loss*. This can be compensated either by increasing the ampere turns or by adapting the yoke design accordingly. The yoke design can be adapted by either increasing the width of the yoke or using a material with higher saturation flux density.

On the other hand, reduced field quality is caused by non-homogenous spatial distribution of saturation in the yoke. One countermeasure is to adapt the pole profile by mainly increasing the pole

width, thus moving the regions with reduced field quality further away from the region of interest. The second method is an active correction of the field quality by using the so called *pole face winding*. They act as a superimposed current dominated correction magnet. The third method is to introduce air slits in the poles (shown in Fig. 21). In H-type dipoles, they aim on equalizing the spatial distribution of saturation. The slits are located at the pole centre where the saturation is small. Due to the reduced iron, the areas around the holes tend to saturate at lower field levels and the effect of saturation is homogenized across the yoke. However, the consequence of this method is a further loss of ampere turns. In window-frame dipoles, the holes are positioned on the sides of poles (indicated in Fig. 22). For higher fields, these parts of the poles are switched off. The flux lines have to then go around the holes and the flux close to the pole is equalized, thus leading to an improved field quality.

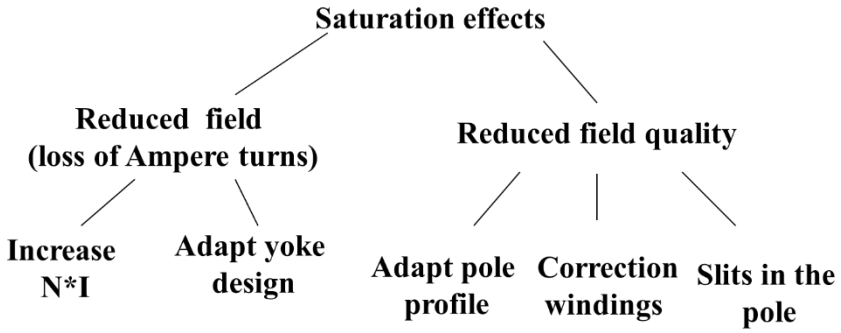


Fig. 20: Consequences of saturation and its countermeasures

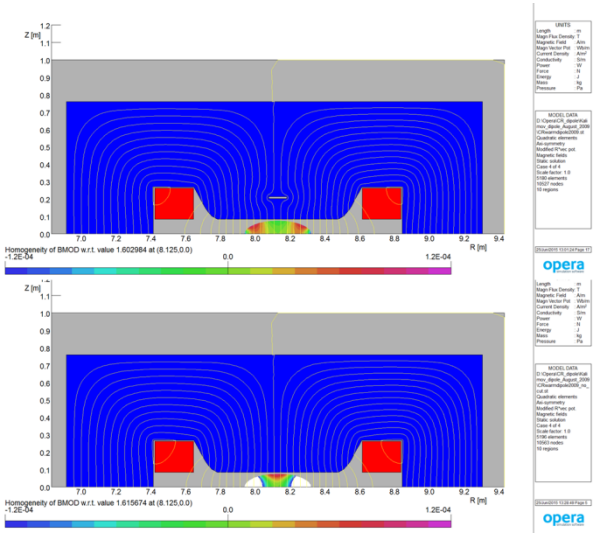


Fig. 21: Methods to minimize the influence of saturation on the field quality: H-type dipole with holes in the centre of the pole.

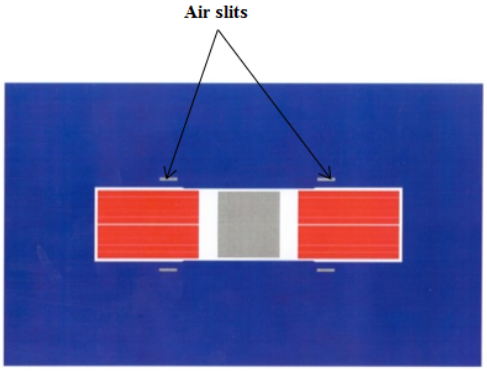


Fig. 22: Methods to minimize the influence of saturation on the field quality: window-frame type dipole with slits on the sides of the pole.

5.3 Curved yokes

As the beam describes a circular path inside the magnet, the sagitta 'h' comes into play. It depends on the radius r and the deflection angle α and is proportional to the square of the magnet length l.

$$h = r \cdot (1 - \cos(\alpha/2)) \propto l^2 .$$

Building a straight dipole is much simpler and therefore less costly than a curved one. But a wider aperture needs to be provided which increases the magnet's cross-section and is a cost driver too.

If a straight magnet is rotated by half of its deflection angle with respect to the incoming beam and shifted by half of the sagitta, its aperture width is used most efficiently (illustrated in Fig. 23). In that case, the mechanical aperture of the magnet needs to be the sum of beam width and sagitta. The ratio of beam width to sagitta is used for deciding between straight and curved yokes.

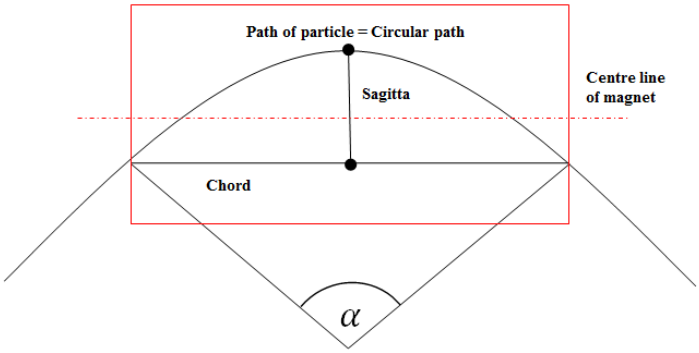


Fig. 23: Curved beam path in a straight dipole magnet

Laminated magnet yokes are typically stacked from fine blanked laminations. For curved magnets, they are stacked with their corresponding radii. Therefore, a curved stacked yoke has the same radii for the poles and back legs but different centres, whereas a massive yoke would have different radii but the same centre for all curved parts (depicted in Fig. 24). The useful aperture of a stacked yoke is therefore reduced in dependence of the deflection angle given by the reduced aperture,

$$b = a \cdot \cos\frac{\alpha}{2}.$$

To limit this effect, magnets with large bending angles are fabricated from several stacks. The ends of the individual stacks are cut under the appropriate angle and fixed together to one yoke.

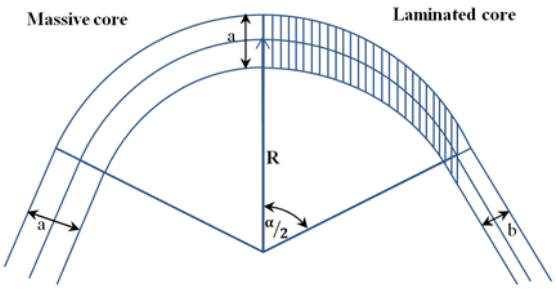


Fig. 24: Comparison of curved yokes: left side shows a massive yoke. All radii have the same centre. The useful aperture is hence constant. Right side shows a laminated yoke. All radii are identical but the centres are different. The useful aperture at the entrance of the magnet is reduced.

6 Special magnets

6.1 Introduction

In the following, two criteria will be used to distinguish special magnets from normal ones. The first are magnets with special features or materials in their designs. These are the following.

- Magnetic septa -> separate report by M. Paraliev.
- Kicker magnets -> separate report by M. Barnes.
- Radiation resistant magnets.
- Special yoke shapes.
- Integrated magnets.
- Magnets with cobalt iron yokes.
- ...

The other kind of special magnets are those which cannot be described by multipoles as shown in section 3.1. These are the following.

- Solenoids.
- Magnetic horns.
- Toroids.
- ...

6.2 Radiation resistant magnets

In accelerators, beam losses always occur that could either be intended or unavoidable. Intended beam losses occur on targets as well as during charge state separation or mass separation. Examples for unavoidable losses are charge exchange reactions on residual gas or electrical septum wires as well as resonances.

The material of the magnet must be chosen with respect to the expected radiation level and intended life time. The most sensitive component is typically the coil insulation.

Standard epoxy resins have radiation resistances of the order of 10^7 Gy. This can be increased by one order of magnitude by using improved plastics as polyimide or isocyanates. If even higher radiation resistances are required, a fully inorganic design must be chosen. A radiation resistance of more than 10^9 Gy can be realized.

Using improved plastics does not influence the conductor and coil design itself (Fig. 25). But the tooling for coil potting must be adapted to the different procedures. This can be for instance cooling of moulds for coil potting.



Fig. 25: Coils for an external beam line at KEK potted with improved plastics. At first glance no difference to standard coils is to be seen (photo K. H. Tanaka).

The fully inorganic solution requires a completely different design (as in Fig. 26). Metal oxide insulated conductors (MIC) are used for the coil. They consist of an inner copper part which carries the current. It is surrounded by a MgO-layer serving as insulator. The structure is held together by another copper sheath. This set-up requires much larger bending radii in comparison to standard copper conductors. Special care must be taken for joints and end caps as MgO is hygroscopic and its insulation properties are lost in humid condition. MIC are available as hollow conductors and without central bore. In high radiation fields, the combination of current and cooling water might lead to radiolysis. Then, MIC without bore holes are used and dedicated radiators are added to the coil structure. The complete coil structure is finally potted in solder.



Fig. 26: Radiation resistant dipole prototype for the SuperFRS pre-separator (photo G. Otto, GSI Darmstadt)

6.3 Special yoke shapes

External requirements may demand special yoke shapes. In most cases, these are limitations by surrounding beamlines and installations of the same accelerator, e.g., beam lines for injection and extraction as well as branching beam lines.

Most common are the C-shaped yokes of the magnetic septa (see separate report by M. Paraliev). But, standard H-type dipoles can also be re-designed to fit the requirements for branching points. Figure 27 shows the 45° high energy beam transport (HEBT) bending dipole for the cancer

therapy facility HIT in Heidelberg. This dipole is used to branch the beam to the different treatment areas. Its yoke has an opening to accommodate the straight beam pipe. The removed iron is magnetically compensated by an iron block added to the yoke in the vicinity of opening.

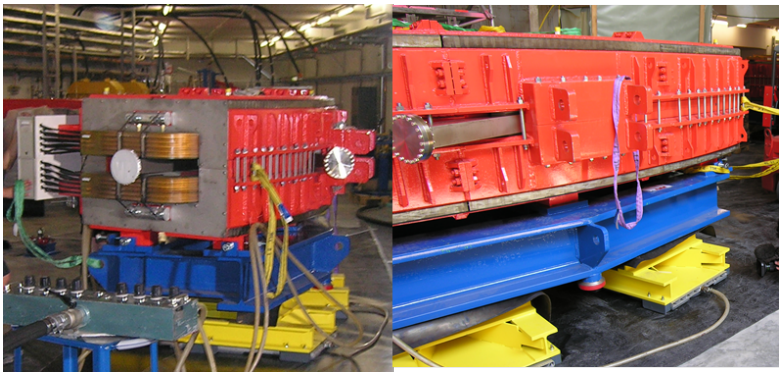


Fig. 27: The 45° HEBT dipole for the cancer therapy facility HIT in Heidelberg. The opening accommodating the straight beam pipe is magnetically compensated by an iron block (Photo C. Will, GSI Darmstadt).

Other examples for specially shaped yokes are the ones used in synchrotron light sources. The synchrotron light is tangentially emitted at the bends of the synchrotrons. Magnetic elements after the bends must not interfere with the beam lines for the synchrotron light. Therefore, special yoke shapes need to be used. Figure 28 shows as example a sextupole of the Australian Light Source.

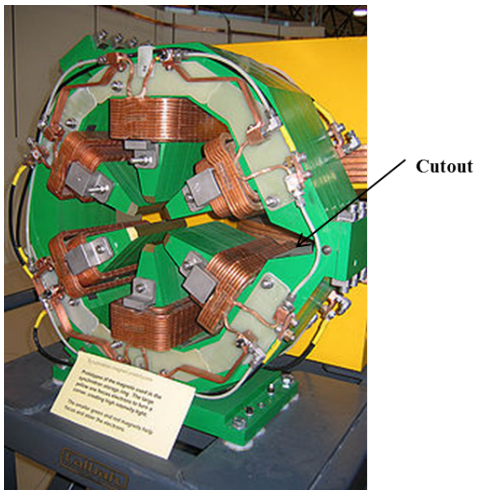
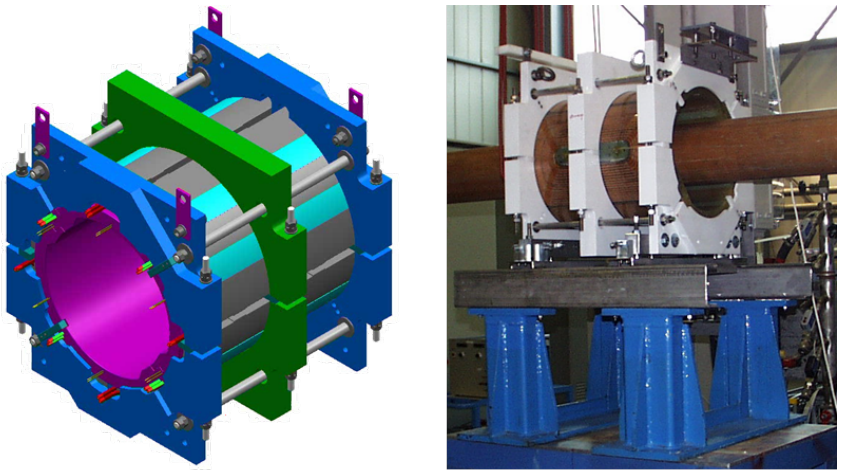


Fig. 28: Prototype sextupole for the storage ring of the Australian Light Source. Its yoke has a cut out at the right side to generate the clearance for the beam line [5].

6.4 Integrated magnets

Several different magnets can be integrated into one body for saving space or improving the overall stability.

The first example shows the corrector multipoles used in the SIS18 at GSI (Fig. 29). The longitudinal space is very limited. Therefore, three multipoles (normal quadrupole, skew quadrupole, and skew sextupole) are integrated into one body. All magnets are iron free. The saddle coils are arranged concentrically.



(a) (b)

Fig. 29: Corrector multipoles for SIS18 at GSI: (a) computer model of a correction magnet containing a normal quadrupoles, a skew quadrupole and a skew sextupole, (b) corrector multipole during testing. The brown tube inside the aperture contains the measurement coil.

Another example is the main magnets for the MAX IV storage ring in Lund (Fig. 30). The magnetic elements (e.g., one dipole, two quadrupoles, three sextupoles and two steering magnets) are integrated into a common yoke. The upper and the lower half of the yoke are precision machined iron blocks. The yoke itself acts as common girder for all magnets as well as for flux guidance. All magnetic elements are well aligned with respect to each other during the manufacturing process. Therefore, the alignment process of the machine itself is accelerated. The position of the magnetic elements within one main magnet is not influenced by ground settlements.

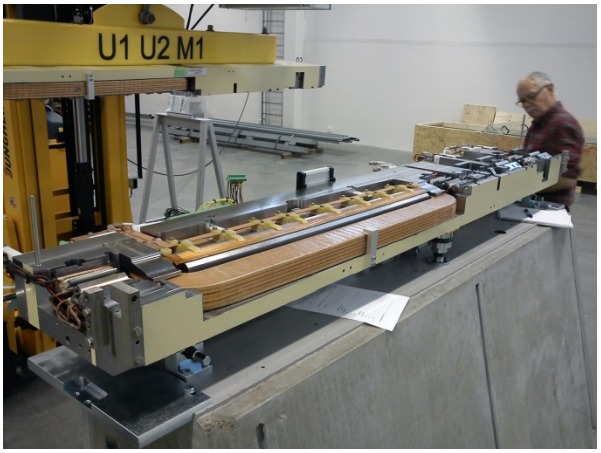
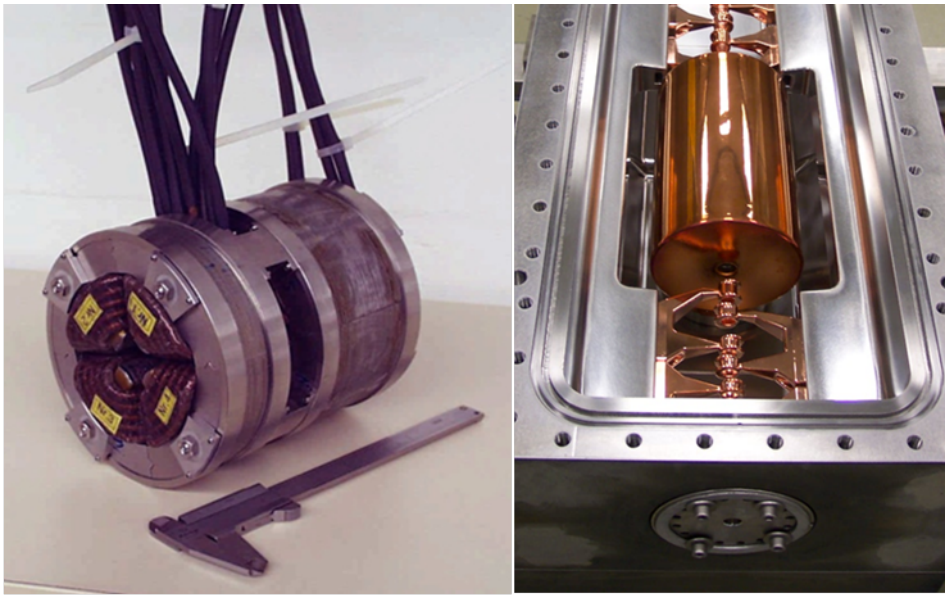


Fig. 30: Main magnet for MAX IV. This type (U1) contains one main dipole, two quadrupoles, three sextupoles. as well as *x*- and *y*-steering magnets [6].

6.5 Special yoke materials

As an example of special yoke materials, cobalt iron (CoFe) will be used here. By adding up to 50% Co to the yoke material, the saturation flux density can be increased to about 2.25 T. This enables compact design with high fields or field gradients. This advantage has to be paid for with a much higher price (up to 200 €/kg). Additionally, the manufacturing needs special care as the material is brittle. Also, care must be taken to avoid activation of Co during operation.

The examples show internal quadrupoles of an IH-type LINAC (seen in Fig. 31). The quadrupole yokes are made of 0.35 mm cobalt iron sheets. The pole base is enlarged to the maximum. Single layer coils are used. Gradients up to 124 T/m inside the 20 mm diameter aperture can be achieved. The quadrupole triplets are finally integrated into copper-plated drift tubes, which is another example for integrated magnets.



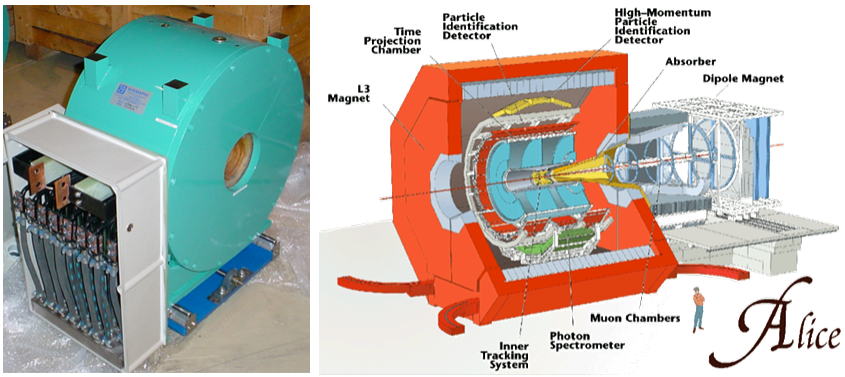
(a) (b)

Fig. 31: (a) Prototype quadrupole doublet for high field gradients with CoFe yoke, (b) fully integrated quadrupole triplet inside an IH-type LINAC.

6.6 Solenoids

In the simplest case, a helical winding represents a solenoid. The main field is parallel to the axis and for long solenoids the magnetic field inside is constant. In contrast to quadrupoles, a solenoid focusses both planes simultaneously. For explaining the focusing effect, both the main field and the stray field have to be taken into account. The stray field at the ends of the solenoid is radially oriented. A charged particle approaching the solenoid, experiences an azimuthal force due to the radial field. The created azimuthal motion causes the main field of the solenoid to produce a force that is directed towards the axis of the magnet. The charged particle beam is focused independently from the *x*- and *y*-plane.

Sizes vary from several tens of centimetres to several tens of metres (Fig. 32). Depending on the required focussing strengths, normal conducting or superconducting solenoids are used. Typical applications are low-energy beam lines in the vicinity of ion sources (mostly normal conducting), linear accelerators (often superconducting), and magnets for particle tracking in experiments (normal or super conducting).



(a) (b)

Fig. 32: (a) The solenoid used in the low energy beam transfer line of the Cancer Therapy Facility in Heidelberg having a diameter of ~0.6m and a length ~0.3m, it weighs 0.5 t, (b) the central solenoid of Alice [7] is ~12 m in diameter and ~12 m in length. Its weight is ~10,000 t. Both magnets have central field of about 0.55 T.

6.7 Magnetic horns

A magnetic horn (or van der Meer horn) is a pulsed high-current focusing device. It is used to focus the secondary beam after a production target. The central path is free of magnetic field and matter. Therefore, the primary beam is not influenced and cannot damage the horn. Off-axis (secondary) particles are focused (shown in Fig. 33).

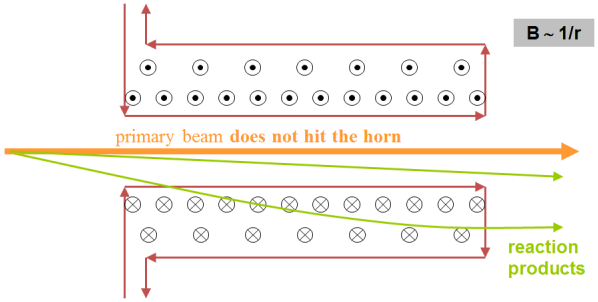
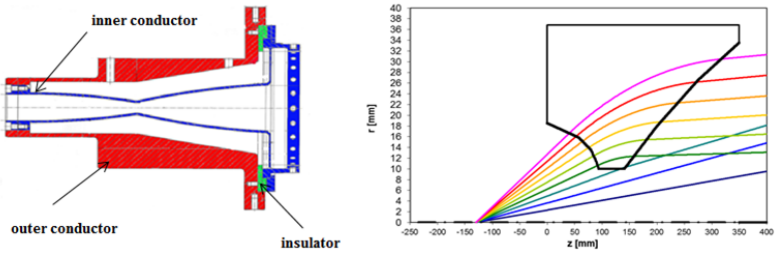


Fig. 33: Principle of a magnetic horn. A rotational magnetic field is generated by a single turn high current set up (fig. K. Knie, GSI Darmstadt).

A magnetic horn will be used in the pbar separator of the FAIR project to focus antiprotons (shown in Fig. 34). The planned operating data are 400 kA and 15 kV. The inner conductor of the horn is shaped in a way that a parallel antiproton beam is generated independent of its original angle to the axis.



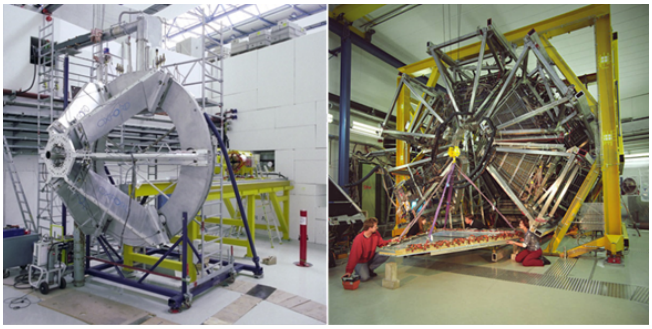
(a) (b)

Fig. 34: (a) Cross-section of the magnetic horn for the pbar separator at FAIR, (b) trajectories of antiprotons with different angles to the axis through the horn. A parallel beam is formed (figure K. Knie, GSI Darmstadt).

6.8 Toroids

Toroids exhibit a donut shaped magnetic field. The magnetic flux is closed and no distinguished poles exist. Outside the toroid, the stray field is weak.

The toroid magnet system of the High Acceptance Dielectron Spectrometer (HADES) at GSI consists of six super-conducting air coils (see Fig. 35). In a volume of 9 m^3 up to 3.6 T can be achieved. The space between the coils is to accommodate the different detector systems. They consist of more than 82,000 individual detectors.



(a) (b)

Fig. 35: (a) The toroid magnet system of the HADES experiment at GSI. Six super-conducting air coils are integrated in the spokes, (b) installation of the detector systems into the toroid.

7 Examples

7.1 Cancer therapy facility HIT in Heidelberg

In this section, examples for magnets described above are depicted with the magnets designed and produced for the cancer therapy facility HIT in Heidelberg.

The HIT centre is a dedicated therapy facility for treating cancer tumours. It is the first of a kind facility in Europe to provide treatments using both protons and carbon ions. Another striking feature of the HIT is its ability to provide a 360° rotating beam for radiation therapy. The HIT facility was in general developed by the GSI Helmholtz Centre for Heavy Ion Research in Darmstadt with contributions from MT Aerospace and a co-operation between Siemens Healthcare and experts at HIT-Betriebs GmbH. The entire project was funded equally by the Heidelberg university hospital and by the German government. The centre has been treating patients since 2008. Being almost half the size of a soccer field, it encompasses a gigantic device (discussed shortly).

HIT has a footprint of approx. 70 m × 70 m. An overview of the facility can be seen in the Fig. 36. Being a compact accelerator, most magnet designs and features described above can be found here. For completeness, some magnetic septa are also shown.

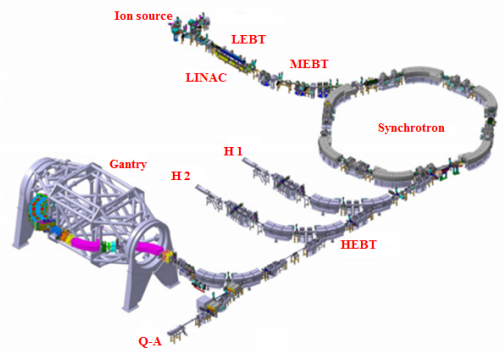


Fig. 36: Cancer therapy facility HIT [8]

The facility is equipped with two ion source branches. The electron cyclotron resonance ion sources produce light ions ranging up to oxygen. The magnets in the low energy beam transport (LEBT), guide the beam to the linear accelerator (LINAC). After the LINAC, the beam reaches an energy of 7 MeV/u which equals about 10% the speed of light. This is followed by a synchrotron with a circumference of about 65 m, where the beam attains at maximum approximately 73% the speed of light which equals a beam rigidity of 6.6 Tm. After the required energy is reached (please see Table 5), the beam is distributed by the HEBT to several beam lines. H1 and H2 are horizontal beam lines for patient treatments. The third branch guides the beam along an isocentric *gantry* (the gigantic device mentioned earlier). There also exists a fourth beam line for quality assurance (Q-A) and R&D purposes.

Table 5: Beam parameters for the HIT

Parameter	Protons	Light ions (e.g., carbon)
Energy	48-221 MeV/u	88-430 MeV/u

Aiming on a high reliability all magnets were designed conservatively. Wherever possible, existing designs and tooling were reused.

7.2 LEBT magnets

The ion beam is electrostatically extracted from the ion sources. Due to its low energy, it can by influenced by relatively small magnetic fields. This is visible in the design of the magnets in the LEBT.

At first the beam is focused by a solenoid already described in section 6.6.

The LEBT quadrupoles, one singlet before and one triplet behind the analysis dipole provide further focusing (see Fig. 37). Being low gradient (3.2 T/m) quadrupoles, these make use of the standard design type I for their yokes. The parallel sided poles are surrounded by a cylindrical coil. To minimize the crosstalk between the magnets, they are equipped with mirror plates.

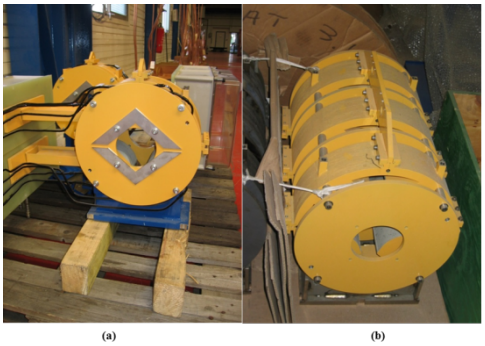


Fig. 37: (a) LEBT quadrupole singlet, (b) quadrupole triplet yoke

The analysis dipole is used for the magnetic separation of the charge state distribution delivered by the ion sources (shown in Fig. 38). It is an H-type dipole with a deflection angle of 90°. Due to the low energy, field (0.2 T) and bending radius are small (400 mm). A production using stamped laminations was geometrically impossible. Rough cut laminations were glued to a rigid block. The final yoke structure with one curved and one straight coil window was milled. The coil is D-shaped to avoid a negative curvature on the inside.

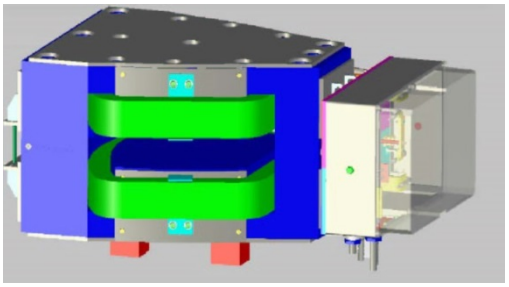


Fig. 38: Analysis dipole for charge separation behind the ion source

The switch yard dipole is used for funnelling the ion beams from the two ion sources to the linear accelerator (seen in Fig. 39). It is a simple, straight H-type dipole with a 0.1 T field.

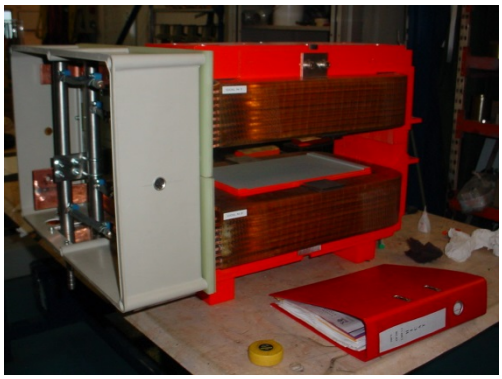


Fig. 39: Switch yard dipole

The LEBT is equipped with three pairs of combined horizontal/vertical LEBT steering magnets (Fig. 40). This allows the compensation of angular errors as well as axis offsets of the beam in both ion source branches and in front of the LINAC. The steering magnets have a window-frame design. The cylindrical coils are wound around the yoke. Due to the low field (0.0252 T), the power is low and no active cooling is required.

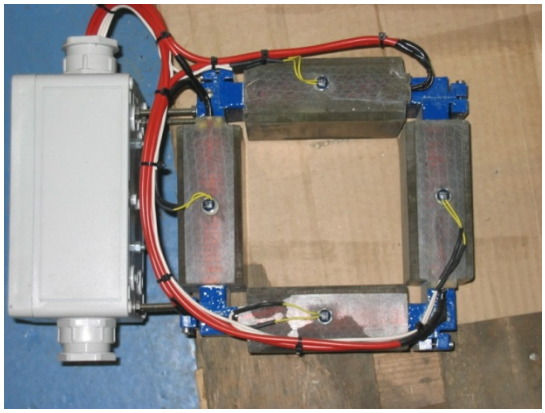


Fig. 40: LEBT steering magnet during pre-assembly

The focusing into the LINAC structure is done by another solenoid of the same type as described in section 6.6.

7.3 LINAC magnets

The LINAC consists of two radio frequency accelerating structures: one Radio Frequency Quadrupole (RFQ) tank and an Interdigital H-mode (IH) tank. The matching section between both tanks is

equipped with a quadrupole doublet. It makes use of the same quadrupoles as inside the IH-tank and directly behind the IH-tank. They were described in section 6.5.

The matching section contains in addition, a combined horizontal/vertical steering magnet (as in Fig. 41). To be as compact as possible it comprises a window-frame design with four single layered cylindrical coils and an extremely short yoke. It is integrated together with the quadrupole doublet into a stainless steel housing.

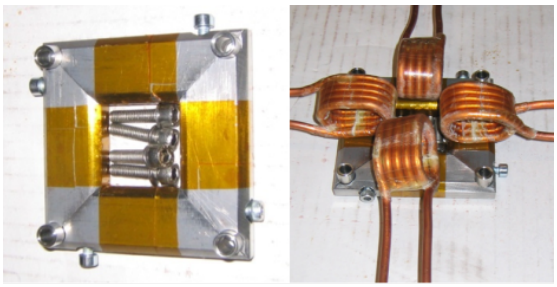


Fig. 41: Matching steering magnet: (a) pre-assembled yoke of the matching steering magnet, (b) yoke and coils during assembly preparation.

7.4 MEBT magnets

The MEBT provides the beam transport to the synchrotron as well as a so called stripping section. Lighter ions are stripped here to higher charge states (e.g., C^{3+} -> C^{6+}).

Two 25° dipoles bend the beam into the synchrotron injection channel and provide charge state separation after the stripper (shown in Fig. 42). A curved H-type design is used. The magnets were stacked from stamped laminations. Their field is 0.57 T.

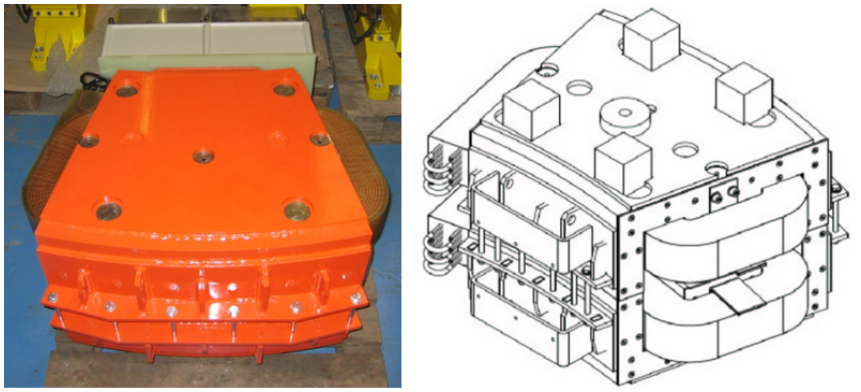
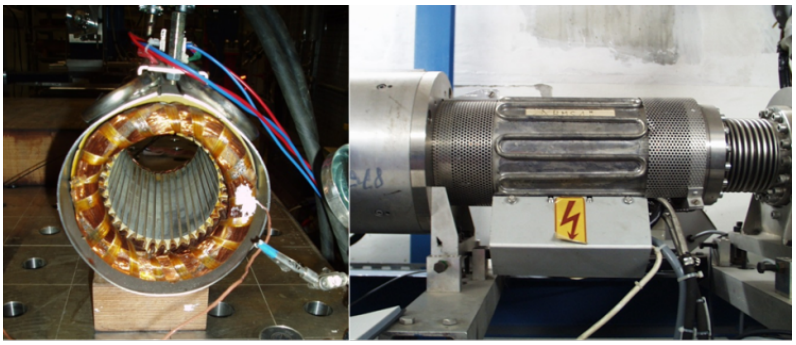


Fig. 42: MEBT dipole between linear accelerator and synchrotron

The quadrupoles in the MEBT make use of the same cross-section as the HEBT quadrupoles and are described with them in section 7.6.

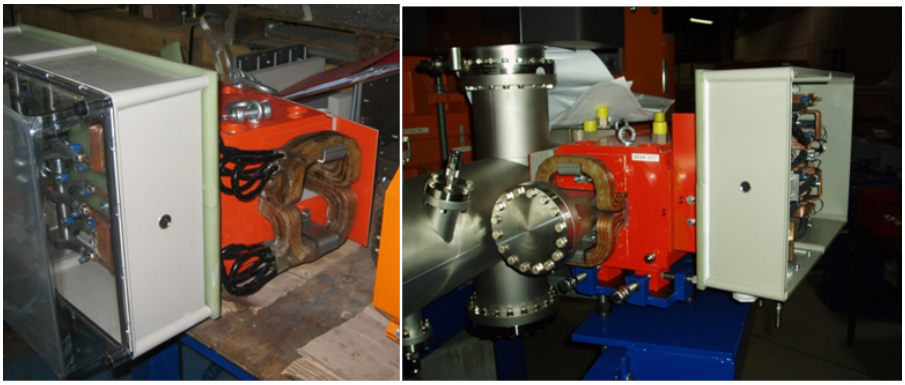
The MEBT is equipped with horizontal/vertical steering magnets, too. A low cost design was intended during their original development at GSI in the 1970s. The stator of a standard electromotor is used as yoke. The coils have a $\cos\theta$ design. Cooling is provided indirectly by a radiator wrapped around the yoke (shown in Fig. 43).



(a) (b)

Fig. 43: (a) Beam pipe steering magnet during test at GSI, (b) magnet integrated in a beam line

The injection into the synchrotron is provided by a curved window-frame C-dipole. It has a curved yoke (15°). The field is 0.42 T. The stray field on the circulating beam is reduced by a soft iron screen. This type of magnet is often called magnetic septum. In the Heidelberg case we called it a inflector dipole (shown in Fig. 44).



(a) (b)

Fig. 44: Inflector dipole into the synchrotron: (a) entrance of the inflector dipole, (b) exit of the inflector dipole during installation. To the left of the magnet, the beam line for the circulation beam can be seen.

7.5 Synchrotron magnets

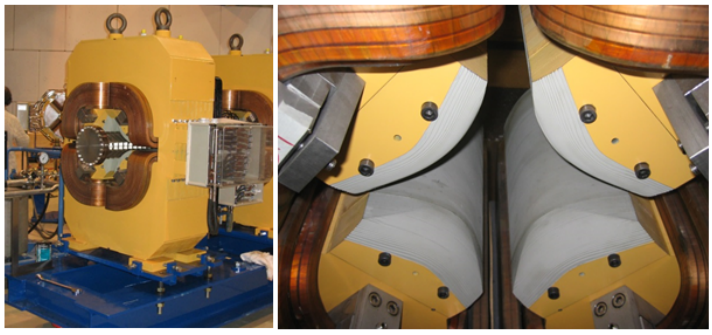
The main dipoles of the synchrotron are H-type dipoles with a deflection angle of 60°. Three curved stacks are connected to the dipole to limit aperture reduction. All six dipoles are operated in series with one power supply. Therefore, magnet-to-magnet identity has to be taken care of. Removable end plates allow the adjustment of the magnetic length by insertion of additional laminations (see Fig. 45). The field is 1.53 T with a ramp rate of 1.53 T/s. Correction coils are integrated into the main coils to provide horizontal steering.



(a) (b)

Fig. 45: (a) Synchrotron dipole with 60° bending angle, (b) insertion of an additional lamination for adjusting the magnetic length.

The synchrotron quadrupole relies on a figure-of-eight design and has a gradient of 7 T/m. The yoke cross-section is identical to the quadrupoles in the synchrotron SIS18 at GSI. The coils were adapted from quadrupoles in the injection channel to SIS18. The twelve quadrupoles are operated with four power supplies having three quadrupoles in series each. As with the dipoles, magnet-to-magnet identity has to be taken care of. Removable end pieces (see Fig. 46) allow the adjustment of the magnetic length by insertion of additional laminations.



(a) (b)

Fig. 46: Synchrotron quadrupoles: (a) two synchrotron quadrupoles on a common girder during pre-assembly, (b) removable pole end pieces.

The sextupoles have a design similar to the standard quadrupole type II. The coils are cranked accordingly (seen in Fig. 47). The yoke consists of three parts, each containing two poles. The gradient amounts to $\frac{d^2B}{dx^2} = 26.7\ \mathrm{T/m^2}$. The cross-section of the yoke is identical to the sextupoles in SIS18, whereas a new coil design has been developed.

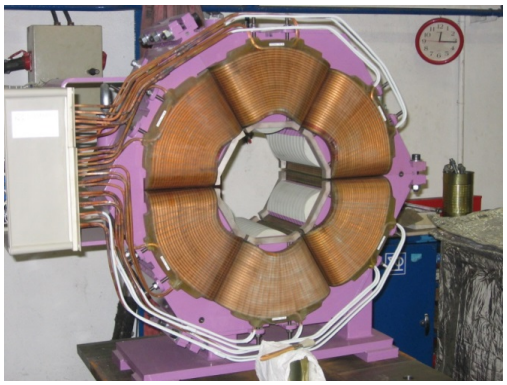


Fig. 47: Sextupole for the synchrotron during production

Vertical steering in the synchrotron is provided by window-frame design steering magnets with cylindrical coils. The field is 0.134 T. The field quality was improved by a spacer in the centre of the coils as seen in Fig. 48.

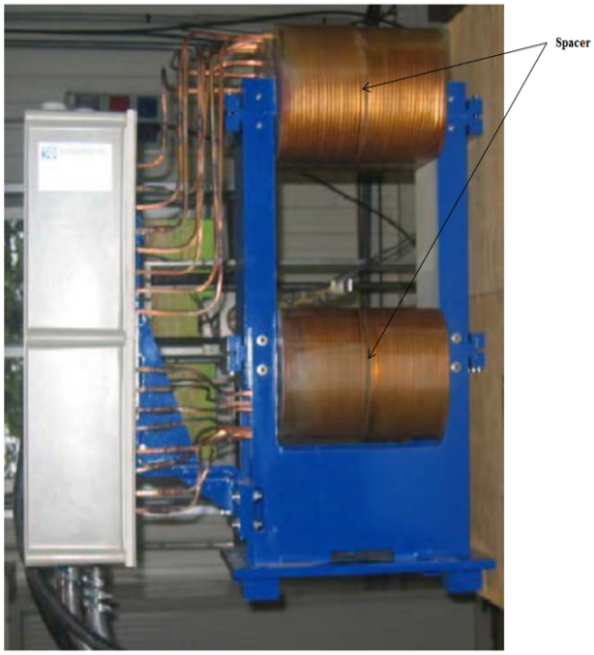


Fig. 48: Synchrotron vertical steering magnet during test

During injection, the beam must be stacked by very fast dipole magnets. For this purpose, the synchrotron is equipped with three so called bumpers (Fig. 49). These are two-turn window-frame dipoles. Because of the fast ramp rate (560 T/s), thin laminations are not sufficient for the suppression of eddy currents: The yokes are manufactured from powder composite materials. The absolute field level is small (0.0195 T).

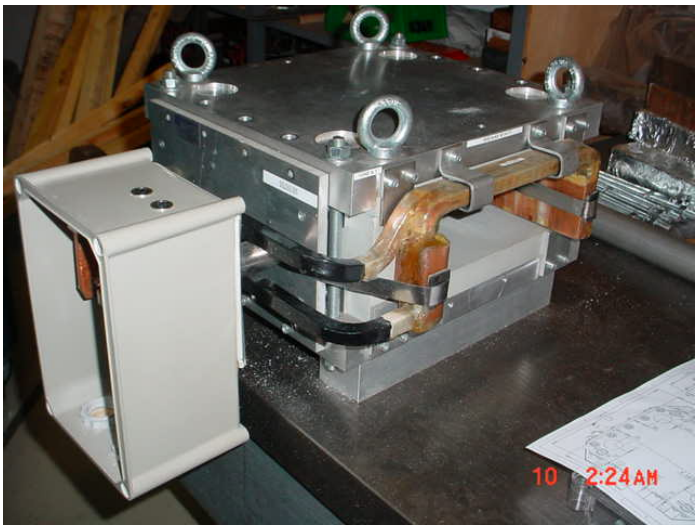


Fig. 49: Bumper magnet

The extraction channel is equipped with two magnetic septa.

The septum I is a straight window-frame dipole with C-shaped yoke. It has a deflection angle of 6.5°. To have enough horizontal space for the deflected beam, the gap widens after about two thirds of the magnet length (see Fig. 50). Coil and vacuum chamber are made accordingly. The coil is soldered from different copper parts for achieving tight bending radii and tolerances.



(a) **(b)**

Fig. 50: (a) Septum I seen from the circulating beam side, (b) vacuum chamber for Septum I. The lower branch is bent for the deflected beam.

Septum II is a curved window-frame dipole with a curved c-shape yoke (see Fig. 51). It has a deflection angle of 13.5° at 0.9 T. It is operated in series connection with the main dipoles of the synchrotron.

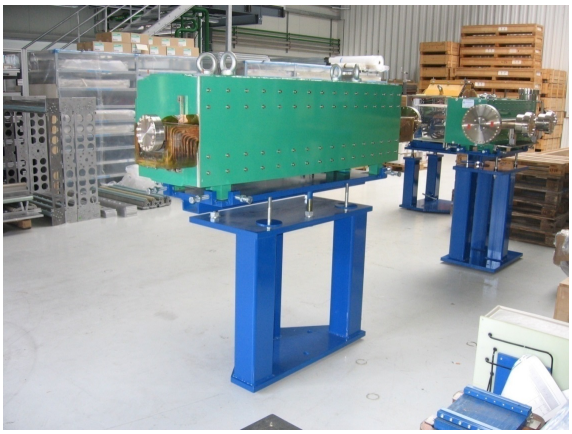
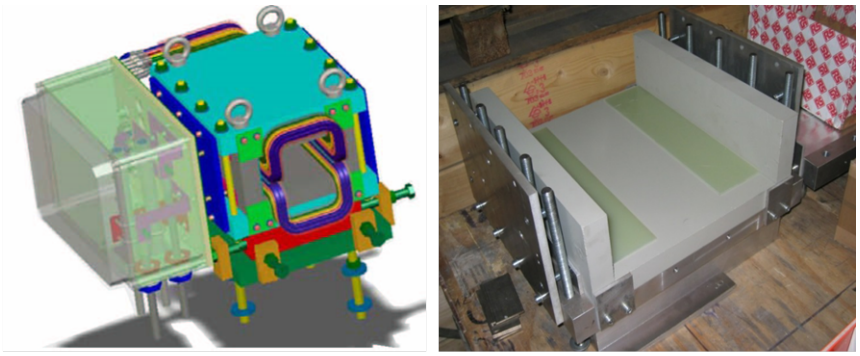


Fig. 51: Septum II

7.6 HEBT magnets

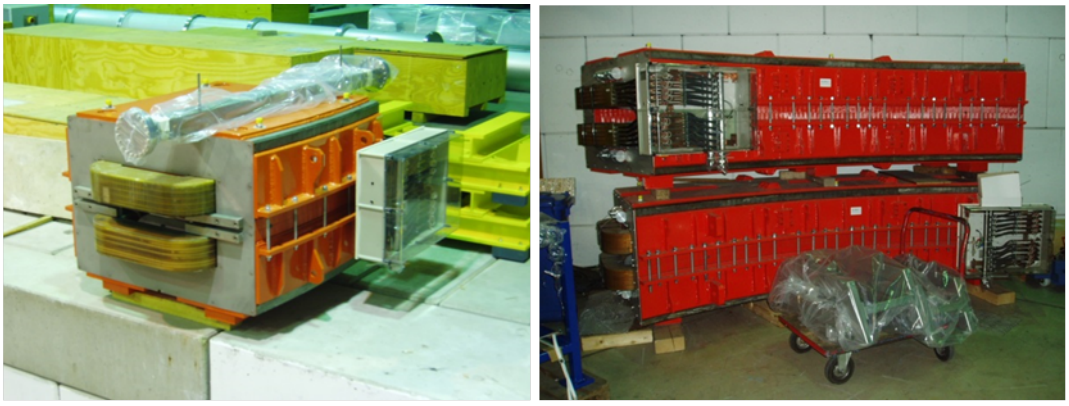
Directly after the synchrotron, the spill abortion bumper (as seen in Fig. 52) is located. In case of any failure it provides a fast switch off of the beam. Only if the magnet is switched on is the beam transported to the HEBT. If it is switched off, the beam is aborted into a dedicated beam dump. It is a fast magnet (1000 T/s) with only a few turns. The yoke is made of powder composite material. It has a deflection angle of 0.69° at 0.2 T.



(a) (b)

Fig. 52: (a) Spill abortion bumper, (b) lower half-yoke spill abortion bumper. It is manufactured from powder composite material. The yoke parts are kept together by a stainless steel structure.

The 15° and 45° dipoles are used in the high energy transfer beam lines (shown in Fig. 53). They have identical cross-sections. They are curved H-type dipoles with a field of 1.51 T. Zero field correction coils are integrated into the main coils for the usage at branching points (see also section 6.3). The 45° dipoles are made from three stacks to limit the aperture reduction. The poles have a hole for the compensation of saturation effects (see also section 5.2).



(a) (b)

Fig. 53: (a) The 15° HEBT dipole, (b) two 45° HEBT dipoles in storage

The quadrupoles in the HEBT, MEBT, and on the gantry make use of the same cross-section (depicted in Fig. 54). They are based on the standard design type II. The pole basis is enlarged to the maximum. The coils are cranked accordingly. The yoke lengths are adapted according to the requirements. The HEBT quadrupoles have a yoke length of 490 mm, whereas the MEBT and the gantry quadrupoles have a yoke length of 290 mm. The yoke cross-section and the coil design were copied from existing GSI beam line magnets.

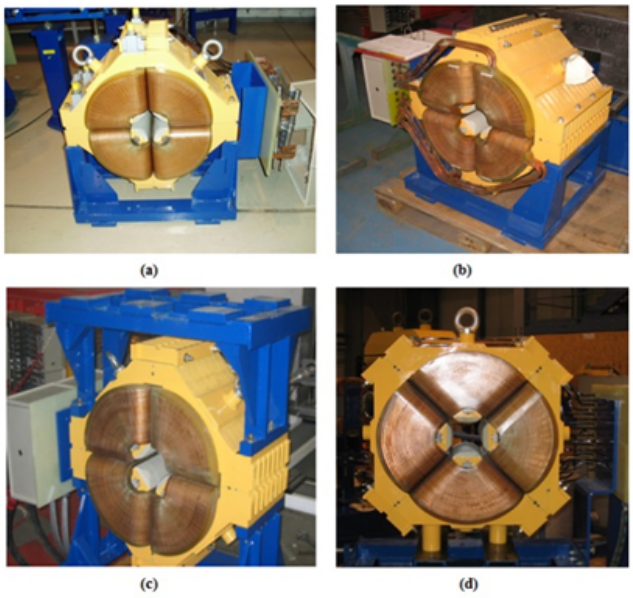


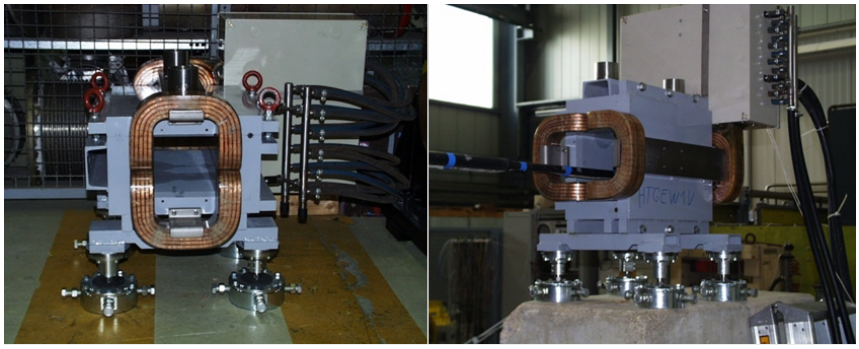
Fig. 54: MEBT, HEBT, and Gantry quadrupoles: (a) MEBT quadrupole (yoke length 290 mm), (b): HEBT quadrupole (yoke length 490 mm), (c) gantry quadrupole with top and bottom girder (yoke length 290 mm), and (d) skew quadrupole for matching section to gantry (yoke length 290 mm). The gradients for both lengths are similar. The shorter ones have 18.8 T/m and the longer ones 19.3 T/m.

Horizontal and vertical steering in the HEBT and on the gantry is provided by simple window-frame dipoles with cylindrical coil (seen in Fig. 55). Yoke and coils for horizontal and vertical steering magnets are identical. The field is 0.1 T.



Fig. 55: HEBT horizontal and vertical steering magnet

Each treatment area, as well as the Q-A area, is equipped with a horizontal and a vertical scanning dipole. They allow the scanning of the ion beam over the tumour cross-section as in a cathode ray tube. Apart from the interface to the girder, horizontal, and vertical scanning dipoles have an identical design. The fast ramped (62 T/s) straight window-frame dipole has a field of 0.31 T. Its lamination thickness is only 0.35 mm, whereas all other magnets have 1 mm laminations. To minimize the eddy current effects, the vacuum chamber is thin walled (0.35 mm). It is corrugated to withstand the atmospheric pressure.



(a) **(b)**

Fig. 56: Scanning dipoles during tests at GSI: (a) horizontal scanning dipole (GSI), (b) vertical scanning dipole

7.7 Gantry

While the treatment areas H1 and H2 deliver the ion beam horizontally, a gantry on the other hand allows the rotation of the beam around the patient. The patient is located in the isocentre. The beam line, including all necessary infrastructures, is rotated around the patient. A sketch of a so-called isocentric gantry is shown in Fig. 57.

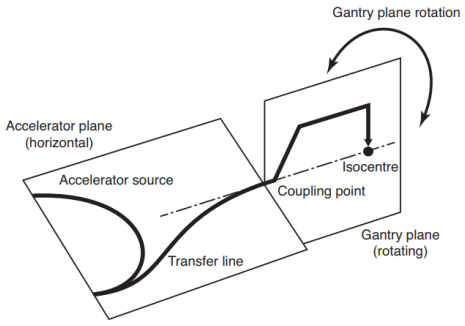


Fig. 57: Principle of an isocentric gantry [9]

With the isocentre, it is implied that the focal point is the same at every position for any rotational angle of the gantry. This eliminates the need for a patient lying on the isocentric axis to be moved vertically, with the gantry rotating horizontally. Another advantage of the gantry is that it has a large good field region.

The size of a gantry is solely determined by the use of normal conducting dipole magnets. These are not only used to transport the beam along the axis. They are additionally used to create an offset between the beam axis and the gantry's rotational axis and to further bend the beam towards the patient lying on the isocentric axis.

As may be recalled from section 5.2, the saturation of iron already starts at 1.5 T, thus restricting the maximum field on the beam axis to less than 2 T (approximately 1.8 T). Considering the beam rigidity of protons at the required energy, as in Table 5 with a field of 1.8 T, the gantry must have a radius of at least 5 to 6 metres, making it huge.

At HIT, as both protons and heavy ions are used and the gantry is significantly larger than conventional ones. With a maximum energy of 425 MeV/u and a rigidity of 6.57 Tm, a bending radius of 3.65 m is reached. With the normal conducting dipole technology, it results in a gantry of size ≈ 19 m in length and ≈ 15 m in diameter spanning up to 3 floors in height. The overall mass is 600 tonnes, 135 tonnes of which are contributed by the magnets that guide the beam to the gantry. Despite its huge rotating mass, a positioning accuracy of about ± 0.5 mm about the tumour is achieved (which is comparable with a proton gantry).

All magnets on the gantry are equipped with girders and alignment feet on the top and at the bottom. By these means, they are restrained inside the gantry structure and the required accuracy can be achieved.

As describe above, the gantry diameter is determined by the dipole field level. To keep the gantry compact, all dipoles on the gantry have a field of 1.81 T. Two 45° dipoles generate the necessary offset distance from the rotational axis (see Fig. 58). They are curved hybrid design dipoles: an H-type yoke is equipped with an additional inner bedstead coil between the poles. The outer coil is of bedstead type too. The yokes are made from two stacks to limit aperture reduction.



Fig. 58: A 45° gantry dipole during pre-assembly

The last magnet is the 90° gantry dipole. It has a window-frame design. It is the magnet with the largest aperture in the facility as it has to accommodate the scanning magnet's deflected beam. The curved yokes are made from three stacks to limit the aperture reduction (shown in Fig. 59). Holes beside the poles limit the influence of saturation (see section 5.2).

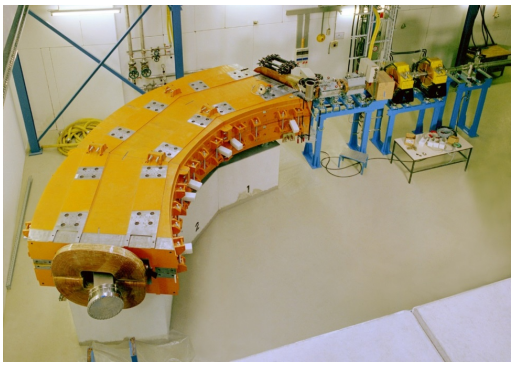


Fig. 59: The 90° gantry dipole during test at GSI

The quadrupoles, steering magnets, and scanning magnets on the gantry are, with the exception of the girders, identical to the HEBT magnets, which are described in section 7.6.

8 Acknowledgements

The author wishes to acknowledge support from numerous colleagues inside and outside GSI, especially from V. Srinivasan and P. Rottländer.

References

[1] J. Franklin, *Classical Electromagnetism*, (Dover Publications, New York, 2017).

[2] J. Tanabe, Iron Dominated Electromagnet Design, US Particle Accelerator School, June 2005, SLAC-R-754, http://www.slac.stanford.edu/cgi-wrap/getdoc/slac-r-754.pdf.

[3] http://www.luvata.com/.

[4] G. Schnell, *Magnete Grundlagen - Aufbau- Anwendungen*, (Verlag Karl-Themig München, 1973).

[5] https://commons.wikimedia.org/wiki/File:Aust.-Synchrotron,-Sextupole-Focusing-Magnet,-14.06.2007.jpg.

[6] M. Johannson, Beam Dynamics meets Magnets- II workshop, Bad Zurzach, 1−4 Dec 2014.

[7] https://cms.cern/detector/bending-particles

[8] D. Ondreka and U. Weinrich, GSI, Darmstadt, Germany, 'The Heidelberg ion therapy (HIT) accelerator coming into operation', Conf.Proc. C0806233 (2008) TUOCG01

[9] H. Owen et al., Technologies for delivery of proton and ion beams for radiotherapy, arXiv:1310.0237

Bibliography

Proceedings of the CAS-CERN Accelerator School: Magnets, Bruges, Belgium, 16-25 June 2009, edited by D. Brandt, CERN-2010-004, http://cdsweb.cern.ch/record/1158462.

Proceedings of the CAS-CERN Accelerator School: Measurement and alignment of Accelerator and Detector Magnets, Anacapri, Italy, 11-17 April 1997, edited by S. Turner, CERN-98-05, http://cds.cern.ch/record/318977

J.T. Tanabe, *Iron Dominated Electromagnets*, (World Scientific Publishing Co., 2005), https://doi.org/10.1142/5823.

IEEE, Proceedings of International Conferences on Magnet Technology, every 2nd year, e.g., 24th International Conference on Magnet Technology (MT-24), Seoul, 2015 [IEEE Trans. Appl. Supercond. 26 (2016) no. 4].

H.E. Knoepfel, *Magnetic Fields* (Wiley and Sons, New York, 2000).

Proceedings of the CAS–CERN Accelerator School: Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018)

Kicker Systems

M.J. Barnes
CERN, Geneva, Switzerland

Abstract
Each stage of an accelerator system has a limited dynamic range and therefore a chain of stages is required to reach high energy. A combination of septa and kicker magnets is frequently used to inject and extract beams from each stage. The kicker magnets typically produce rectangular field pulses with fast rise and/or fall times; however, the field strength is relatively low. To compensate for their relatively low field strength, the kicker magnets are generally combined with electromagnetic septa. The septa provide relatively strong field strength but are either DC or slow pulsed. This paper discusses injection and extraction systems with particular emphasis on the hardware required for the kicker system. Design options and computer aided design tools are also discussed. In addition, the interaction of the beam with the kicker system, primarily the kicker magnet, is considered.

Keywords
Injection; extraction; kicker; pulse power; beam coupling impedance; modulators.

1 Introduction

An accelerator stage has limited dynamic range: a chain of accelerator stages is required to reach high energy. Thus beam transfer into (injection) and out of (extraction) an accelerator is required. The design of the injection and extraction systems aims to achieve the following:

- minimize beam loss;
- place the newly injected or extracted particles onto the correct trajectory, with the correct phase space parameters.

A combination of septa and kickers is frequently used for injection and extraction. Septa can be electrostatic or magnetic: they provide slower field rise and fall times, but a stronger field, compared to kicker magnets. Some septa are designed to be operated with DC. Kicker magnets provide fast field rise and fall times, but relatively weak fields.

In general, a septum (plural: septa) is a partition that separates two cavities or spaces. In a particle accelerator, a septum is a device which separates two field regions. Important features of septa are an ideally homogeneous (electric or magnetic) field in one region, for deflecting the beam, and a low fringe field (ideally zero magnetic and electric field) next to the septum so as not to affect the circulating beam. Hence, a septum provides a space separation of circulating and injected/extracted beams. In contrast, a kicker magnet provides time selection (separation) of the beam to be injected/extracted: a kicker system is used for fast, single-turn, injection and extraction.

The processes of injection and extraction are covered in the proceedings of this CERN Accelerator School. The paper S*epta* discusses the hardware associated with septa. The present paper discusses the hardware associated with kicker magnets.

The field produced by a kicker magnet must rise/fall within the time period between the beam bunches (see sections 2 and 3). In addition, the magnetic field must not significantly deviate from the

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-005.229

flat top of the pulse or from zero between pulses (i.e., very small ripple/excursions): this is discussed further in the paper *Injection: Hadron Beams*, in the proceedings of this CERN Accelerator School. Typical field rise/fall times range from tens to hundreds of nanoseconds and pulse widths generally range from tens of nanoseconds to tens of microseconds. If a kicker exhibits a time-varying structure in the field pulse shape, this can translate into small offsets with respect to the closed orbit (betatron oscillations). Thus a fast, low-ripple, kicker system is generally required.

2 Single-turn (fast) injection

Figure 1 shows an example of fast single-turn injection in one plane. The injected beam passes through the homogeneous field region (gap) of the septum: the circulating beam is in the field-free region (i.e., there is space separation of the injected and circulating beams). The septum deflects the injected beam onto the closed orbit at the centre of the kicker magnet; the kicker magnet compensates the remaining angle. The septum and kicker are either side of a quadrupole (defocusing in the injection plane) which provides some of the required deflection and minimizes the required strength of the kicker magnet.

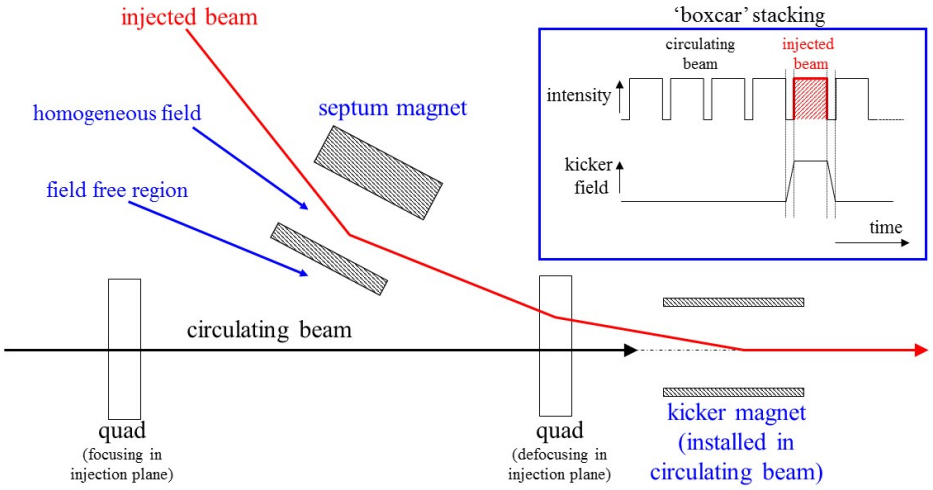


Fig. 1: Fast single-turn injection in one plane

When a kicker magnet is installed in a circular accelerator the circulating beam is typically in the aperture of the magnet. Thus the kicker field must rise from zero to full field in the time interval between the circulating beam and the start of the injected beam (Fig. 1, top right) and fall from full field to zero field in the time interval between the end of the injected beam and the subsequent circulating beam (Fig. 1, top right). The kicker system is described in more detail in section 5.

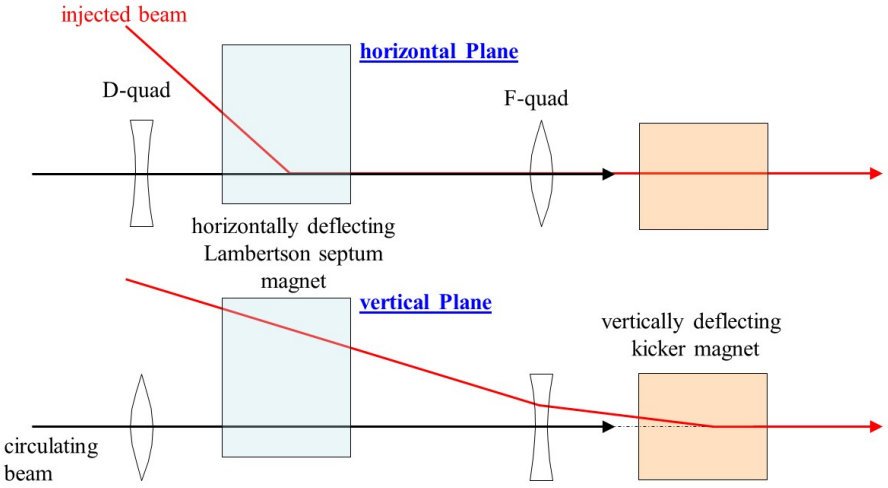


Fig. 2: Fast single-turn injection in two planes

Figure 2 shows an example of fast single-turn injection in two planes: a Lambertson septum is used for a two-plane injection scheme. The injected beam passes through the homogeneous field region of the septum: the circulating beam is in the field-free region of the septum. In the example shown in Fig. 2 the septum deflects the beam horizontally and the downstream kicker magnet deflects the beam vertically onto the closed orbit of the circulating beam. The septum and kicker are either side of an F-quadrupole (horizontally focusing but vertically defocusing) to minimize the required strength of the kicker magnet. The Lambertson septum magnet is discussed in more detail in the proceedings of this CAS, in paper *Septa*.

3 Single-turn (fast) extraction

Extraction is the process of ejecting a particle beam from an accelerator and into a transfer line or a beam dump at the appropriate time while minimizing beam loss and placing the extracted particles onto the correct trajectory, with the correct phase space parameters. Extraction usually occurs at a higher energy than injection, hence stronger elements (e.g., larger $\int B.\mathrm{d}l$) are required. At high energies many kicker and septum modules may be needed. To reduce the required strength of the kicker magnet, a closed orbit bump can be applied to bring the circulating beam near to the septum.

Figure 3 shows an example of fast single-turn extraction in one plane. The kicker magnet deflects the entire beam into the septum in a single turn [time selection (separation) of beam to be extracted]. The extracted beam passes through the homogeneous field region of the septum: the circulating beam, prior to extraction, is in the field-free region of the septum (space separation of the circulating and extracted beams). The septum deflects the entire kicked beam into the transfer line.

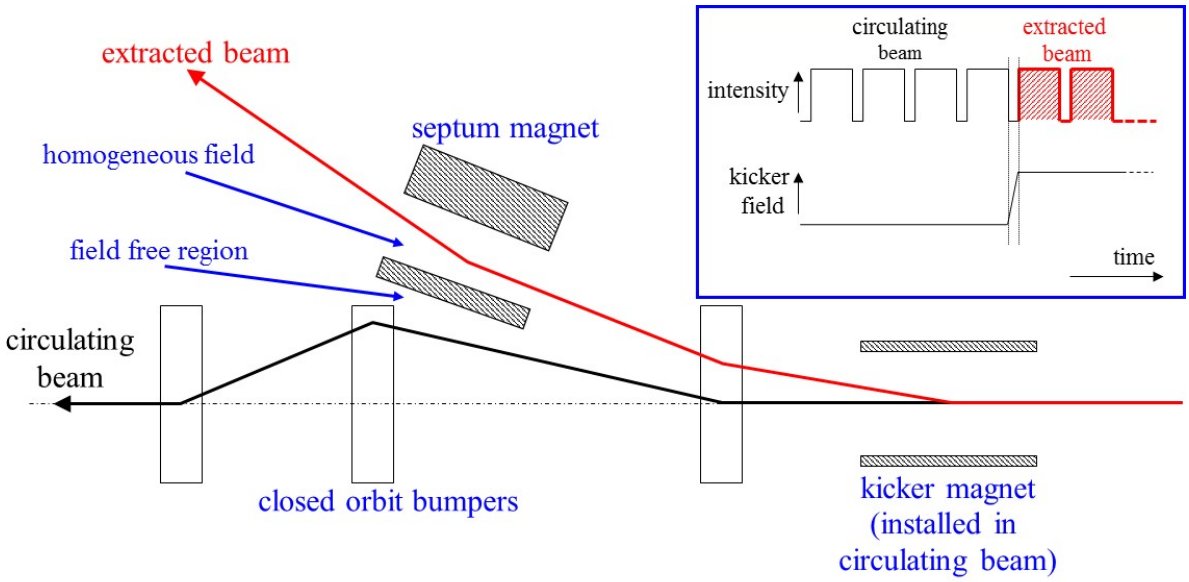


Fig. 3: Fast single-turn extraction in one plane

When a kicker magnet is installed in a circular accelerator the circulating beam is typically in the aperture of the magnet. Thus, the kicker field must rise from zero to full field in a beam-free time interval deliberately created in the circulating beam (Fig. 3, top right). The entire beam is generally extracted and hence a fast fall time is typically not required; however, sometimes, bunch-by-bunch transfers are made and then the field of the kicker magnets must have fast rise and fall times [1].

4 Deflection of beam

The force exerted on a charged particle, moving through electric and magnetic fields, is known as the Lorentz force. The Lorentz force is given by Eq. (1):

$$F = q\left[E + \left(v \times B\right)\right] . \tag{1}$$

In this equation:

F	is the force (N);
E	is the electric field (V/m);
B	is the magnetic field (T);
q	is the charge of the particle (C), and $q = Q \cdot e$;
e	is elementary charge ($1.602 \cdot 10^{-19}$ C);
Q	is a multiple of the elementary charge (e.g., Q=1 for a proton);
v	is the instantaneous velocity of the particle (m/s);
×	is the vector cross product.

Using the small angle approximation, $\tan(\theta) = \theta$, the deflection of a charged particle beam in a magnetic field is given by Eq. (2) [2]:

$$\theta_{B,x} = \left[\frac{0.2998 \cdot Q}{p}\right] \cdot \int_{z_0}^{z_1} \left|B_y\right| dz = \left[\frac{0.2998 \cdot l_{eff} \cdot Q}{p}\right] \cdot \left|B_y\right| . \tag{2}$$

In this equation:

B_y	is the magnetic flux density in the y-direction (T);
p	is the beam momentum (GeV/c);

l_{eff} is the effective length of the magnet [usually different from the mechanical length, due to fringe fields at the ends of the magnet] (m);
$\theta_{\text{B,x}}$ is the deflection angle, in the x-direction, due to magnetic field B_{y} (radians).

Using the small angle approximation $\tan(\theta) = \theta$, the deflection of a charged particle beam in an electric field is given by Eq. (3) [2]:

$$\theta_{E,x} = \left[\frac{Q}{(p \cdot 10^9) \cdot \beta \cdot} \cdot \int_{z_0}^{z_1} |E_x| dz \right] = \left[\frac{|V| \cdot l_{eff} \cdot Q}{d \cdot (p \cdot 10^9) \cdot \beta} \right] . \quad (3)$$

In this equation:

V is the potential difference between plates (V);
d is the separation of the plates (m);
E_{x} is the electric field in the x-direction (V/m);
β is a unit-less quantity that specifies the fraction of the speed of light at which the particles travel (v/c);
$\theta_{\text{E,x}}$ is the deflection angle, in the x-direction, due to electric field E_{x} (radians).

5 Kicker system

5.1 Overview

Figure 4 shows a simplified schematic of a kicker system. The main sub-systems ('components') of a kicker system are:

- pulse forming line (PFL) (coaxial cable) or pulse forming network (PFN) (lumped elements);
- kicker magnet;
- fast, high power, switch(es);
- resonant charging power supply (RCPS);
- transmission line(s) [coaxial cable(s)];
- resistive terminators.

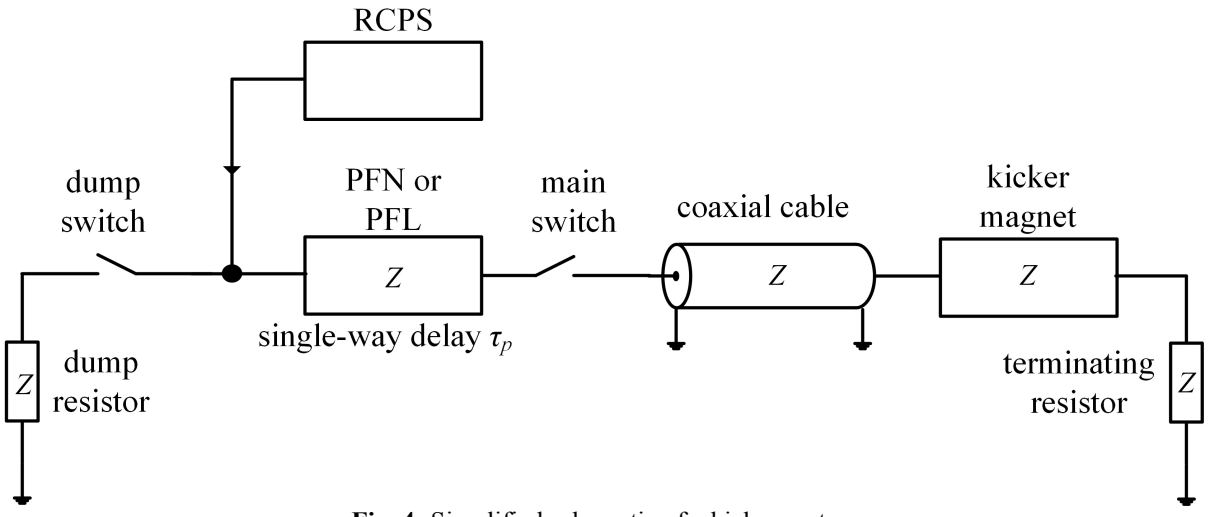


Fig. 4: Simplified schematic of a kicker system

5.2 **Pulse forming circuit: general case**

Figure 5 shows a simplified schematic of a pulse-forming circuit: the switch is initially open and the coaxial cable (PFL) is pre-charged, through the large valued resistor or inductor, to a voltage V.

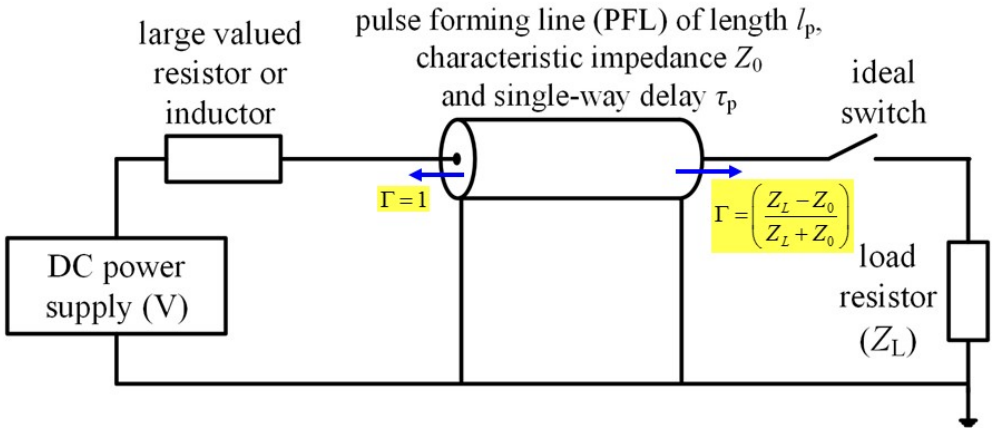


Fig. 5: Simplified schematic of a pulse-forming circuit

At time $t = 0$, when the ideal switch closes, the load voltage is given by Eq. (4):

$$V_L = \left(\frac{Z_L}{Z_0 + Z_L}\right) V = \alpha V \ , \tag{4}$$

In this equation:

$$\alpha = \left(\frac{Z_L}{Z_0 + Z_L}\right);$$

V_L is the load voltage (V);
V is the initial voltage to which the PFL is charged (V);
Z_L is the load impedance (Ω);
Z_0 is the characteristic impedance of the PFL (Ω).

Figure 6 shows a lattice diagram for the general case (impedances not necessarily matched) for the voltage on the PFL. At time $t = 0$, when the switch closes, a voltage pulse of '$(\alpha - 1)V$' propagates from the load end of the PFL towards the charging end. At the charging end of the PFL the reflection coefficient (Γ) is +1 and hence a voltage of '$(\alpha - 1)V$' is reflected back towards the load end of the PFL. At the load end of the PFL the reflection coefficient is given by Eq. (5):

$$\Gamma = \left(\frac{Z_L - Z_0}{Z_L + Z_0}\right) = \beta \ , \tag{5}$$

and hence a voltage of '$\beta(\alpha - 1)V$' is reflected back towards the charging end of the PFL, etc.

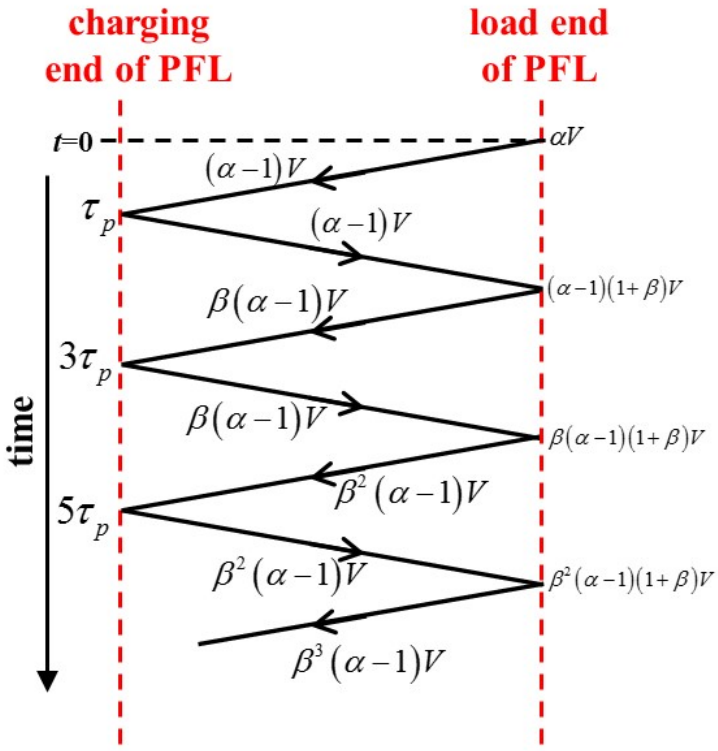


Fig. 6: Lattice diagram for the pulse forming circuit of Fig. 5 (general case)

Impedances need to be matched to avoid reflections, i.e., from Eq. (5), when $Z_L = Z_0 \rightarrow \beta = 0$. When the impedances are matched:

- PFN/PFL is charged to a voltage V by the RCPS;
- the main switch (MS) closes (Fig. 4) and, for a matched system, a pulse of magnitude $V/2$ is launched, through the coaxial cable, towards the kicker magnet. A voltage pulse of magnitude $-V/2$ propagates from the load end of the PFN/PFL towards the charging end;
- once the flat top of the current pulse reaches the (matched) terminating resistor (TMR), full field has been established in the kicker magnet (section 5.3.3);
- the duration of the pulse in the magnet can be controlled in length, between 0 and $2\tau_p$, by adjusting the timing of the dump switch (DS) (Fig. 4) relative to the MS.

Note: if the magnet termination is a short circuit, the magnet current is doubled but the 'fill time' of the magnet is doubled too (section 5.3.6). In this case, the DS may be an inverse diode: the inverse diode 'automatically' conducts when the PFN voltage reverses, at the charging end of the PFL/PFN, but there is no control over the pulse length in the magnet.

5.3 Kicker magnet

5.3.1 History

Figure 7 shows a 1970s 'plunging' kicker magnet, which was hydraulically operated [3]: the aperture was too small for the kicker to be in the beam line during circulating beam. Developments leading to higher current pulses permitted larger apertures: kicker magnets developed later at CERN were not hydraulically operated.

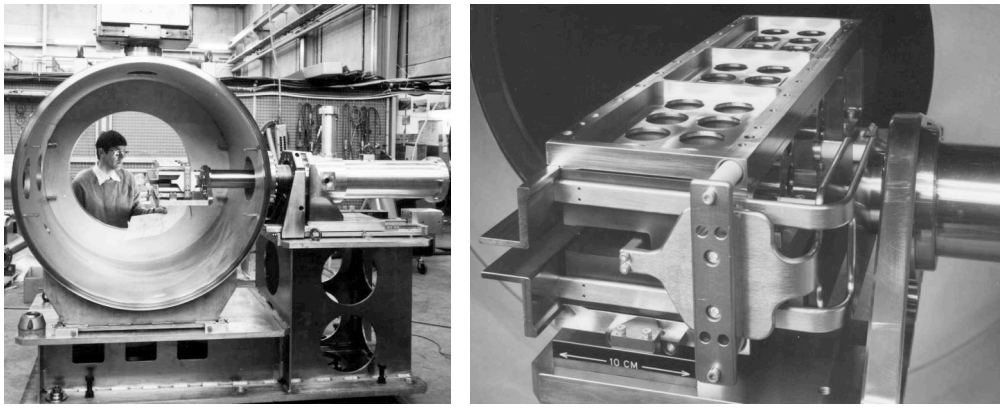


Fig. 7: A 1970s plunging kicker magnet for ejection from the CERN Antiproton Accumulator (AA)

5.3.2 Overview

Figure 8 shows a cross-section of a typical C-core kicker magnet with the return conductor closing the aperture. Fast kicker magnets generally have a ferrite yoke and are transmission line type magnets with a rectangular-shaped aperture of dimensions H_{ap} by V_{ap} (Fig. 8).

Assuming a high relative permeability for the yoke, the flux density in the aperture (B_y) of the kicker is given by Eq. (6):

$$B_y \cong \mu_0 \left(\frac{N \cdot I}{V_{ap}} \right), \tag{6}$$

where:

μ_0	is permeability of free space ($4\pi \times 10^{-7}$ H/m);
N	is the number of turns (usually 1 for a kicker magnet);
I	is magnet current (A);
V_{ap}	is the distance between the inner edges of the 'legs' of the ferrite (m).

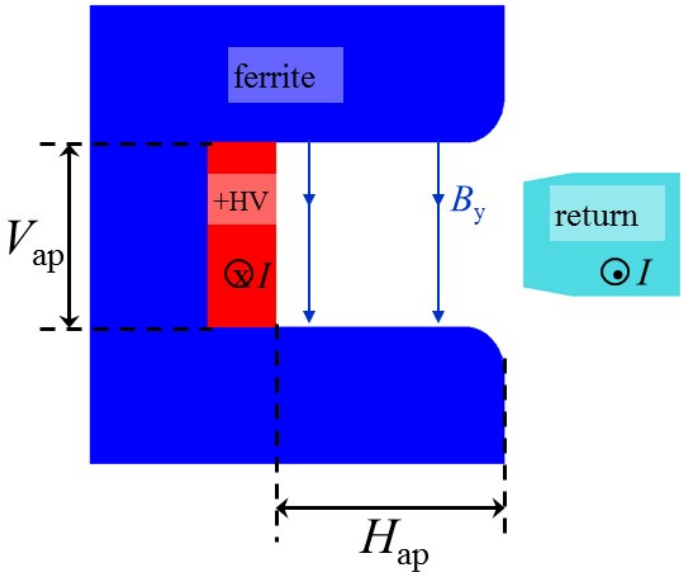


Fig. 8: Cross-section of a typical C-core kicker magnet

The skin effect and proximity effect result in current flow on the inside surface of both the high voltage (HV) and return conductors. Hence, assuming a high relative permeability for the yoke, inductance is given by Eq. (7):

$$L_{m/m} \cong \mu_0 \left(\frac{N^2 \cdot H_{ap}}{V_{ap}} \right) , \tag{7}$$

In this equation:

H_{ap} is the distance between the inner edge of the HV conductor and the end of the leg of the magnetic material (m);
$L_{m/m}$ is inductance per metre length of the kicker magnet (H/m).

Since kicker magnets generally need to be fast, they usually have a single-turn coil ($N = 1$): multi-turn coils are used only for slower, lumped-inductance, kicker magnets.

The deflections of a charged particle in a magnetic field and an electric field are given by Eq. (2) and Eq. (3), respectively. For a transmission-line kicker magnet, where ferrite C-cores are sandwiched between HV capacitance plates (see section 5.3.3.2), end effects result in an effective length of each end cell approximately $\left(V_{ap}/4\right)$greater than its physical length [4]. Thus the effective length of the kicker magnet is increased by an amount approximately equal to $\left(V_{ap}/2\right)$, in comparison to the physical length of the aperture.

5.3.3 Design options for kicker magnets

Design options for kicker magnets include the following [3].

- Type: 'lumped inductance' or 'transmission line' (with specific characteristic impedance (Z))
- Machine vacuum: install in or external to machine vacuum
- Aperture: window frame, closed C-core or open C-core
- Termination: matched impedance or short circuited

5.3.3.1 Lumped-inductance kicker magnet

Although a lumped-type magnet has a simple structure, in many cases it cannot be applied to a fast kicker system because of its impedance mismatch and its slow response [5]. The lumped-inductance kicker is generally useable only when a rise time above a few hundreds of nanoseconds is required. The lumped-inductance kicker either has a resistor in series with the kicker magnet input or else the resistor is omitted. In both cases, the kicker magnet only sees voltage during pulse rise and fall: the magnet voltage during field fall time is the opposite polarity compared too field rise. With a short-circuit termination, magnet current is doubled for a given PFN/PFL voltage and system characteristic impedance.

For a magnet inductance L_m with a resistor in series with its input, combined with a pulse generator of impedance Z, the rise of the magnet current is exponential with a time constant t_{cl} given by Eq. (8):

$$t_{cl} = \left(\frac{L_m}{Z + R} \right) . \tag{8}$$

The 5% to 95% rise time of the current is as long as three time constants. To help to overcome the long rise time, a capacitor can be connected in parallel with, at the entrance of, the lumped-inductance magnet, but this can provoke some overshoot.

5.3.3.2 Transmission-line kicker magnet

To overcome the long rise time, the first transmission-line kicker magnet was developed at CERN in the early 1960s [5]. A transmission-line magnet consists of few to many 'cells' to approximate a broadband coaxial cable (Fig. 9). Ferrite C-cores are sandwiched between HV capacitance plates: plates connected to ground are interleaved between the HV plates. The HV and ground plates form a capacitor to ground (Fig. 9). One C-core, together with its ground and HV capacitance plates, is termed a cell. Each cell conceptually begins and ends in the middle of the HV capacitance plates. The 'fill time' of the magnet (τ_m) is the delay required for the pulse to travel either one way, if the magnet is terminated in its characteristic impedance, or two ways, if the magnet is terminated in a short circuit, through the '*n*' magnet cells.

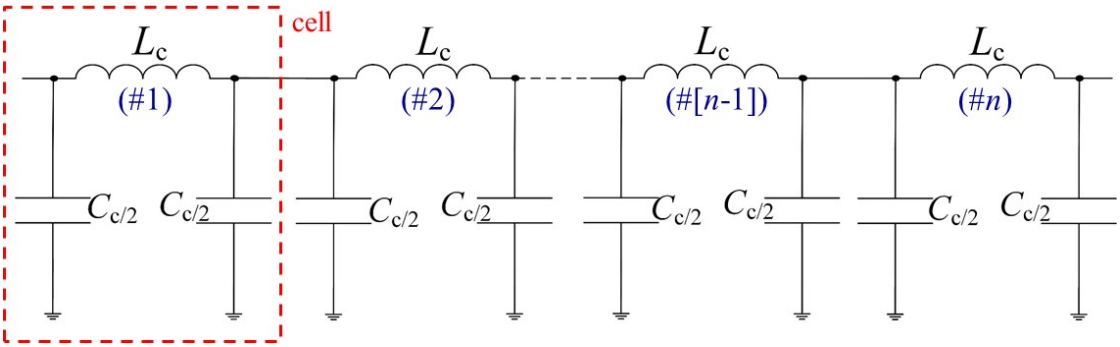


Fig. 9: Simplified equivalent electric circuit of a transmission-line kicker magnet

The characteristic impedance (Z) of the kicker magnet, which is matched to the impedance of the generator to minimize reflections [6, 7] (see section 5.2), is given by Eq. (9):

$$Z = \sqrt{\frac{L_c}{C_c}} \,, \tag{9}$$

In this equation:

L_c is the inductance of a cell of the kicker magnet (H);
C_c is the capacitance of a cell of the kicker magnet (F).

The single-way delay of the kicker magnet is given by Eq. (10):

$$\tau_m = n\sqrt{L_c \cdot C_c} = n\left(\frac{L_c}{Z}\right) = \left(\frac{L_m}{Z}\right) \,, \tag{10}$$

In this equation:

neglecting end effects, $L_m = n(L_c)$,

L_m is the total inductance of the kicker magnet (H).

For a kicker magnet terminated with a matched resistor, field rise time begins with the start of the rising edge of the voltage pulse at the entrance of the kicker magnet and finishes with the end of the rising edge of the pulse at the exit of the magnet. Flux is given by the time integral of the difference between the voltage at the magnet entrance and exit [Eq. (11)]:

$$\Phi = \int \left(V_{in}(t) - V_{out}(t)\right) \mathrm{d}t \,, \tag{11}$$

In this equation:

Φ is flux (V·s);
$V_{in}(t)$ is the voltage at the entrance to the kicker magnet (V);
$V_{out}(t)$ is the voltage at the exit of the kicker magnet (V).

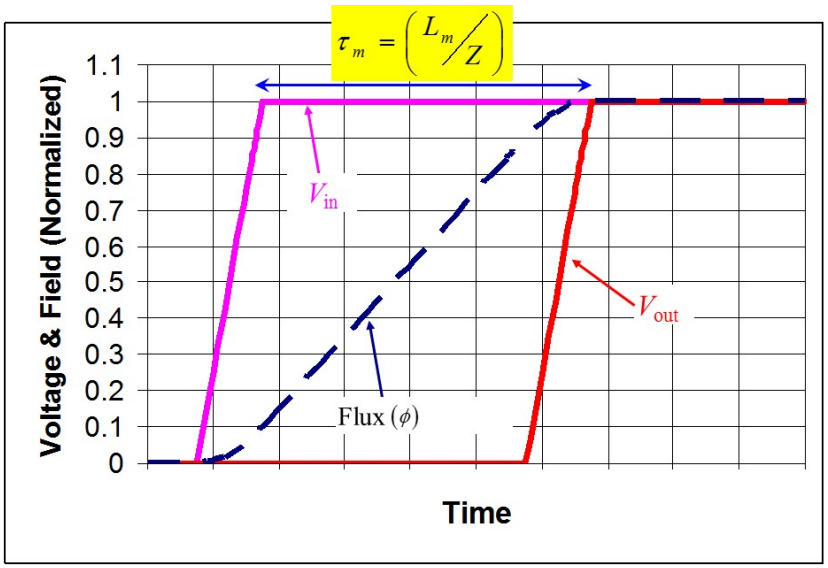


Fig. 10: Flux in an 'ideal' transmission-line kicker magnet

The flux builds up until the end of the voltage rise at the exit of the kicker magnet (Fig. 10): thus, for an ideal transmission-line kicker magnet terminated in its characteristic impedance, flux rise time is given by the sum of the voltage rise time and the single-way delay of the magnet. The single-way delay, in this case, corresponds to the magnet fill time. Therefore it is important that the pulse does not degrade while travelling through the kicker magnet. Hence, the magnet cut-off frequency is a key parameter, especially with field rise times below ~100 ns. Cut-off frequency (f_c) depends on series inductance (L_{cs}) associated with the cell capacitor (C_c) [3]:

$$f_c = \frac{1}{\pi \cdot \sqrt{\left(L_c + 4L_{cs}\right) \cdot C_c}} = \frac{Z}{\pi \cdot \sqrt{\left(L_c + 4L_{cs}\right) \cdot L_c}} \quad . \tag{12}$$

Thus, to achieve a high cut-off frequency, L_{cs} should be kept as low as possible and the cell length small, i.e., low cell inductance and cell capacitance. However, cells cannot be too short because an adequate distance is required between the HV and ground capacitance plates to avoid voltage breakdown. In addition, many very short cells would significantly increase the cost and complexity of the kicker magnet.

Figure 11 shows the results of a low-voltage measurement on each of the HV capacitance plates of a transmission-line kicker magnet. The fast rise time of the input voltage pulse, used for the measurements, contains frequency components above the cut-off frequency of the cells. Thus there is an increase in the rise time of the voltage pulse between the entrance HV capacitance plate and the second HV capacitance plate; in addition there is significant ripple on the pulses.

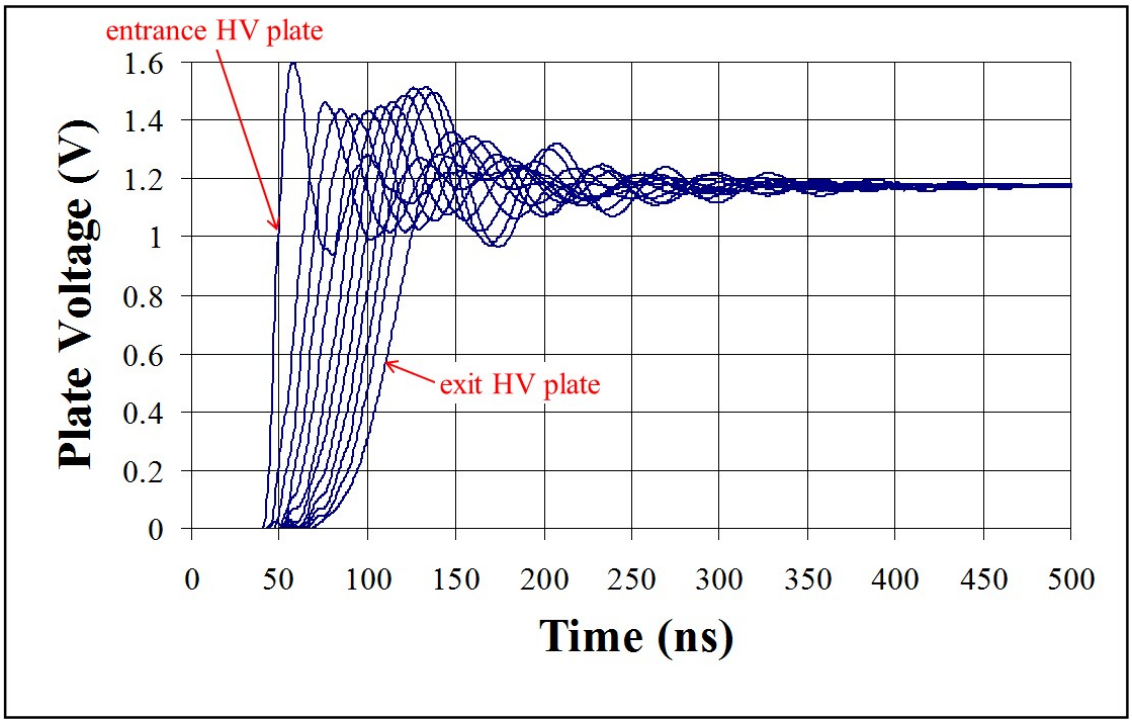


Fig. 11: Low voltage measurement on each of twelve HV capacitance plates of an eleven-cell transmission-line kicker magnet.

The choice of the characteristic impedance, for a transmission-line magnet, depends upon the required field rise/fall time and the available length for the kicker magnet. In general the highest impedance, up to the impedance of commercially available coaxial cable (50 Ω), is used while still respecting the available length for the kicker magnet and chosen level of PFN voltage [3]. The higher the characteristic impedance, the higher the magnet cut-off frequency (Eq. (12)): a high cut-off frequency reduces the field ripple.

A vacuum dielectric transmission-line kicker magnet can require a very large area of metal plates to form the capacitance, especially for a low-impedance system. Since the minimum separation of the capacitor plates is determined by practical considerations, such as voltage breakdown and cost, the capacitance is determined by the area of the plates; however, Eq. 9 shows that the capacitance (area) must increase by a factor of 4 for a decrease in impedance by a factor of 2 (neglecting edge effects). It has been shown that it is feasible to design a 'hybrid' kicker magnet, which is an effective transmission kicker magnet with a fast rise time, using only 3 to 5 large cells [8], but which requires large capacitance values. However, it is impractical for lower impedance magnets or for magnets with larger cell inductances to use vacuum dielectric capacitors. Hence, with the space limitation at many accelerator facilities, the application of higher permittivity dielectric capacitor media is necessary [7, 9].

5.3.4 Machine vacuum

The minimum-sized aperture for a kicker magnet can be achieved if the magnet is in vacuum: if the magnet is outside vacuum, a chamber must be inserted in the kicker aperture, thus increasing the dimensions of the aperture. The minimum value of both H_{ap} and V_{ap} (Fig. 8) are determined by beam parameters. Equation (7) shows that magnet inductance is proportional to H_{ap} and inversely proportional to V_{ap}. However, if V_{ap} is increased, although the magnet inductance is reduced, this is at the expense of increased current to obtain a given flux density (Eq. 6): increased current also requires, for given impedance, increased PFN/PFL voltage and thus increased insulation requirements.

Advantages of putting a transmission-line kicker magnet in the machine vacuum are the following.

- Aperture dimensions are minimized, therefore the number of magnets and/or voltage and current are minimized for a given $\int B.\mathrm{d}l$ and rise time.
- Machine vacuum is a reliable dielectric (70 kV/cm is OK [3]) and generally 'recovers' after a flashover, whereas a solid dielectric, outside vacuum, may not recover.

Disadvantages of putting a transmission-line kicker magnet in the machine vacuum are the following.

- The kicker magnet is costly and time consuming to construct (all parts must be appropriately cleaned and handled, a vacuum tank is required, and suitable pumps are needed. If the kicker magnet must be baked out, to achieve the required vacuum level, the design may need to allow for thermal expansion during bake out, etc.).
- In the event of failure of something inside the vacuum tank, the vacuum must be broken to repair the fault. Afterwards, pumping, bake out (if appropriate), and HV conditioning are required—a time-consuming process.

The beam-coupling impedance may be an issue irrespective of whether the magnet is in vacuum. Hence a beam screen (see section 5.5) in the kicker aperture, may be required: this will require an increase in aperture dimensions in any case.

5.3.5 Kicker magnet aperture

Normally a magnetic circuit is used which contains magnetic material [3, 5]: without magnetic material the effective value of V_{ap} (Fig. 8) is greatly increased, therefore requiring more current to achieve the required field (Eq. 6). In addition, magnetic material improves field uniformity. Nickel-zinc (NiZn) ferrite, with initial $\mu_{\mathrm{r}} \approx 1000$, is typically used [1, 3]. If the ferrite is in machine vacuum (section 5.3.4) the out-gassing properties must be acceptable: in addition, the density of the ferrite must be such that its water absorbent characteristics are not too high and no undesirable lubricants must be used during the grinding process [1]. At CERN, 8C11 ferrite (Ferroxcube) and CMD5005 ferrite (National Magnetics Group, Inc.) are currently used, although other types of ferrite are under consideration. The 8C11 and CMD5005 ferrites have the following properties:

- field rise can track current rise to within ~1 ns [3];
- low remnant field;
- low out-gassing rate, after bake out.

Figure 12 shows a full-aperture C-core magnet (Left Hand Side (LHS)) and a window-frame magnet ((Right Hand Side (RHS)). The C-core magnet shown in Fig. 8 has its aperture closed by the return conductor. The C-core magnet on the LHS of Fig. 12 has the return conductor behind the yoke: this is for beam gymnastic reasons and has the effect of increasing the effective width ($H_{\mathrm{ap(eff)}}$) of the magnet aperture, of given dimensions, from H_{ap} to a value given approximately by Eq. (13) [10, 11]:

$$H_{ap(eff)} \approx \left(H_{ap} + V_{ap}/2\right) . \tag{13}$$

Similarly, the inductance of a cell of the full-aperture kicker magnet is increased by a factor of $\left(H_{ap(eff)}/H_{ap}\right)$ with respect to the return conductor closing the aperture of the kicker magnet.

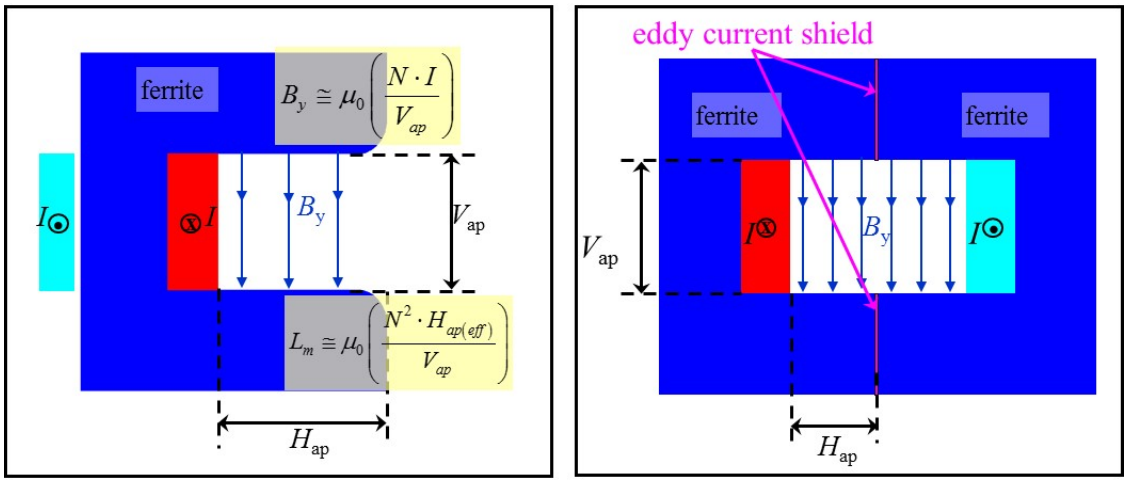


Fig. 12: Full-aperture C-core magnet (LHS) and window frame magnet (RHS)

To reduce magnet fill time by a factor of two, FNAL and KEK use a window-frame topology (RHS of Fig. 12): it can be considered as two symmetrical C-magnets energized independently. However, the window-frame magnet requires two generators to achieve the reduced magnet fill time [5]. It is necessary to include eddy current 'shields' between the two ferrite C-cores to reduce beam-coupling impedance (see section 5.5).

C-core kicker magnets are generally used at CERN.

5.3.6 Kicker magnet termination

When space along the beam line is at a premium, a short-circuit termination has the advantage over a matched resistive termination of doubling current and hence field (for a given system impedance) [3]: in addition, a short-circuit termination reduces the time during which the kicker magnet is exposed to high voltage. However, disadvantages include:

- for a given magnet length and impedance, fill time of the kicker magnet is doubled: to establish full field in the magnet, the current/voltage pulse must propagate from the input to the output of the magnet and then reflect back to the input;
- the kicker magnet experiences voltage of both polarities: one polarity during field rise and the opposite polarity during field fall (the incident voltage wavefront is inverted at the short-circuit termination [reflection coefficient is −1]);
- if the DS is used to control pulse length it must be bidirectional (unidirectional DS, acting as an inverse diode, is suitable for a fixed length pulse);
- beam-coupling impedance can be influenced (see section 5.5).

5.4 Kicker magnet design tools

Circuit simulation and finite element codes greatly assist the goal of obtaining high performance kicker systems. Simulation of circuits which include almost all known parasitic elements and non-linearities is now possible [5].

Two- and three-dimensional (3D) finite element codes now include AC and transient analysis with eddy currents. These tools are used for kicker magnet design to predict magnetic field, cell inductance, electric field [13–15], and capacitance. In order to obtain realistic predictions for inductance and magnetic field distribution, in the aperture of the kicker magnet, it is necessary that the skin effect and proximity effect, in the HV and 'ground' (return) conductors (Fig. 13), be properly accounted for: this requires an AC or transient analysis to be carried out. Codes such as *Opera2D* and *Opera3D* [16]

are also used to study the shape of the ferrite and conductors and thus optimize the field homogeneity [14, 15]; the current distribution is calculated by the code. The LHS of Fig. 13 shows a transmission-line kicker magnet used to deflect the beam horizontally. The magnet was modelled using *Opera2D* to predict the magnetic and electric field distribution. The predictions were post-processed and the total deflection calculated; the RHS of Fig. 13 shows a plot of deflection uniformity for the optimized geometry of ferrite and return conductor [14, 15].

The penetration of the pulsed field through a beam screen (section 5.5), in a magnet aperture, can also be accurately calculated and the frequency dependence of magnet inductance, due to eddy currents in the screen, predicted [17]. An equivalent circuit can then be fitted to the resulting predictions to account for the frequency dependence in an analogue circuit simulation [18].

An AC analysis can also be used to predict the frequency dependence of both inductance and resistance, i.e., skin effect and proximity effect, due to eddy currents in PFN coils [18, 19] (see section 5.8). An equivalent circuit can be fitted to the resulting predictions to account for the frequency dependence in an analogue circuit simulation. The electric circuit models can also include displacement current during turn on of a multi-gap thyratron (section 5.9.1).

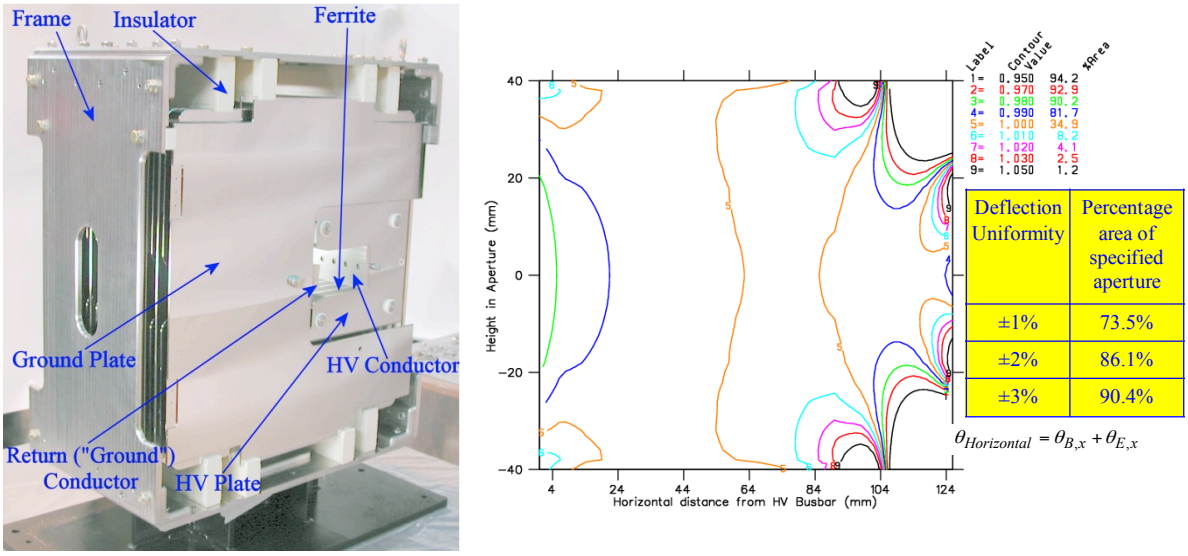


Fig. 13: Transmission-line kicker magnet and corresponding deflection uniformity plot

In addition, capacitance to ground of a HV plate, in a kicker magnet, is influenced by insulators and nearby ground planes such as the:

- ground plate;
- magnet frame;
- return conductor.

Boundary element software such as *Coulomb*, a 3D code from Integrated Engineering Software [20], can be used to accurately predict the capacitance of a cell of a kicker magnet. The predictions resulting from *Coulomb* have been carefully checked against measurements and found to be in good agreement [4].

Thermal analysis packages such as *ANSYS* [21] are used to assess transient and steady-state temperatures of, for example, the yoke of a kicker magnet (section 5.6). *ANSYS* can also be used to assess mechanical stress on magnet components during rapid heating or cooling.

Powerful simulations tools such as *CST Particle Studio* [22], which can model frequency-dependent permittivity and permeability (e.g., of ferrite), can be used to predict the beam-coupling impedance for complex structures such as kicker magnets, and determine beam-induced power

deposition in volumes of the structure (section 5.5). The high frequency structure simulator (*HFSS*) [21] can be used for beam-coupling impedance and field homogeneity simulations (section 5.10).

5.5 **Beam-coupling impedance**

High-intensity accelerators are very sensitive to longitudinal and transverse beam-coupling impedance. Kicker magnets, with their ferrite yoke, can result in considerable beam-coupling impedance [23−27], and can be one of the dominant sources of impedance [28], which can limit the intensity reach of an accelerator. The real component of the longitudinal beam-coupling impedance, together with the beam spectrum, results in beam induced power deposition in the ferrite yoke [23−27, 29−31], sometimes heating the yoke beyond the Curie temperature. If the yoke exceeds its Curie temperature, it will temporarily lose its magnetic properties: as a result, beam could be miskicked. Hence, it would be necessary to allow the yoke to cool below its Curie temperature before using the system to inject or extract beam: because of long thermal time constants this could require several hours [32]. In extreme cases beam-induced power can affect the generator as well [33].

There are formulae for estimating the beam-coupling impedances of kicker magnets e.g., Refs. [34−37]. Nevertheless, *CST Particle Studio* [22] is an important software tool for predicting beam-coupling impedance, especially for complex geometries and dispersive materials, understanding the source of both broadband impedance and impedance resonances, and optimizing a design [25−27]. Beam-coupling impedance measurements can be made using a coaxial wire method, which is also an invaluable tool [12, 23]. Standard wire measurement techniques are described in [38]. Wire measurements rely on the fact that an ultra-relativistic beam has a very similar electromagnetic field distribution to that of a transverse electromagnetic (TEM) line [23]. Generally speaking, the wire diameter should be as small as possible to get a high line impedance [23].

Some beam-coupling impedance resonances are attributable to the image path of the beam not being continuous, for example, a gap between the end of the kicker magnet and the vacuum tank. In this case, these geometrical resonances can be eliminated by inserting transition pieces inside the tank to electrically connect the tank flange to the kicker magnet at each end [29, 42−44]. When several kicker magnets are installed in a common tank, the transition between magnets is carried out in a similar way [5].

Longitudinal beam-coupling impedance of the ferrite yoke of a kicker magnet is reduced by providing a beam screen in the magnet aperture [12, 17, 23−27, 30, 39−41]. The metallic parts of the screen conducts the beam image current and hence the electromagnetic fields penetrating the ferrite are reduced in magnitude but, at the same time, the screen must not significantly attenuate the pulsed field. An alumina chamber is often placed in the aperture of a kicker magnet. A thin metallic coating is frequently applied to the inner surface of the alumina chamber to both screen the kicker magnet yoke from the beam [e.g., 45−48], and also to prevent build-up of static charge on the chamber [47]. It is important that the metallization does not unduly degrade field rise/fall times or the flat top of the field pulse in the kicker magnets. This is of particular concern for kicker magnets that have a fast field rise and/or fall time. The impact of the metallization on the transient field response can be studied using finite element method (FEM) simulations [48]. For lumped element kicker magnets, analytic formulae also allow a rapid evaluation of the influence of the metallic coating upon field rise and fall. The time constant for the metallized coating, obtained from these formulae, can then be incorporated in a *PSpice* model of the kicker system so that arbitrary excitation waveforms can be considered and the system optimized for the required magnetic field rise and fall times [48].

The LHS of Fig. 14 shows longitudinal beam-coupling impedance reduction techniques for the MKE kickers [23], used for extraction from the CERN super proton synchrotron (SPS) ring. Two sets of silver fingers are printed directly on the ferrite, by serigraphy: each set of fingers is connected to a high-voltage plate. The serigraphy fingers cannot be continuous between the two high-voltage plates, as this would create eddy-current loops and hence can significantly increase field rise time. The spacing

between the two sets of fingers, for the MKE, is 20 mm: this spacing is chosen for high-voltage reasons. The length of the MKE ferrite block is 230 mm. The serigraphy does not reduce the available aperture.

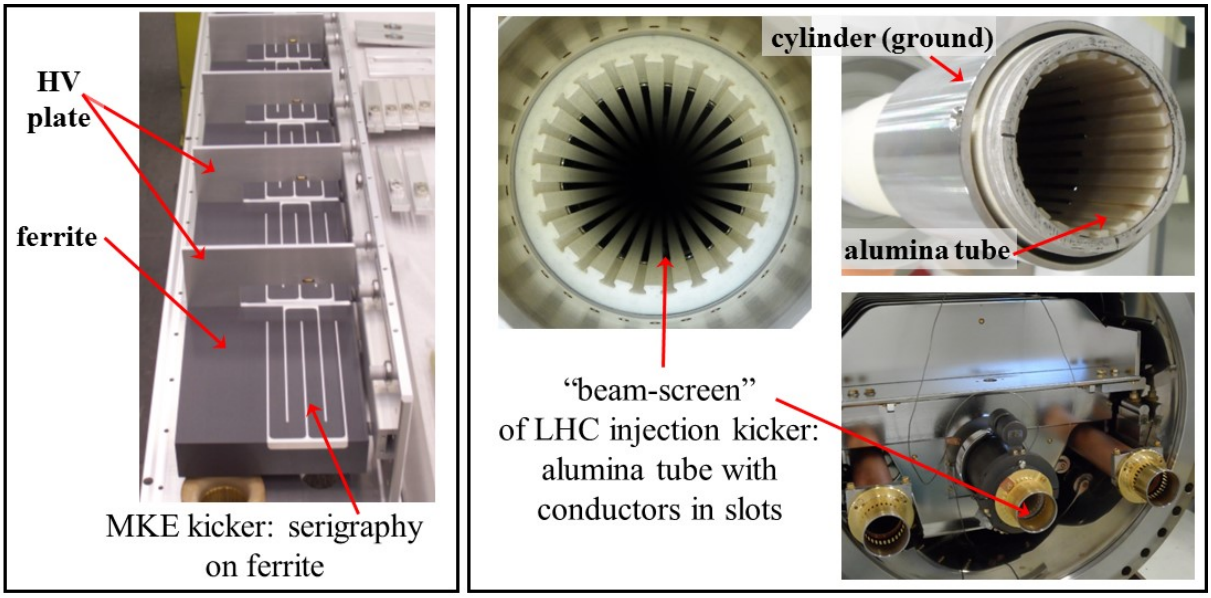


Fig. 14: Beam-coupling impedance reduction techniques for the MKE (LHS) and MKI (RHS) kickers

Figure 15 shows longitudinal beam-coupling impedance for several systems; unless mentioned otherwise, the impedance is derived from measurements: the serigraphy results in a strong reduction of the real part of the longitudinal beam-coupling impedance over a broad frequency range. At the same time the serigraphy introduces a quarter-wavelength resonance [25]: the resonant frequency is dependent upon properties of the ferrite (permittivity and permeability [which are frequency dependent]), and the length of the serigraphy fingers. This resonance is clearly visible in Fig. 15, at approximately 50 MHz, in the measured impedance of the serigraphed MKE-L magnet: the length of the serigraphy fingers, for this case, is 200 mm. Hence, to limit beam-induced power deposition, the length of the serigraphy fingers should be optimized to avoid a high-impedance resonance close to the frequency of a beam harmonic [25]. Furthermore, care must be taken to ensure that an impedance resonance does not result in beam instabilities.

The thickness of the serigraphy must be sufficient to achieve the objective of adequately reducing the longitudinal beam-coupling impedance but thin enough such that field rise time is not significantly increased [49]. Recent *CST* simulations show that serigraphed fingers would not be effective at reducing beam-induced power deposition when applied to individual cells of a kicker magnet that has relatively short cells: for example, the CERN SPS injection kickers (MKP) [27] the MKP-L ferrite length is only 26 mm per cell: approximately 1/9th of the MKE ferrite length per cell. With short cell length, to achieve sufficient coupling between the two sets of fingers, the serigraphy must extend over a significant portion of the length of the yoke. Thus, the serigraphy must cross the MKP HV plates, which sandwich the ferrite cores. Hence, insulation is required between the magnet and the serigraphy: for the MKP it is expected that the serigraphy would be applied to alumina plates [27, 50].

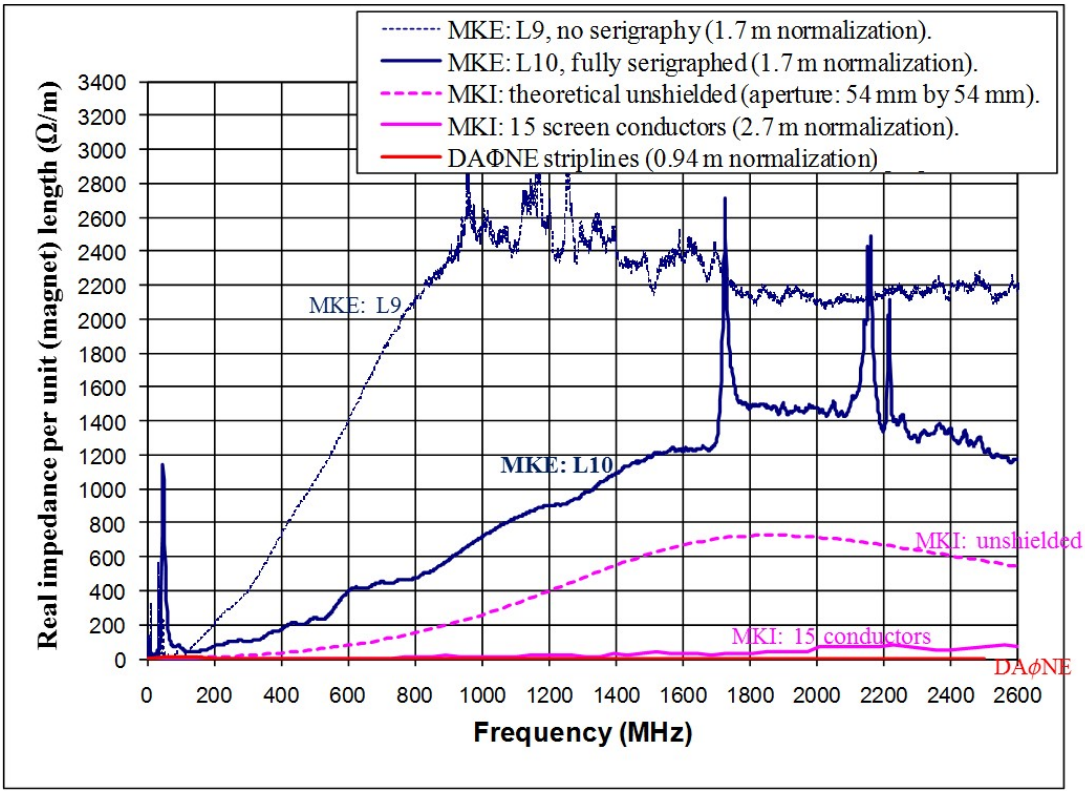


Fig. 15: Real component of longitudinal beam-coupling impedance for several systems, terminated in their characteristic impedance.

Longitudinal beam-coupling impedance measurements on the SPS extraction kicker magnets show that the high impedance resonance at ~50 MHz (Fig. 15) is only present with serigraphy [12]. Detailed measurements show that this resonance is also influenced by the external circuit on the output of the magnet (resistor termination or short circuit): for an SPS MKE, with the pulse input end open circuit, terminating the kicker magnet output in a short circuit, rather than its characteristic impedance, results in an increase in the expected beam induced power deposition by almost 10%.

Longitudinal beam-coupling impedance measurements, on a kicker magnet terminated in a short circuit, with coaxial cables on its input, have been carried out [44]: when the MS thyratron is in the off state (Fig. 4), the transmission cables between the MS thyratron and the input to the kicker magnet are effectively terminated in an open circuit at the MS end. The measurements show that below ~45 MHz impedance resonances were produced by electrical resonant modes of the kicker system: the resonances correspond to that of a half-wavelength resonator, i.e., with one end open circuit and the other short circuit. Above 45 MHz there is not strong coupling to the electrical circuit of the kicker system hence, above 45 MHz, the longitudinal beam-coupling impedance spectrum did not depend strongly upon the electrical termination on the output of the magnet [44]. Beam based measurements of kicker impedances also show that a kicker magnet with coaxial cables on its input and terminated in a short circuit exhibit impedance resonances, many of which are supressed by a resistive termination [51]. Similarly, measurements carried out with the output of the magnet terminated in its characteristic impedance show that the presence or absence of coaxial cable on the input of the magnet, terminated remotely in an open circuit, has a small, but measurable, effect upon beam-coupling impedance.

A saturating inductor can be placed at the electrical input to the kicker magnet to reduce the effect of displacement current, during turn on of a multi-gap thyratron, upon field rise time (section 5.9.1). Longitudinal beam-coupling impedance measurements have been carried out, on a kicker magnet terminated in a short circuit, to assess the influence of the saturating inductor: resonances due to the input transmission cables are effectively eliminated by the presence of the saturating inductor [44]. In

addition, a series resistor−capacitor (known as a speed-up network), connected from the kicker magnet side of the saturating reactor to ground, can be used to reduce the magnitude of the remaining impedance resonances [44].

The RHS of Fig. 14 shows longitudinal beam-coupling impedance reduction techniques implemented for the MKI kickers, used for injection into the CERN large hadron collider (LHC) rings. A 3 m long alumina tube is manufactured with 24 slots on its inside diameter [17]; however, this tube results in approximately 15 mm loss of aperture, and thus the magnet aperture dimensions must be increased to accommodate it, therefore reducing the magnetic field strength for a given current (see section 4). Carefully radiused conductors are inserted into the slots: the alumina tube acts as a support for these conductors and in addition provides high-voltage insulation. At one end of the alumina tube there is capacitive coupling between the beam screen conductors and an outside metallization, which is connected to a metallic cylinder that is at ground potential (Fig. 14). The capacitive coupling provides a continuous path for the beam image current while preventing eddy-current loops: eddy-current loops can significantly increase field rise time. Three-dimensional electromagnetic simulations have been carried out to optimize the geometry of the beam screen. Nevertheless the design of the upstream end of the beam screen (Fig. 17), results in an open-ended half-wavelength resonator; this is due to the region where the screen conductors overlap with the outer metallization [52].

Each MKI kicker magnet has toroidal ferrite rings mounted around each end of the alumina tube, outside of the aperture of the kicker magnet (Fig. 16). The original purpose of the rings mounted on the end where the screen conductors are connected directly to the beam pipe was to damp low-frequency resonances [53].

During long shutdown 1 (LS1) of the LHC, the beam screen of the MKI kicker magnets was upgraded to have 24, rather than 15, screen conductors [53]. *CST* simulations also show that the longitudinal distribution of volume losses of the post-LS1 MKI kicker magnets is non-uniform: the power deposition in the yoke is predominantly at the capacitively coupled end of the magnet, with losses being relatively low in the down-stream half of the yoke [52]. The design of the upstream end of the beam screen and ferrite rings can be optimized to reduce the beam-induced power deposition in the ferrite yoke, at the expense of increased power deposition in the upstream ferrite rings: nevertheless, redistributing power deposition in this way can be advantageous (see section 5.6).

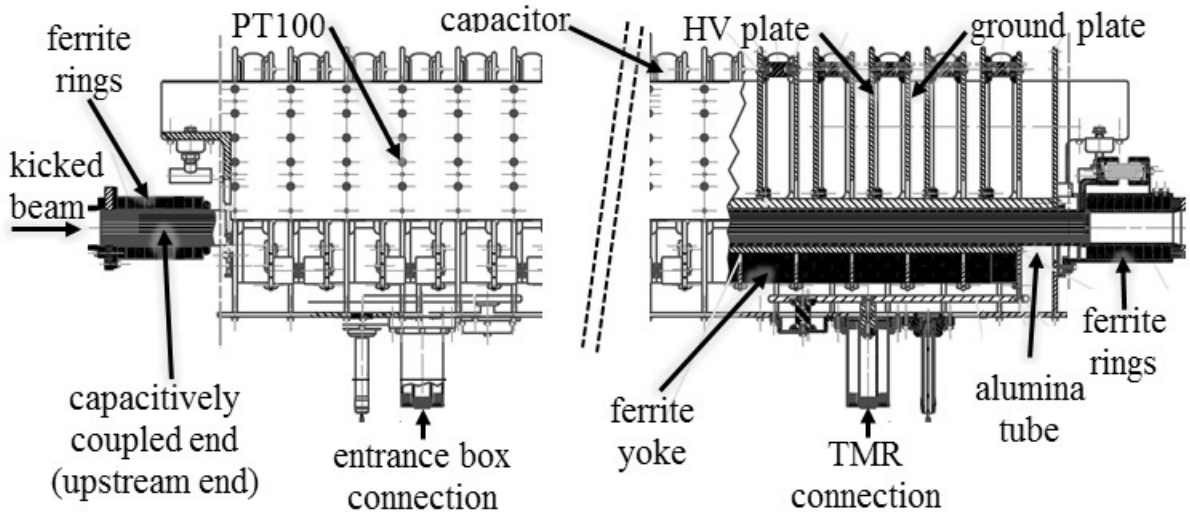


Fig. 16: MKI kicker magnet

Since the MKE and MKI beam screens (serigraphy for the MKE kicker magnets and conductors for the MKI kicker magnets) are in the beam aperture, a fast changing magnetic field (during field rise and fall) induces a high voltage on the beam screen. The magnitude of voltage induced is dependent

upon the position of the conductor in the aperture of the kicker magnet. For the MKI kicker magnets, extensive 3D electromagnetic simulations have been carried out, using the code TOSCA [16], to study electric fields on the surface of the alumina tube. The predictions, in conjunction with operational experience, are used to determine a safe upper operating limit for the electric field [54, 55]. Based on the results of these simulations, the electric field gradient associated with the conductors at highest voltage is reduced by optimizing the length of the screen conductors, such that those at lower voltage 'shade' those at higher voltage [40]: hence, as shown in Fig. 17, the screen conductors towards the HV busbar (bottom) of the alumina tube have the shortest length. In addition, a vacuum gap was introduced between the outside of the alumina tube and the metallic cylinder (Fig. 17), to significantly reduce the electric field at this end of the screen conductors [54]. As a result of the high permittivity of the alumina tube, this gap reduces the capacitance of the capacitively coupled end of the beam screen and hence causes the half-wavelength resonances, resulting from the overlap of the screen conductors with the metallic cylinder, to increase in frequency [56].

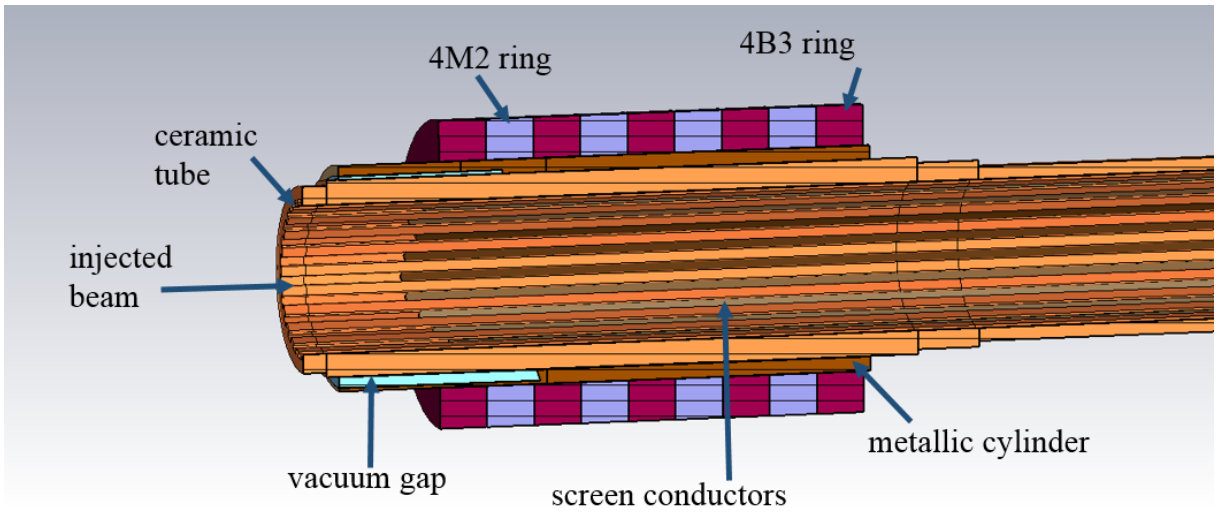


Fig. 17: Upstream end of beam screen of MKI kicker magnet

In summary, the design of a beam screen must be optimized for achieving the required field rise/fall time, good high-voltage behaviour, and reducing the real part of the longitudinal beam-coupling impedance over a broad frequency range. The real component of longitudinal beam-coupling impedance can be significantly reduced by:

- incorporating a metallized alumina chamber;
- silver serigraphy of ferrites (painted fingers): there is negligible loss of aperture, but a high impedance resonance is introduced by the serigraphy;
- include beam screen conductors within the kicker magnet aperture: however this can result in approximately 15 mm loss of aperture and is thus usually not applicable as a retrofit to an existing kicker magnet;
- using striplines instead of a ferrite loaded magnet: Fig. 15 shows that the DAΦNE striplines had a very low longitudinal impedance [57]. However using striplines is not feasible for obtaining a large deflection angle in a limited length.

Even with a well-designed beam screen, with high bunch intensity and short bunch lengths, integrated over many hours of a good physics fill, the beam-coupling impedance of the kicker magnet yoke can lead to significant beam induced heating of the yoke: thus, enhanced cooling may be necessary.

5.6 Yoke temperature

5.6.1 Cooling

To limit longitudinal beam-coupling impedance, while allowing a fast magnetic field rise time and good high-voltage behaviour, a beam screen can be included in the aperture of the kicker magnet (section 5.5). Nevertheless, the beam screen may not sufficiently limit beam-induced heating of the yoke. If the yoke exceeds its Curie temperature it temporarily loses its magnetic properties, preventing further injection/extraction until the yoke cools down: because of a long thermal time constant this could require several hours [32]. Kicker magnets are frequently installed in vacuum: convection is negligible and, depending upon the emissivity of the inside of the vacuum tank, thermal radiation is limited [58, 59].

To limit the temperature rise of the SPS extraction kicker magnets, due to beam-induced power deposition, the ferrite yoke is indirectly cooled using water [29, 60]. The cooling is implemented using water-cooled aluminium nitride plates, which are in thermal contact with the ferrites: aluminium nitride is a good thermal conductor but is also electrically insulating, which is necessary as the ferrite is at pulsed high voltage. The cooling allows the beam-induced power deposition to be doubled, with respect to a non-cooled system, before the Curie temperature of the ferrite yoke is reached [29].

For the MKIs, as a result of the low thermal emissivity of the internal surface of the vacuum tanks [58], cooling of the ferrite yoke by thermal radiation is limited. Hence, High Luminosity LHC (HL-LHC) type beam would result in temperatures of the upstream ferrite yokes and upstream ferrite rings well above their Curie temperatures [61], despite the presence of an effective beam screen. Thus extensive research has been carried out to find a high emissivity coating, for the internal surface of the vacuum tank, which is compatible with ultra-high vacuum, thermal bake outs, and high voltage, and does not peel or flake. Several surface finishes have been considered: laser treatment, carbon coating, multilayer optical coatings deposited by magnetron sputtering, and coatings deposited by thermal spray (e.g., plasma spray and flame spray). However, although increasing the emissivity reduces both the maximum temperature and the thermal time constant, so that the yokes cool down more rapidly, the temperatures are not reduced sufficiently for HL-LHC type beam [61]. Hence, other cooling schemes have been studied.

Aluminium nitride plates in good thermal contact with the ferrite yoke have also been studied: predictions show that the yoke would be below its Curie temperature. However, *CST* simulations show that the ferrite rings, at the upstream end of the kicker magnet, experience a significant proportion of the total beam-induced power deposition, for a structure that has relatively little volume. Hence additional cooling is required for the upstream ferrite rings as their temperatures are otherwise predicted to exceed their Curie point: these rings already have a relatively high Curie temperature of 200°C and 250°C for 4M2 and 4B3 ferrite, respectively [61].

Recent beam-coupling impedance studies show that a change in the design of the upstream end of the magnet can reduce significantly the beam-induced power deposition in the ferrite yokes by moving it into the upstream ferrite rings (see section 5.5). Detailed *ANSYS* simulations show that, with this new proposed design, for HL-LHC type beam, the high temperature of the upstream yokes is caused mainly by the heat conducted and radiated from the ferrite rings. Thus, effective cooling of the ferrite rings significantly reduces the temperature of the upstream ferrite yokes too [61]: direct cooling of the ferrite rings is possible as they are not at pulsed high voltage. Nevertheless, the cooling system will be in a pulsed high voltage and ultra-high vacuum ($<10^{-9}$ mbar) environment: hence, one must ensure that the risk of liquid leakage into machine vacuum is negligible.

5.6.2 Measurement of yoke temperature

PT100 temperature sensors are used in several of the kicker magnets installed in the CERN accelerator rings: these are generally shielded to avoid beam induced electromagnetic interference. The

PT100 sensors cannot be in direct contact with the magnet yoke, as the yoke is at pulsed high voltage [32, 60]: hence, during operation, the temperature indicated by a PT100 may be significantly less than the actual yoke temperature. In addition, for the MKIs, the temperature distribution throughout the yoke is highly non-linear [52, 61]. Thermal studies have been carried out to correlate temperatures measured by the PT100s with the yoke temperatures. However, there is always an uncertainty because this correlation is also dependant on the heating and cooling rates, which varies during LHC operation.

Either 8C11 or CMD5005 ferrite is presently used for the MKI yoke: the permeability rapidly reduces for temperatures above the Curie temperature (~125°C). In addition to reducing the strength of the magnetic field, a decrease in the inductance of an MKI results in a decrease in the propagation delay of the kicker magnet and the rise time of the current in the TMR. By pulsing the MKI magnets without beam in the accelerator, the propagation delay of a magnet and the rise time of the current in the TMR can be measured: this is carried out immediately after beam in the LHC is dumped—when, after a long duration physics run with high intensity beam, the yoke of the MKIs is still relatively warm. *PSpice* simulations, where the inductance or capacitance of every cell is modelled as reduced by a given percentage, show that the percentage reduction of magnet delay is almost five times greater than the percentage decrease in TMR current rise time: hence, the magnet delay is the preferred of the two diagnostics. In general, a plot of magnet propagation delay versus measured temperature, or TMR current rise time versus measured temperature, is relatively linear: however once the ferrite starts to lose its permeability, there is a relatively rapid increase in the slope of the curves. Hence, these electrical measurements allow one to deduce whether or not the ferrite yoke is approaching/above the Curie temperature [62].

5.7 Electron cloud

Significant pressure rise, due to electron cloud, occurs in and nearby the MKIs (Fig. 18): the predominant gas desorbed from surfaces is H_2. Conditioning of surfaces reduces electron cloud, and thus dynamic pressure rise, but further conditioning is often required when beam parameters (e.g., bunch spacing, length, and intensity) are pushed.

Voltage is induced on the MKI screen conductors during field rise (to 30 kV) and fall (to −17 kV) [54]. High pressure, at the upstream (capacitively coupled) end, can result in breakdown/flashover. Hence, it is desirable to minimize dynamic pressure in and around the MKIs. The alumina tube of each MKI has a secondary electron yield (SEY) of ~10 when first installed and, together with metallic surfaces facing the beam, requires conditioning with beam [63].

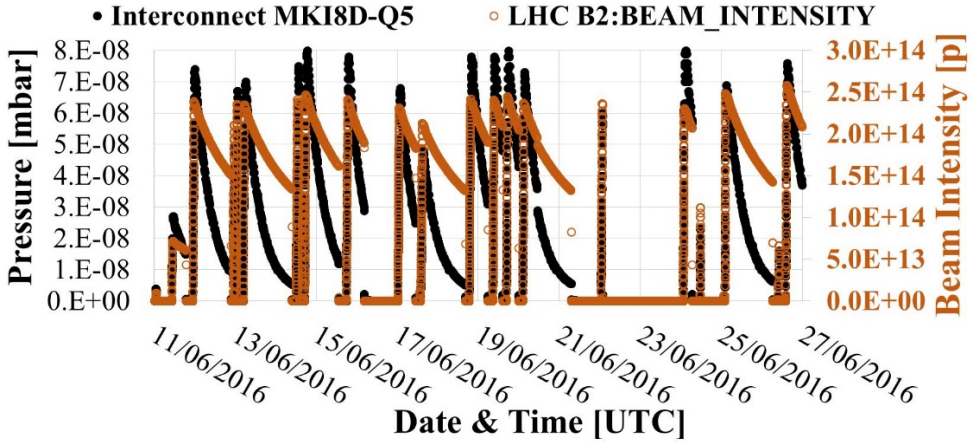


Fig. 18: Example of dynamic pressure in an MKI8D interconnect and beam intensity during June 2016

In order to mitigate electron multipacting, a coating of Cr_2O_3 is applied, by Polyteknik (Denmark), to the inside of the alumina tube by magnetron sputtering. Measurements show that naked, high-purity, alumina has a maximum SEY (δ_{max}) of ~9. A Cr_2O_3 coating, applied by magnetron sputtering, reduces

δ_{max} to approximately 2.3: bombarding the surface with electrons further reduces δ_{max} to less than 1.4 (Fig. 19) [63]. In addition, a set of samples sputtered with 50 nm of Cr_2O_3 was installed in a setup in the CERN SPS, during the winter stop of 2016−2017, for tests with beam. These tests confirmed that relatively rapid conditioning of the Cr_2O_3 occurred in the presence of beam.

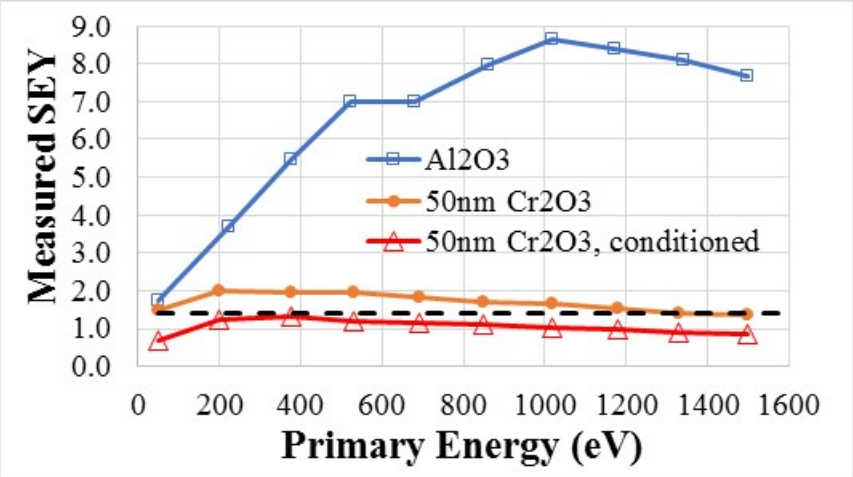


Fig. 19: Measured SEY of high purity alumina and alumina with a 50 nm Cr_2O_3 coating: bombarding Cr_2O_3 with electrons reduces δ_{max} to below 1.4.

Cr_2O_3 was originally studied, for coating of the inside of an MKI alumina tube, based on papers published by Sudarshan and Cross [64] and by Shioiri [65], which reported a ~30% increase in voltage of surface flashover in vacuum, in comparison with uncoated alumina. Reference [65] concludes that the effectiveness of Cr_2O_3, at increasing surface flashover in vacuum, is due to:

- a reduction in secondary electron avalanche;
- a reduction of surface charging due to decreased surface resistance (10^{10}–10^{11} Ω·cm).

5.8 Pulse-forming line/network

The simplest configuration for a pulse-forming circuit is a PFL (transmission line/coaxial cable). The PFL gives fast and ripple-free pulses, but low attenuation is essential, especially with longer pulses, to keep droop and 'cable tail' within specification [3]. Attenuation is adversely affected by the use of semiconductor layers to improve voltage rating [3]. Hence, for PFL voltages above 50 kV, SF6 pressurized polyethylene (PE) tape cables have been historically used: however the availability of SF6 gas filled coaxial PFL cables is a concern for both existing and future kicker systems. PFL becomes costly, bulky, and the droop becomes significant (≥ 1%) for pulses exceeding approximately 3 μs duration.

Where low droop and long pulses are required, a PFN is used: a PFN is an artificial coaxial cable made of lumped elements. The SPS extraction PFNs at CERN [66], which are approximately 40 years old, have approximately 20 cells, connected in series in a single line (Fig. 20, RHS), which are individually 'adjustable': these PFNs have 'corners' (Fig. 20, LHS), therefore mutual inductance between cell inductances is not well defined and can result in ripple on the pulse. Adjusting the pulse flat-top is difficult and time consuming.

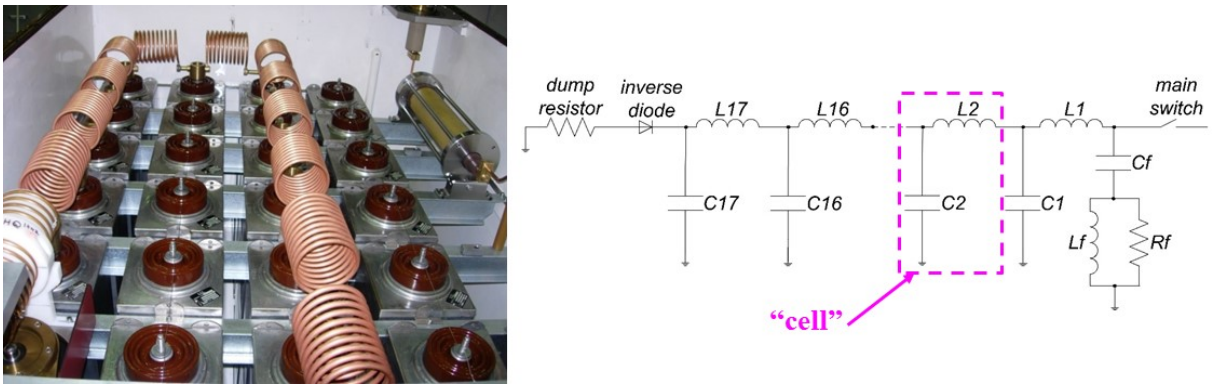


Fig. 20: The SPS extraction PFNs at CERN have approximately 20 cells, which are individually 'adjustable'

The preservation of the transverse emittance of the proton beam at injection into the LHC is crucial for luminosity performance. The transfer and injection process is important in this respect, and injection offsets are a well-known source of error [67]. To limit the beam emittance blow up due to injection oscillations, the reflections and the flat-top ripple of the field pulse must be lower than ±0.5%, a very demanding requirement [9].

Each of the two LHC injection kicker systems has four 5 Ω PFNs [68]. Each PFN consists of two lumped element delay lines, each of 10 Ω, connected in parallel [68]. Each 10 Ω line consists of 26 central cells plus two end cells. A cell consists of a series inductor, a damping resistor connected in parallel, and a capacitor connected to ground (Fig. 21, LHS). The inductors are part of a single continuous coil, 4.356 m long, with 198 turns and a pitch of 22 mm [68]. The central cell inductors are made of seven turns each. The nominal MS and DS end-cell inductors have nine turns and five turns respectively, but are built with one extra turn to allow some adjustment to compensate for end effects. The 26 central cells of the coils are not adjustable and therefore are defined with high precision: the coil conductor is a copper tube wound on a rigid fibreglass coil former. Both delay lines are mounted in a rectangular tank (Fig. 21, RHS), with mild steel walls, that is filled with insulating silicone fluid. Each line is surrounded by a 3 mm thick, Ω-shaped, aluminium shield, which has an inner radius of 140 mm. Two thyratron switches, a MS and a DS, are connected to the PFN (see also Fig. 4).

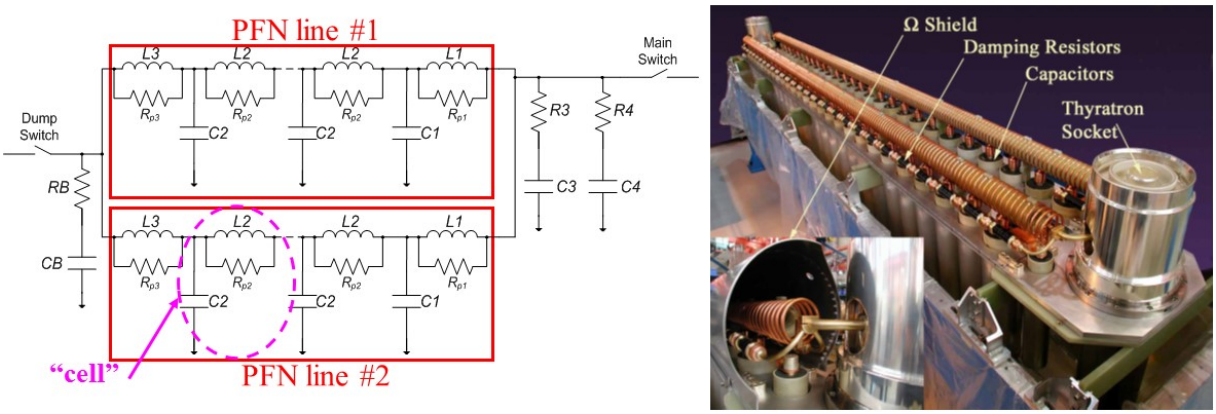


Fig. 21: Each LHC injection PFN consists of two parallel 10 Ω lines with precision wound coils

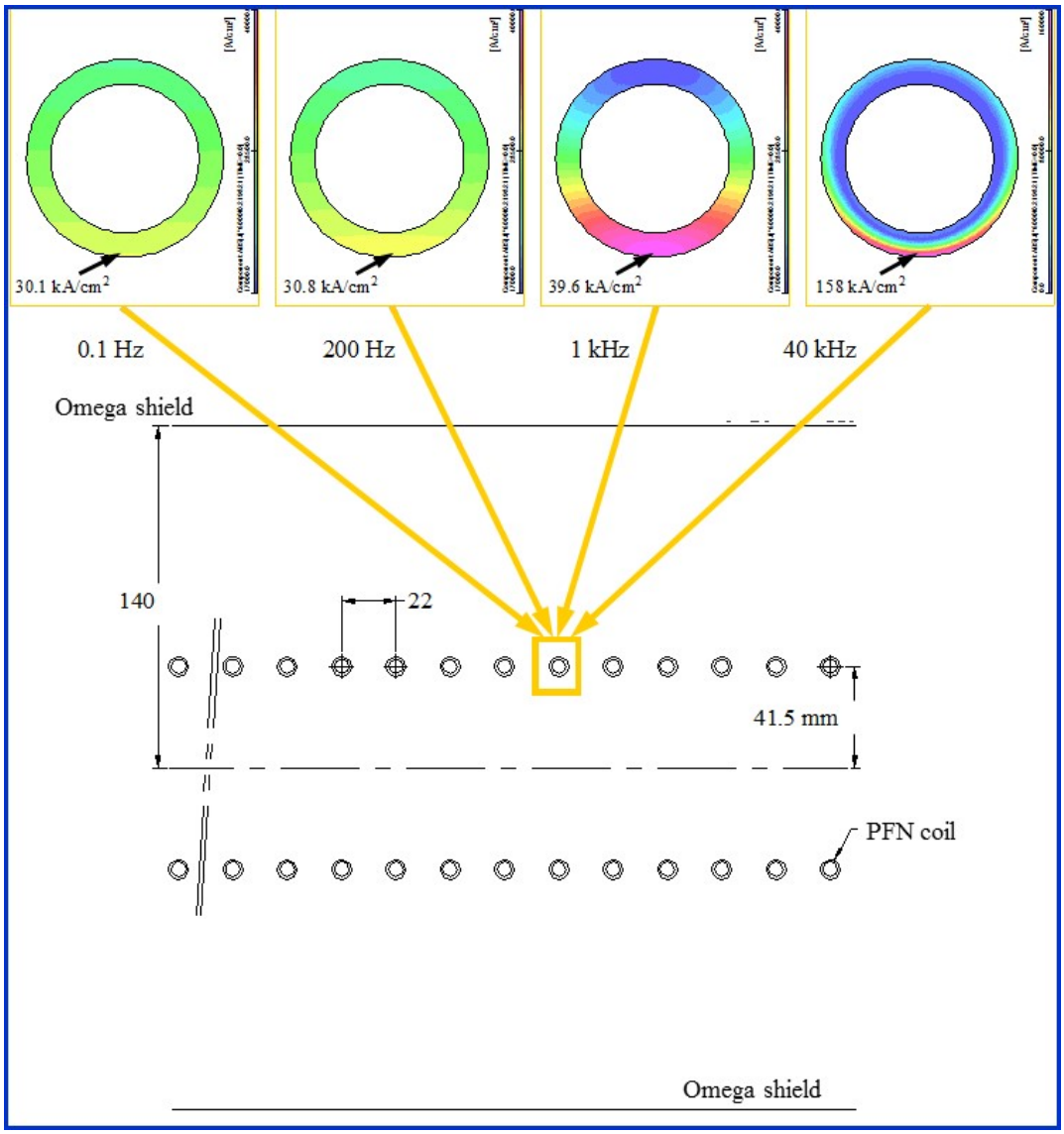


Fig. 22: *Opera2D* prediction for current distribution in LHC injection PFN for a current of 6 kA

Opera2D simulations have been carried out to assess the frequency dependence of inductance and resistance of the coil of the LHC injection PFN [19]. Figure 22 shows the predicted current distribution in the PFN coil at frequencies of 0.1 Hz, 200 Hz, 1 kHz, and 40 kHz: the nominal DC current density is 27.3 kA/cm^2, for a total current of 6 kA.

Figure 23 shows the predicted inductance of a 7-turn cell versus frequency for a mean radius of the coil of 41.5 mm. The 'Grover limits' refer to values calculated from equations [19]. The reduction in inductance as frequency is increased from DC to a few hundred Hertz is mainly due to shielding attributable to the Ω-shield. The reaction field from the eddy currents induced in the Ω-shield reduces the flux density along the axis of the coil from 0.343 T near DC to 0.315 T at a few hundred Hertz, for a current of 6 kA. As the frequency is increased beyond a few hundred Hertz the inductance decreases, mainly due to skin effect and proximity effect within the coil. Conduction losses along the PFN coils result in a droop of the pulse of approximately 0.5% in the kicker magnet. *PSpice* simulations, subsequently confirmed by measurements, show that conduction losses in the PFN coil can be compensated for by increasing the PFN capacitor values linearly from the MS end to the DS end [68].

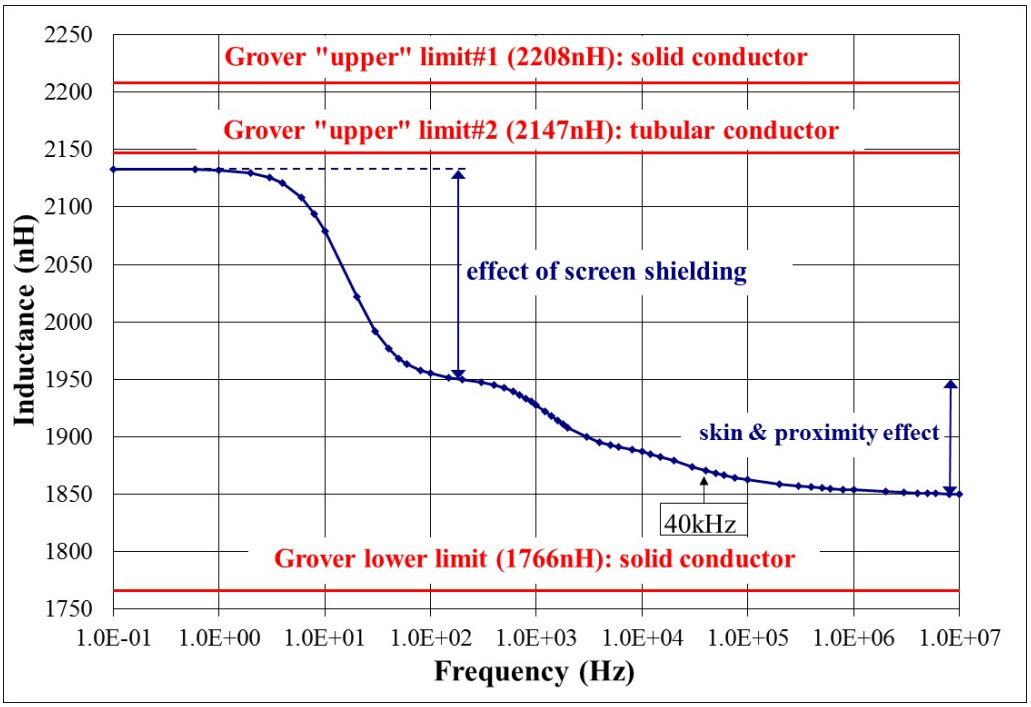


Fig. 23: *Opera2D* prediction for inductance of a 7-turn PFN cell, versus frequency, for a mean radius of the coil of 41.5 mm.

5.9 Power switches

5.9.1 Thyratrons

Despite the advances in high-power solid-state switches, deuterium thyratrons are still widely used as the power switch in kicker systems. The deuterium thyratron is a high peak power device, which uses deuterium gas as the switching medium. The switching action is achieved by a transfer from the insulating properties of neutral gas to the conducting properties of ionized gas [69]. Voltage breakdown in the deuterium-filled gap is initiated by free charges (electrons and ions) crossing the gap under the influence of an electric field. If sufficient energy is available, gas molecules are ionized producing more free charges. The positive ions are accelerated towards the lower potential electrode and cause the release of secondary electrons. Under the right circumstances, the processes become self-sustaining and voltage breakdown occurs [69].

Thyratron commutation is achieved by introducing plasma into the grid/anode region via slots in the grid structure. The plasma is created in the cathode/grid region by a fast-rising trigger pulse applied to the grid(s), which then diffuses to the grid slots where it comes under the influence of the anode field [69].

Three-gap thyratrons can hold off 80 kV and switch 6 kA of current with a 30 ns rise time (10% to 90%) [~150k A/μs]. The individual gaps in a three-gap thyratron break down in sequence [70, 71]. Initially the gap closest to the cathode conducts and the full PFN voltage is shared between the centre and anode gaps. Approximately 50 ns later the centre gap starts to conduct and the full PFN voltage builds up across the anode gap [70]. The voltage redistribution between the parasitic capacitance of each of the gaps is associated with a flow of displacement current [71]. The displacement current also flows in the external circuit, and hence through the kicker magnet, and can increase the effective rise time of the kick [72]. Appropriately placed saturating inductors can be used to reduce both the effect of the displacement current and the rise time of the main current pulse [71, 73].

Special care must be taken with the high-power thyratron:

- coaxial housings are used to achieve low inductance but there must be adequate insulation to the housing;
- current flow in the connections around the thyratron should be symmetrical. An asymmetrical magnetic field can impose forces on the internal plasma preventing uniform current density which may affect performance and lifetime [69];
- the appropriate thyratron must be selected (e.g., suitable rating for peak current and average current) and properly applied for the anticipated short circuit and fault conditions;
- the thyratron must be adequately cooled to ensure that the maximum specified envelope temperature is not be exceeded;
- erratic turn on (turn on without a trigger being applied) can result in operational problems: erratic turn on is reduced significantly by 'fast' (~ms) charging of the PFN/PFL immediately before the kicker magnet is required to be pulsed (see below);
- the reservoir voltage of the thyratron should be optimized. The gas density in the grid/anode gap must be maximized, to minimize current rise time and switching losses and maximize thyratron lifetime, consistent with achieving a low rate of erratic turn on.

In order to reduce as much as possible the number of erratic turn ons, RCPSs are used to charge PFNs for kicker magnets [3, 74]. The number of erratic turns on of a thyratron switch is dependent upon several variables, which include the duration for which there is a high voltage across the thyratron; thus, in order to minimize the number of erratics, it is advantageous to minimize this time period. In general, an RCPS can also allow the reservoir voltage to be increased, if the rate of erratic turn on is still acceptable, and therefore help to extend thyratron lifetime.

The fast RCPS developed for the LHC injection kicker systems charges two PFNs to 54 kV in approximately 800 μs: approximately 600 μs later the PFNs can be discharged into the kicker magnets [75]. However, in reality, the LHC injection PFN are fully charged for between 2 ms and 3 ms before the MSs are triggered. During 2015 and 2016, for the eight LHC injection kicker magnets, there was a total of approximately 250,000 MS thyratron injection pulses, one of which resulted in a MS thyratron erratic during resonant charging [63].

To achieve significantly improved performance and lifetime a thyratron can be double pulsed to turn it on: the first pulse, applied to grid 1, pre-ionizes the cathode grid space, which prepares the cathode region for conducting the main current pulse. The second pulse, which is applied to grid 2, should be delayed by at least 500 ns with respect to the first pulse, provides a fast rising voltage to ensure precise triggering (1–2 ns) of the thyratron [69].

As mentioned in section 5.2, the length of the pulse in the kicker magnet can be controlled in length, between 0 and $2\tau_p$, by adjusting the timing of the thyratron DS (Fig. 4) relative to the thyratron MS. However, the wavefront generated by the DS turn-on must travel through the PFL/PFN: as a result of attenuation and dispersion in the PFL/PFN, once the wavefront reaches the MS it has a slower falling edge than the corresponding rising edge that was generated by the MS turn on. Hence, the field fall time of the kicker magnet would be longer than the rise time. If a fast field fall time is required a 'clipper switch' thyratron can be used: this thyratron would be connected from the PFN/PFL side of the MS to ground. This clipper switch can also be used to reduce the duration of the field pulse; hence, the DS thyratron would typically be replaced with an inverse diode. An example of a clipper switch, combined with an inverse diode at the remote end of a PFN, is shown in reference [76].

Control of the duration of the field pulse, either by a DS or a clipper switch, can also be an important protective measure during a fault condition. For example, for the LHC injection system, if there is an erratic (untriggered turn on) of an MS thyratron, already circulating beam could be miskicked: hence, in case of an MS erratic in this system, the remaining MS thyratrons and all the DS thyratrons,

of the same system, are triggered: triggering the DS thyratrons reduces the field pulse duration and hence the potential number of bunches which are miskicked.

As mentioned above, a thyratron switch is susceptible to erratic (uncontrolled) turn on, which can result in circulating beam being miskicked and hence may cause considerable damage to accelerator components. Power semiconductor switches potentially have advantages, over thyratrons, of reduced erratic rate, increased lifetime, and reduced maintenance.

5.9.2 Power semiconductor switches

Despite progress in the capabilities of fast semiconductors switches, they still have limited voltage and/or current capability in comparison with thyratrons. In addition, semiconductors may be influenced by ionizing radiation and neutron flow. Depending on the type of radiation, the semiconductor component type and its working conditions, the radiation effects can be cumulative with relatively slow deterioration of semiconductor performance and/or sudden malfunction or failure: the latter is called single event effects (SEE) [77, 78]. The cumulative effects, which result from total ionizing dose and displacement damage, are responsible for the modification of component parameters such as leakage current, bipolar transistor gain, opto-coupler efficiency, field-effect transistor (FET) threshold voltage, voltage reference value, etc. On the other hand, a SEE provokes sudden malfunctions, which, in the case of HV power semiconductors, usually leads to component failure: this is known as single event burnout (SEB) [79-81]. For a given radiation dose the SEB rate is strongly dependent on the applied voltage: there is a very steep increase of the failure rate when the applied voltage is higher than a certain percentage of the components' rated voltage. The general recommendation for high voltage components, to maintain a reasonably low SEB rate, is to apply a maximum DC voltage of approximately 50% of the semiconductor rating. For the LHC abort kickers (section 5.9.2.1) the power semiconductors can be affected by radiation and neutron flow from the LHC [77].

Presently a range of power semiconductors are either in use in the pulse power generators of various kicker systems at CERN, or are in prototype generators. A number of examples of applications are described in the following sections.

5.9.2.1 LHC dump (abort) kicker system

In some applications thyratron switches cannot be used, e.g., for the dump (abort) kickers in the LHC where no self-firing is allowed [82]. The nominal LHC beam consists of 2808 bunches per beam, with 1.5×10^{11} protons per bunch. Once accelerated to 7 TeV the stored energy per beam is 470 MJ. A beam dump, an 8 m long graphite block, is the only element that can safely absorb this energy. The LHC beam dumping system (LBDS) must protect the LHC from damage by reliably and safely extracting and absorbing the circulating beams when requested. The LBDS consists, per beam, of 15 horizontally deflecting extraction kicker magnets (MKD), 15 vertically deflecting septum magnets and 10 dilution kicker magnets (MKB), followed by several hundred meters of beam transfer line before the extracted beam reaches the dump block [83].

The MKD kicker system is the most safety critical and complex element of the LBDS as a fault in the system can lead to significant damage of LHC equipment. The extraction of the beam from the LHC is at a fixed angle of 0.275 mrad, by the MKD magnets, hence the MKD system needs to track the beam energy (between 450 GeV and 7 TeV). The maximum operating voltage of the MKD system is 30 kV to generate a pulsed current of 19 kA with a flat top of 95 μs. The MKDs should not turn on erratically during LHC operation, and must reliably switch on within a 3 μs particle-free abort gap when required [84]. To minimize spontaneous self-triggering, semiconductors are used for the power switches: fast high current thyristors, from ABB and Dynex, are utilized. The semiconductor switches permit the required wide dynamic range of operation [82].

Each semiconductor switch module for the LHC abort kickers consists of ten series gate turn off (GTO) thyristors (Fig. 24) that have been optimized for turn on. The maximum forward voltage rating of each GTO is 4.5 kV and the switch module is operated over a voltage range from 2.2 kV (450 GeV) to 30 kV (7 TeV) [82]. The current switched is in the range from 1.2 kA to 19 kA: the specified maximum rate of rise of current is 20 kA/μs, which corresponds to ~1/8th of the capability of a thyratron. However, recent tests on similar GTOs, with a high-current gate drive, have demonstrated a capability of more than 32 kA/μs. In order to achieve high reliability each generator has 2 parallel branches, either of which can carry full load current. At least 14 out of the 15 kicker magnets must operate simultaneously on request to safely extract the beam from the LHC. Any spontaneous firing or fault of one of the pulse generators is detected and the other generators are triggered [82].



Fig. 24: Stack of-high power GTOs (LHS) and individual GTOs (RHS)

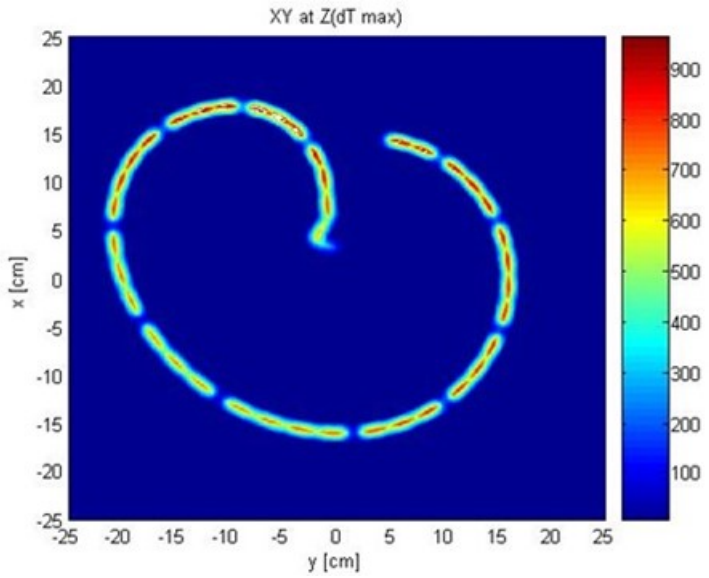


Fig. 25: Lissajous figure of LHC extracted beam on the dump block – right hand side scale shows temperature (°C).

The function of dilution kickers is to sweep the extracted beam in a Lissajous figure (Fig. 25) on the dump block: the circumference of the Lissajous figure is approximately 1.2 m. Separate horizontal

and vertical MKB systems are used to generate sine and cosine-like current shapes over 90 μs. The peak deflection angle is 0.28 mrad (for 450 GeV to 7 TeV beam). Each MKB generator comprises a capacitor coaxially connected to the magnet inductance. The capacitor is discharged through a solid-state closing switch: the discharge pulse is an attenuated sinusoidal oscillation of 24 kA maximum amplitude (at 7 TeV) with a period of 70 μs.

5.9.2.2 Inductive adder

5.9.2.2.1 Compact linear collider damping ring

The compact linear collider (CLIC) would be a high-energy electron−positron collider. It could provide very clean experimental environments and steady production of all particles within the accessible TeV energy range. To achieve high luminosity at the interaction point, it is essential that the beams have very low transverse emittance: the pre-damping ring (PDR) and damping ring (DR) damp the beam emittance to extremely low values in all three planes [85].

Stripline kickers are required to inject beam into and extract beam from the PDRs and DRs [86]. Jitter in the magnitude of the kick waveform causes beam jitter at the interaction point [87]. Hence, in particular, the DR extraction kicker must have a very small magnitude of jitter: the 1 GHz specifications call for a 12.5 kV pulse of 900 ns flat-top duration, with a combined ripple and droop of not more than ±0.02% [85]. An inductive adder (IA) has been selected as a promising means of achieving the specifications for the DR kicker [88, 89]. The IA is a solid-state modulator, which can provide relatively short and precise pulses: it is modular and thus the design can be adapted to other requirements: Fig. 26 shows a simplified electrical schematic of an IA. Photographs of prototype IAs are shown in Fig. 27: the LHS shows a 5-layer IA and the RHS a 20-layer IA.

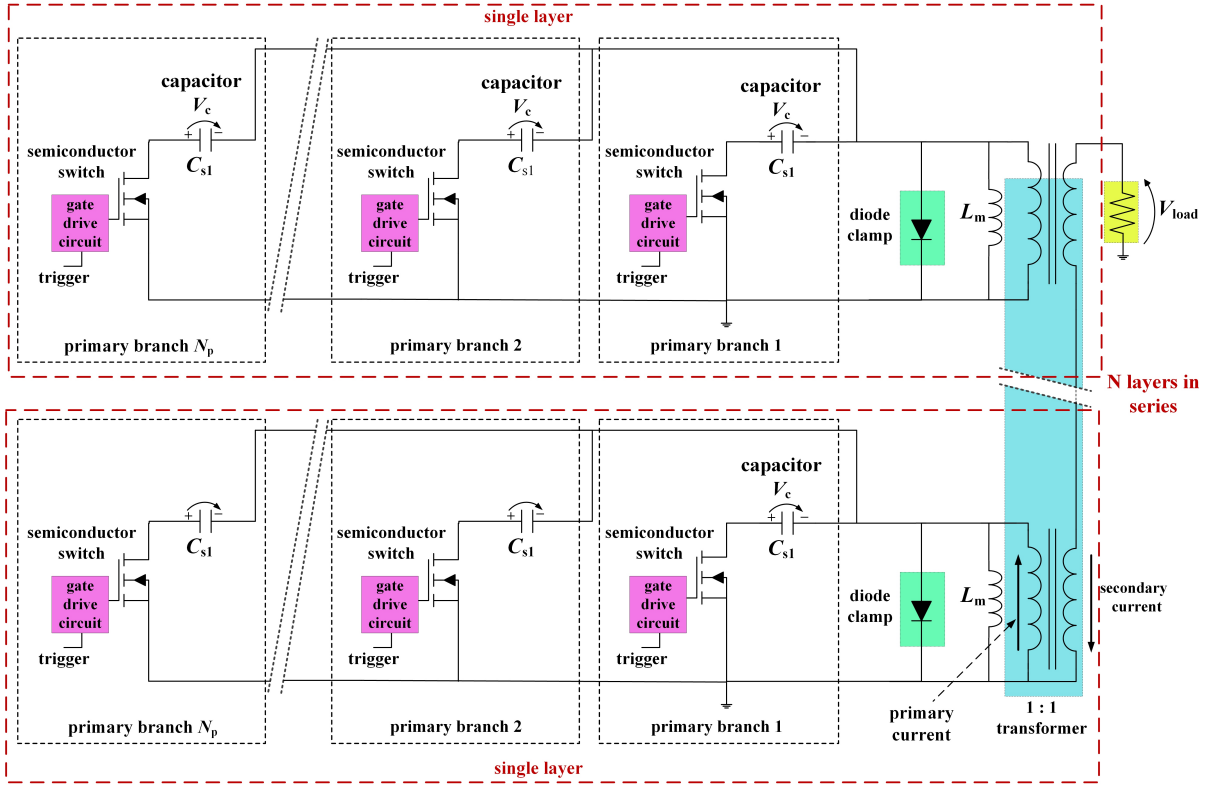


Fig. 26: Simplified schematic of an inductive adder showing *N*-layers in series on the secondary and *N*p parallel primary branches per layer.

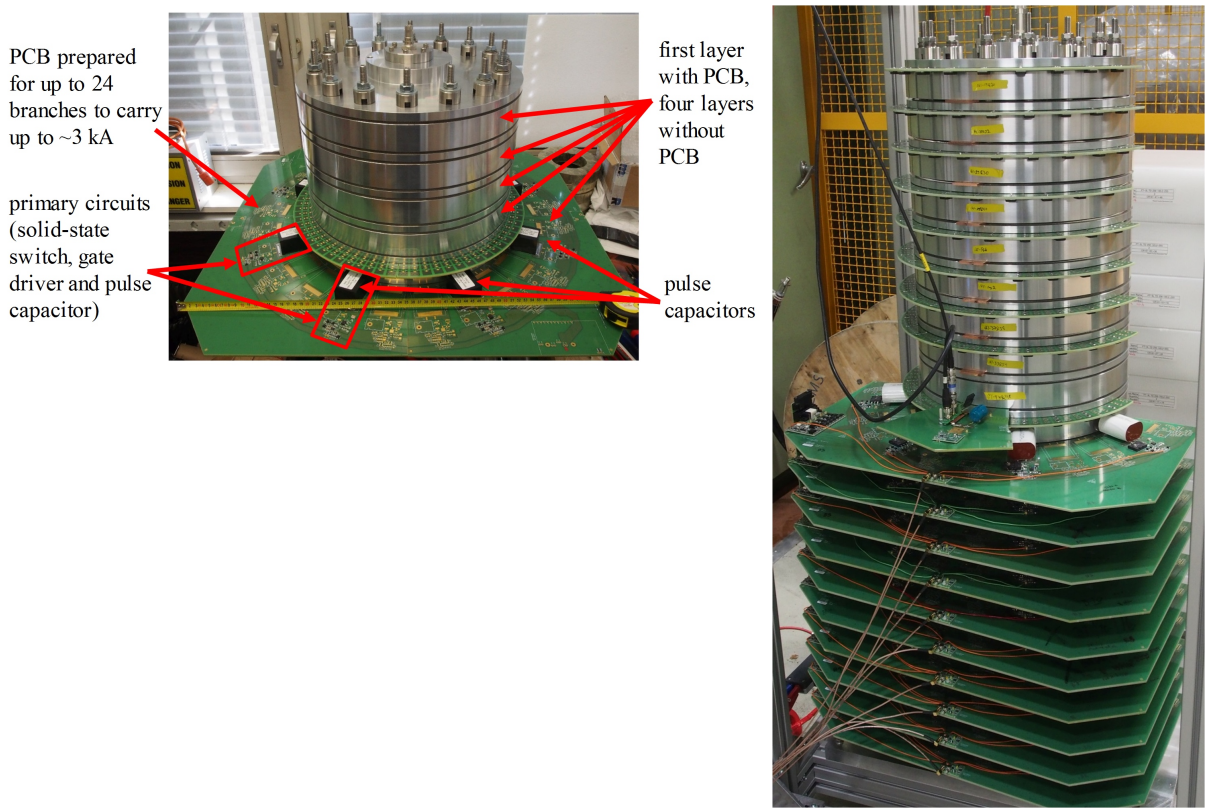


Fig. 27: Prototype inductive adders at CERN: 5-layer IA (LHS) and 20-layer IA (RHS)

An inductive adder (Fig. 26) consists of multiple layers (also known as stages), each of which has a transformer: the transformer usually has a 1:1 turn ratio in order to ensure that it is suitable for fast pulses. The single turn primary totally encloses a magnetic core; hence, the leakage inductance of this geometry is negligible [90]. The secondary winding of each of these transformers is connected in series: hence a step-up voltage ratio of 1:N is achieved by using N-layers, with adequate voltage isolation. The primary circuit typically has many parallel branches: each branch generally contains a single capacitor and a single power semiconductor switch. All the power semiconductor switches and gate drive circuits are referenced to ground and there are no electronics referenced directly to the high-voltage output pulse. In general, the capacitors of all the layers can be charged with a single power supply—this is not, however, necessary and in some applications the capacitors are deliberately charged to different voltages to allow modulation of the output pulse [91].

An IA has good built-in fault tolerance and redundancy, because if one or more solid-state switches of a layer of the inductive adder stack fails to turn on, the magnitude of the output pulse of the IA is reduced by only the voltage of a single layer. However, in order that the failure of one power semiconductor switch does not result in a short circuit on the output of a power supply, and hence failure of the inductive adder to produce adequate output pulses, the capacitors of each layer must be sufficiently decoupled from the power supply using resistors. Similarly a second power supply can be used to provide redundancy in the case of failure of one supply.

Since the IA is of a modular design, extra layers can be added to improve redundancy and/or increase voltage rating. In addition, the modular construction allows for a good scalability, adaption of individual components to the particular application, and a path for future upgrades. A drawback of the IA is that, as a result of it using magnetic cores for the transformer, it has limitations concerning the available pulse duration: however many kicker magnet applications at CERN require pulse durations of only up to 2.6 μs, which is achievable with an inductive adder.

As a result of the low source impedance, the source voltage does not need to be doubled, unalike a matched impedance PFL or PFN. The solid-state switches typically used in the IA are either metal-oxide-semiconductor field effect transistors (MOSFETs) or insulated-gate bipolar transistors (IGBTs). These solid-state switches can be opened when conducting full load current, hence only a portion of the stored energy is delivered to the load during the pulse (therefore a PFL or PFN is not required). In addition, opening of the solid-state switches potentially limits the duration of fault current in the event of a magnet (load) electrical breakdown.

Considerable research and development has been carried out on the IA for the CLIC DR [89]. A significant advantage of the IA over some alternative technologies is the ability to modulate the output [89]. Recent measurements on a prototype IA, equipped with 18-layers including a special analogue modulation layer, are shown in Fig. 28 [92].

Fig. 28 shows the best measured flat-top waveform for two different measurements of an 18-layer IA: the orange curve (top) is an average of 1000 measured pulses and the blue curve (bottom) is an average of 100 measured pulses. The offset difference of the measurements was most probably caused by thermal drift in components or in the high-voltage power supply, which charges the pulse capacitors [92]. For the blue curve, the measured flat-top stability is ±0.03% (±1.80 V) over the full 900 ns flat-top duration, at 6.3 kV. However, the flat-top stability is ±0.02% (±1.35 V) over the first 160 ns and the last 160 ns of the flat-top duration, as labelled in Fig. 28 [92]. This fulfils the CLIC DR requirements for the flat-top stability of the pulse modulator, for both 1 GHz and 2 GHz specifications of the damping rings [85], at approximately a half of the nominal required voltage.

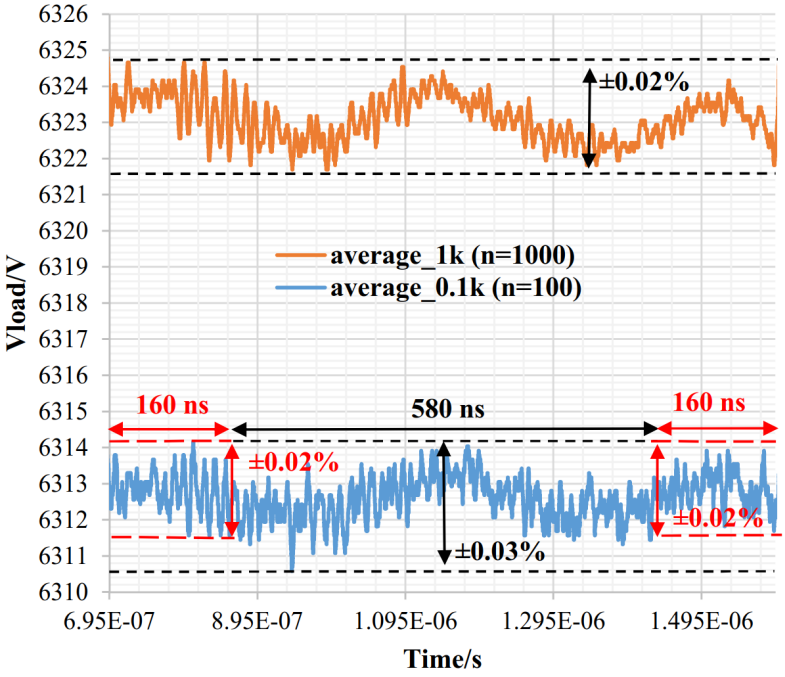


Fig. 28: Measured load voltage with active analogue modulation. The orange curve (top) and blue curves (bottom) are an average of 1000 pulses and 100 pulses, respectively [92].

It should be noted that theoretically a double kicker system can be used to reduce the influence of flat-top ripple upon the beam deflection (see section 5.11).

5.9.2.2.2 CLIC Combiner Ring

The specifications for the CLIC Combiner Ring (CR) extraction kickers call for a burst rate of up to 688 kHz for 140 μs, every 20 ms [85]. The required pulse voltage, of 10 kV into 50 Ω, corresponds to

200 A. The rise and fall time should be no more than 150 ns and the flat-top pulse duration will be between 245 ns and 450 ns, with 96 pulses per burst. Hence, the average repetition rate is 4.8 kHz [93].

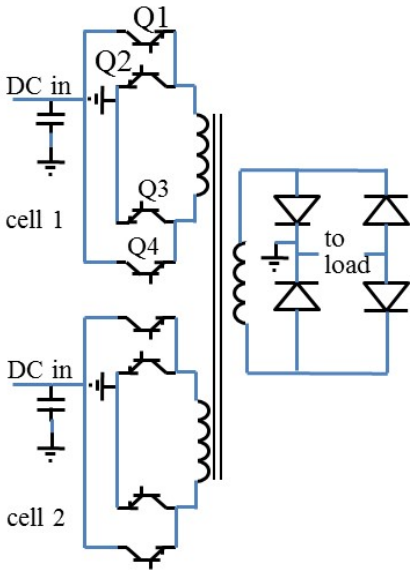


Fig. 29: Novel variant of the typical IA, employing a full-bridge inverter [93]

The energy transferred during a pulse is approximately 1 J, depending upon the flat-top duration, which corresponds to 96 J per burst. The maximum allowable droop during the burst is ~0.25%, corresponding to 25 V. If this energy were stored in a single capacitor, the 10 kV capacitor would need to store at least 19 kJ, which requires a large capacitance value (384 μF). An alternative approach is to use an inductive adder with appropriate droop correction. Assuming that the magnetic core used for the IA must support the voltage × time integral of a burst, this corresponds to a total of ~0.6 V·s. Assuming a 0.5 T saturation flux density for the magnet material, this gives a total cross-sectional area of the core of ~1.2 m^2 [93], which is very large. To reset the cores during the burst, between pulses, requires a reset current of ~100 A, which is similar to the magnitude of the pulse current.

Collaborators at Stanford Linear Accelerator (SLAC) developed a novel variant of the IA topology, employing a full-bridge inverter: Fig. 29 shows a two-cell example. The two cells produce alternating positive and negative polarity pulses, but the load output is single polarity. The flux alternates in direction in the magnetic core: hence, the required cross-sectional area of core material can be reduced by almost two orders of magnitude. Tests and measurements on an initial low-voltage prototype show that long pulse trains can be achieved [93].

5.9.2.2.3 FCC Injection

An IA is also being studied for the injection system of a 100 TeV centre-of-mass future circular collider (FCC) study. This IA design is based on silicon carbide (SiC) MOSFET switches [94]. The FCC injection kicker system would need to deflect 3.3 TeV beam, and the required field rise and fall times are 430 ns: to allow a reasonable magnet fill time, the rise and fall times of the current pulse from the generator need to be <75 ns. A potential drawback of the IA is the output transformer and therefore the risk of saturation of the magnetic core. However, with a maximum flat-top pulse duration of 2 μs, for machine protection reasons, the FCC injection kicker can use an IA for the pulse generator.

Although the available length for the kicker magnets is 150 m, for beam impedance and stability reasons it is desirable to use a significantly shorter length, but this increases the current required from the IA. As a compromise, a current of 2.5 kA and a magnetic length of ~30 m have initially been chosen: the corresponding magnet deflection angle is 0.18 mrad [95]. A system impedance of 6.25 Ω has been selected: this results in an output voltage of 15.7 kV for the inductive adder [94].

5.9.2.2.4 CERN PS KFA45

Semiconductor switch technology is being considered for the CERN PS KFA45 kicker system: it is a 45-year old PFL based system and it is becoming very challenging to find a commercial replacement for the 80 kV SF6 gas filled coaxial PFL cables. Initial specifications for a replacement pulse generator for the PS KFA45 system included a current of 1.52 kA with 2.6 µs flat-top duration and a kicker magnet impedance of 26.3 Ω, terminated in its characteristic impedance: during the flat top the allowable droop and ripple is ±2%. The current and impedance correspond to an output voltage of 40 kV for an IA. The nominal single-way delay of the kicker magnet is 35 ns. The required rise and fall times, defined between 1% and 99% of the field pulse, are 46 ns and 96 ns, respectively [96]. Hence, it is presently assumed that fast switching MOSFETs would be required for the power semiconductor switches. However commercially available MOSFETs presently have a maximum voltage rating of 1.2 kV to 1.7 kV: hence in the region of 50 layers would be required for the IA.

Studies have commenced for an IA system to replace the PFL and thyratron switches of the CERN KFA45 system. These simulations show that minimizing the delay of the IA stack is of paramount performance to achieving fast rise time: thus the height of each layer must be optimized, as must the primary inductance and coupling capacitance between primary and secondary. This will require careful consideration and characterization of the magnetic material to be used and the choice and layout of primary components, especially the pulse capacitors and power semiconductor switches [97].

Recent specifications for the CERN PS KFA45 kicker system require that the kicker magnet is terminated in a short circuit: this will double the required current to more than 3 kA. Modifications of the IA are under consideration so that it can operate with a kicker magnet terminated in a short circuit.

5.9.2.3 Marx generator

In a Marx generator n capacitors are charged in parallel from a relatively low-voltage DC power supply, Udc, and discharged in series into the load, producing an output voltage pulse (V_o) with an amplitude approximately equal to the number of stages (n) times the input voltage (Udc), $V_o = n \cdot Udc$, an idea originally developed by Erwin Marx in the 1920s. In the original Marx generator, spark gaps were used to connect the cell capacitors in series, and resistor and inductors between stages were used to limit charging current, and self-discharge paths [98]. With the improvement in solid-state switch characteristics, major progress has been made on the development of Marx generator type topologies based on semiconductors [99−104].

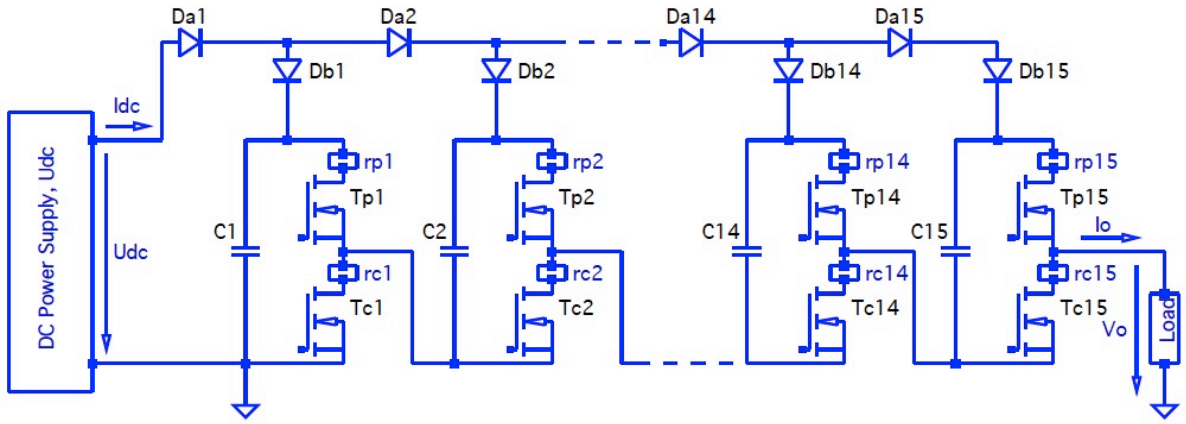


Fig. 30: Simplified schematic of a solid-state Marx generator [105]

The simplest pulse characteristic that a Marx generator can produce is an RC-decay: in this case the capacitance of the erected Marx discharges into a resistive load. The advantage of using the Marx topology, rather than a single capacitor at high voltage, is that the maximum voltage across components

in a Marx cell is the input power supply voltage, *Udc*. Therefore, relatively low-voltage components can be used: since each cell comprises low-voltage components, it is relatively easy to minimize parasitic inductance through the use of closely coupled image planes [99]. In addition, high voltage only exists in the Marx generator when the Marx is erected to produce an output pulse. Also, the energy inside the Marx is stored in several capacitors, instead of one, which has advantages in terms of cost and safety.

Fig. 30 shows a simplified circuit of a solid-state Marx generator [105]. The Marx comprises an energy storing capacitor *Ci*, two set of diodes *Dai* and *Dbi*, and a half-bridge with two MOSFET switches. MOSFETs labelled *Tpi*, where "*i*" is an integer, in the top of the half bridge, apply pulses to the load. The MOSFETs in the bottom of the half-bridge, labelled *Tci*, are for charging capacitors *Ci* between pulses and for defining the falling edge of the output pulse.

One advantage of the Marx generator over the inductive adder is that the Marx does not require magnetic cores to directly generate the output pulse: hence, there is the possibility to generate longer pulses, if the energy stored in the capacitors of the Marx is sufficiently large: this is also dependent upon the permissible droop of the output pulse.

Capacitors *Ci* are charged in parallel from power supply *Udc* through diodes *Dai*, *Dbi*, and MOSFETs *Tci*: during the charging period the load is pulled down to 0 V by turning on MOSFETs *Tci*. The pulse is applied to the load by first turning off MOSFETs *Tci* and then by turning on MOSFETs *Tpi*, to connect capacitors *Ci* in series. The diodes *Dbi* are included for safety reasons, as they limit the discharge of each *Ci* capacitor into the *Ci+1* onwards capacitors, reducing the damage if a *Ci+1* capacitor fails [106].

5.9.2.3.1 CLIC RF breakdown studies

CERN collaborated with *Instituto Superior de Engenharia de Lisboa* (ISEL) and *Energy Pulse Systems* (EPS), Portugal, to develop a high repetition-rate, semiconductor-based, Marx generator. The Marx has been designed to power a spark-gap system: this system allows the study of breakdown rate phenomena for different materials and surface finishes, for RF structures for the proposed CLIC. This Marx circuit is based on SiC MOSFETs, with 15 compact stages, capable of running with capacitive type loads during normal operation, and also in short-circuit mode at reduced repetition rate. The Marx generates pulses of up to 10 kV with widths from 200 ns to 100 µs, at frequencies up to 1 kHz and pulse voltage rise time of less than 100 ns (Fig. 31) with a load capacitance of 150 pF [106].

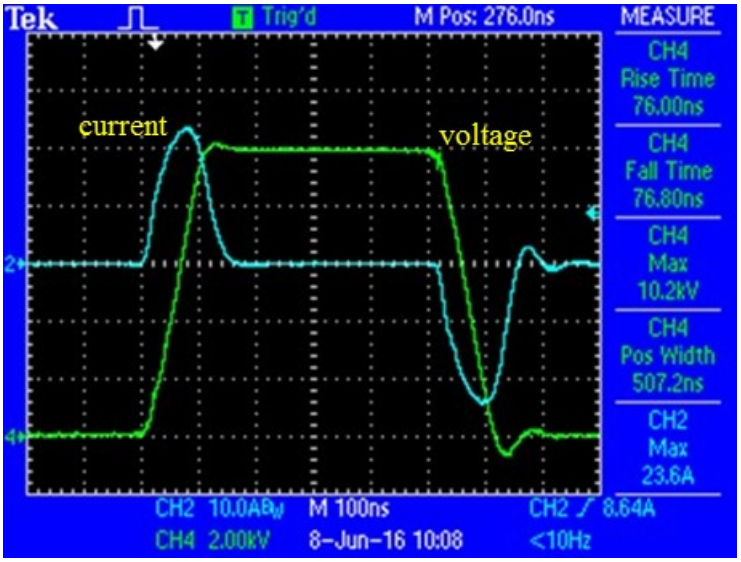


Fig. 31: Measured pulse into a 150 pF load, voltage (green)—2 kV/div, and current (blue)—10 A/div, 100 ns/div [106].

5.9.2.3.2 CERN PS KFA45

As mentioned in section 5.9.2.2.4, the CERN PS KFA45 kicker system is a 45-year-old PFL based system and it is becoming very challenging to find a commercial replacement for the 80 kV SF6 gas filled coaxial PFL cables. Hence, it would be highly advantageous to replace both the PFL and thyratron by a semiconductor based pulse generator, e.g., an IA or a Marx generator.

The specifications of the output pulse from the pulse generator are 40 kV, 3 μs flat-top duration, 30 ns rise and fall times, and 1 Hz repetition rate: it is proposed that the output of the magnet would be terminated in a short circuit and, hence, the load current would be 3.2 kA. The rise and fall time specifications are particularly challenging. A theoretical study for a Marx generator topology, using approximately 50 stages, of 800 V per stage, has commenced; each stage would consist of 24 SiC MOSFETs in parallel. Each MOSFET would conduct approximately 140 A pulses. A 4-stage prototype has been designed and tested, with almost full current at 3 kV output pulse voltage: this prototype allows the proposed design to be evaluated, in order to assess the technological advantages of such a system and limitations for the scale up to full voltage [107].

A Marx stage consists of 24 SiC MOSFETs in parallel, arranged in eight so-called 'switch units' of three MOSFETs per switch unit. A 60 μF capacitance is associated with each switch unit, giving 480 μF per stage. The capacitor bank associated with each switch unit comprises two parallel 30 μF capacitors: this parallel arrangement decreases the series inductance and increases the total current carrying capability. Hence, there are sixteen capacitors in total per stage.

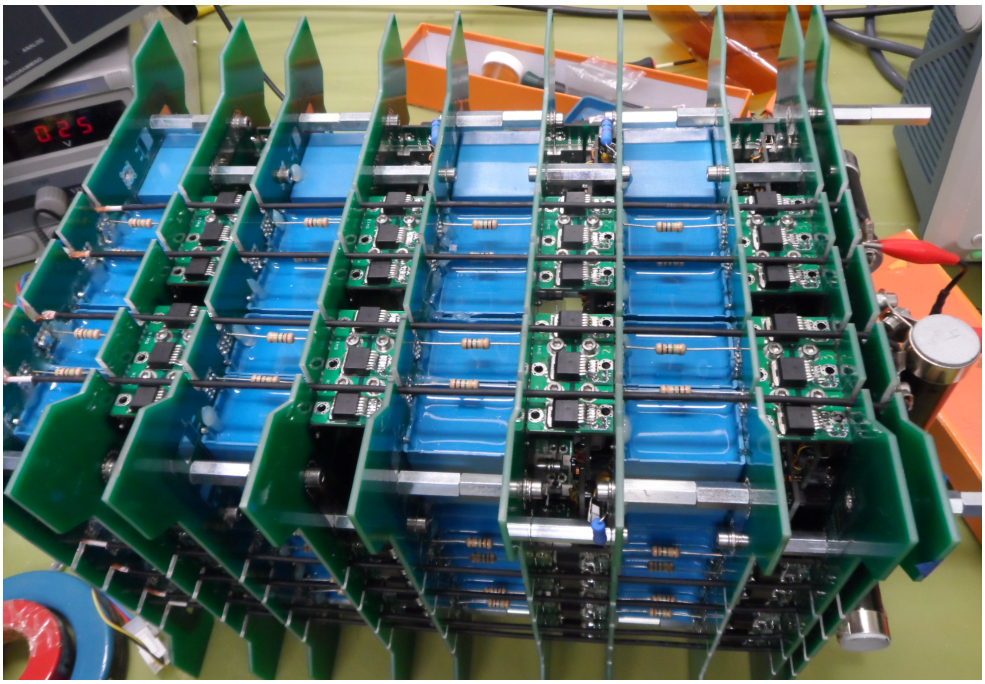


Fig. 32: Four Marx stages: four return conductors per side

The four stages of the Marx generator are assembled in a coaxial type structure, as shown in Fig. 32, in order to limit the loop inductance. The current flows to the load through the MOSFETs and capacitors, placed around the sides of this structure, and returns back through external wires, four per side. A four stage Marx generator was tested to obtain a load current pulse of 2.7 kA, and 3.2 kV, into a 1.18 Ω resistive load: the measured load voltage pulse is shown in Fig. 33. The measured rise and fall times, 10% to 90%, were both ~50 ns. The measured droop is ~1.6% during the 1.5 μs flat top: the expected droop during this period, estimated from an effective capacitance of 120 μF and 1.18 Ω load resistance, is 1.1%.

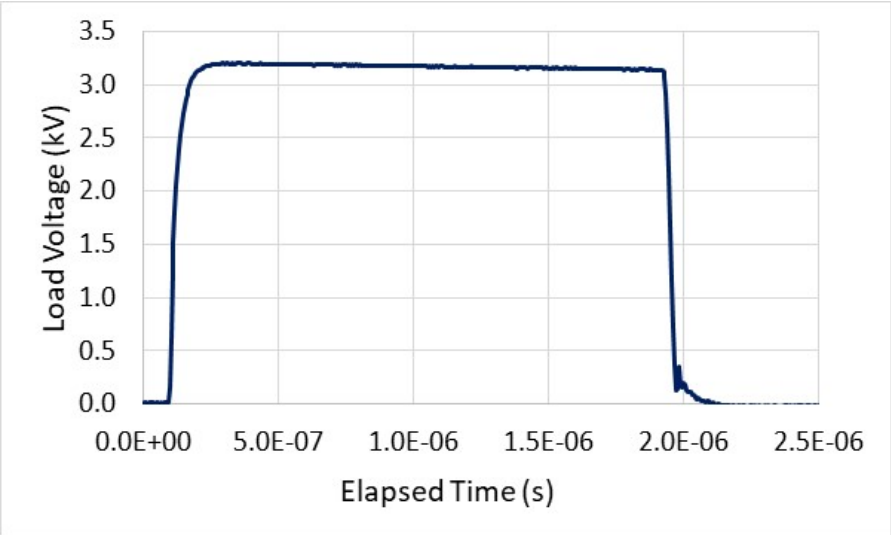


Fig. 33: Measured load voltage pulse for four Marx stages (pre-charge = 800 V/stage)

Although, the initial measurements and test results are promising, there are ongoing tests to determine whether the Marx generator is a good candidate for high current pulsed power generators for kicker magnet applications. In addition, the scaling up of this concept has many complexities including the separation between the main circuit and wires carrying the return current [107]. Furthermore modifications of the Marx generator are under consideration so that it can operate with a kicker magnet terminated in a short circuit. Future tests will include long-term stability and reliability.

5.9.2.4 Series switch topology

The simplest solid-state modulator topology consists of a capacitor bank and a stack of solid-state switches connected in series. The modulator can also be connected to the load with a step-up transformer. Figure 34 shows a simplified schematic of a series switch modulator topology. The advantages of the series switch topology are that it is relatively simple and it can be used to generate both short and long pulses. If there is not an output transformer, the maximum pulse length is a function of the value of the capacitor bank, the load impedance, and the acceptable droop [89].

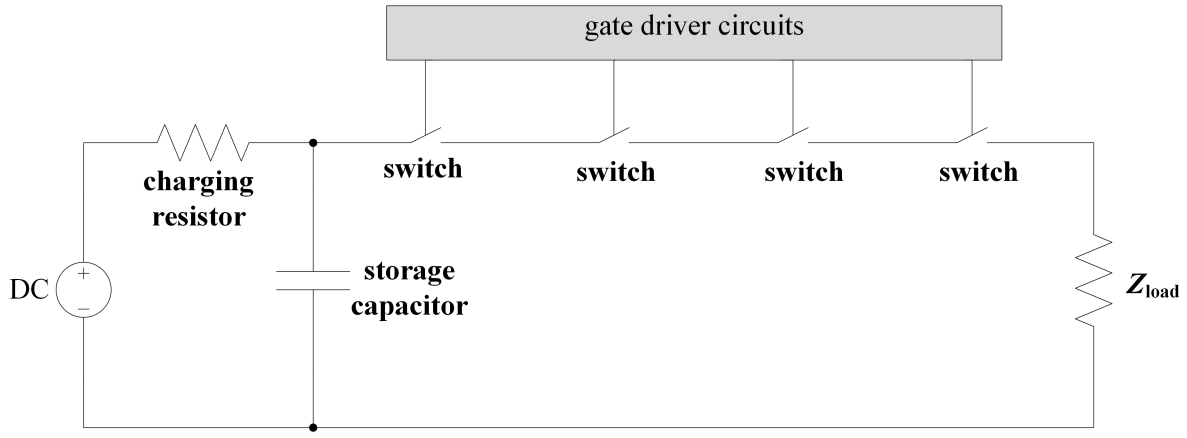


Fig. 34: Simplified schematic of a series switch topology [89]

The disadvantages of the series switch topology include the fact that the gate drive circuit of the power switches are electrically floating and not referenced to the ground: this requires use of transformers or fibre optics between any ground referenced control circuits and the gate driver for the power semiconductor switches. The timing of the trigger pulses has to be well synchronised at the gates

of the power switches to keep the voltage stress over the switches to an acceptable level. Special care must be taken in designing the fibre optic link for the trigger signals, in order to avoid time jitter between the stacked switches [108]. In addition, if a single switch in the stack fails, the voltage stress over other switches is increased and this may cause the other switches to fail, unless appropriate redundancy is included.

A pulse power modulator developed at Saskatchewan Accelerator Laboratory generated up to 20 kV pulses with 10 A of load current. The flat-top length could be adjusted from 0.5 μs to 6 μs, the repetition rate was up to 360 Hz and the rise and fall times were 100 ns and 1 μs, respectively [109]. A bipolar pulse modulator was developed at TRIUMF. This modulator generated ±15 kV pulses with adjustable pulse length from 25 ns to 120 ns. The rise and fall times were less than 32 ns, from 10% to 90% of the pulse voltage, and the repetition rate was fixed at 930 kHz. To achieve these specifications the pulses were reflected back and forth on a low loss coaxial cable [110].

A pulse power modulator designed for a kicker for the ISAC facility at TRIUMF generated unipolar −3.5 kV pulses with a repetition rate up to 52 kHz. The pulse width was adjustable from 350 ns to longer than 10 s and the rise and fall times were 63 ns, from 10% to 90%. To achieve pulse durations of up to 10 s a novel gate driver circuit was developed [111].

A MOSFET stack based pulse modulator for muon lifetime measurements at Paul Scherrer Institute, Switzerland, generated pulses up to ±12.5 kV for deflector plates of a kicker system. The load consisted of deflector plates, the capacitance of which was in total 108 pF. The length of the pulse flat top was adjustable from 200 ns as a minimum, and the rise and fall times less than 40 ns from 10% to 90% of the pulse voltage. The repetition rate was up to 77 kHz [108, 112].

A pulse power modulator operating at over 2 MHz repetition rate and based on stacked switch technology was developed in Extreme Energy Density Institute at Nagaoka University of Technology in Japan. This modulator delivers 5 kV pulses to a 70 Ω load. The peak current was 75 A, operation frequency up to 2.1 MHz, rise and fall times of 33 ns and 43 ns, respectively, and the pulse width of the load voltage was 240 ns [113].

A series switch modulator based on MOSFET switches has also been built to generate very short pulses. An example is a prototype pulse modulator for the damping ring kicker of the ILC. This pulse modulator produced 4.8 kV pulses with a pulse width of 14 ns and with rise and fall times close to 5 ns. In order to achieve the specifications for pulse width, the MOSFET stack was designed to be its own PFL [114].

5.10 **Striplines**

The CLIC study is exploring the scheme for an electron−positron collider with high luminosity and a nominal centre-of-mass energy of 3 TeV. The CLIC pre-damping rings and damping rings (DR) will produce, through synchrotron radiation, ultra-low emittance beam with high bunch charge, necessary for the luminosity performance of the collider. The role of the DRs is to provide the final stage of damping to the required low emittance beam, at a repetition rate of 50 Hz. The normalized emittance requirements are extremely low, 500 nm in the horizontal plane and 5 nm in the vertical plane. Once the beam is damped to the final emittance values, it is extracted from the DRs and injected into the booster linac, where the beam will be accelerated from 2.86 GeV to 9 GeV. Hence, the extraction system from the CLIC DR must not result in blow up of the beam emittance: any ripple or droop on the extraction kicker pulse will produce beam size jitter which could be propagated up to the interaction point of the collider. The CLIC DR have strict requirements for low beam-coupling impedance [85]: hence, stripline technology is an excellent choice [115−117].

Important challenges for the CLIC DR kickers are [116, 117]:

(i) excellent field homogeneity, with only ±0.01% of inhomogeneity allowed over 1 mm radius, in the centre of the aperture;

(ii) good power transmission, with extremely low ripple, by achieving good impedance matching of the striplines in the deflecting (odd) mode;

(iii) very low beam-coupling impedance, which should be below 0.05 Ω/turn in the longitudinal plane and 200 kΩ/m in the transverse plane: the beam-coupling impedance is important during the non-deflecting (even) mode.

Injection and extraction stripline kickers are used for fast injection/extraction applications, such as bunch-by-bunch manipulation, with a field rise time of a few ns. This is the case for the tapered striplines designed for the injection upgrade of the Φ factory DAΦNE [57], as well as the kicker system proposed for the ILC DR beam extraction [118]. For the CLIC DR extraction kicker, the curved and tapered electrodes used in DAΦNE would give higher field inhomogeneity than is permissible. Hence, a novel electrode shape has been developed for the CLIC DR extraction kickers that has allowed to achieve the excellent field homogeneity, good matching of both odd and even mode characteristic impedances, and a decrease in the beam-coupling impedance at low frequencies. These studies were carried out by using *HFSS* [21] and *CST* simulation codes [22].

The striplines have both an odd and an even mode impedance: the odd mode is when both electrodes are driven to opposite polarity voltages, to extract beam from the CLIC DR, whereas the even mode is when the electrodes are not driven by pulse generators. The characteristic impedance of both odd and even modes should ideally be optimized to 50 Ω. However, for coupled electrodes this is not possible to achieve. The difference between the even mode and odd mode characteristic impedances is due to the capacitance between the electrodes (C_{12}), and an infinite separation between the electrodes would be needed to uncouple the electrodes, i.e., to ensure that $C_{12} = 0$. Since the beam-coupling impedance is proportional to the even mode characteristic impedance, the even mode impedance is optimized to be 50 Ω. Nevertheless, it is desirable that the odd mode impedance is as close to 50 Ω as possible, to be able to match their impedance to that of commercially available coaxial cables.

Various shapes of electrodes were studied and optimized for beam-coupling impedance, odd mode impedance as close to the even mode impedance as possible, and excellent field homogeneity. In addition the electrode supports, feedthroughs, and manufacturing tolerances were studied. A novel shape of electrode, called a half-moon electrode, was selected as the optimum shape [117].

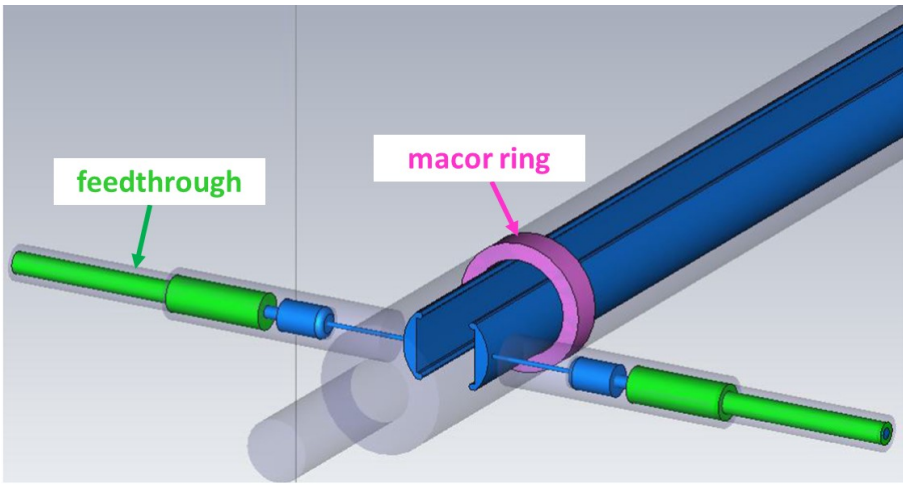


Fig. 35: *CST* model used to optimize the supports and the feedthroughs for the half-moon electrodes

For the optimized half-moon electrode shape, the odd mode characteristic impedance is 40.9 Ω. Since the inductive adder will be connected to each electrode with commercial coaxial cable of 50 Ω impedance, and each electrode will be terminated with 50 Ω (Z_L in Fig. 36), there is an impedance mismatch for the odd mode at both the input and output of the electrodes: impedance mismatches can increase the settling time of the pulse [117]. A new idea was proposed to match the load side characteristic impedance for both the odd and even modes of excitation of the striplines: this is achieved by adding resistor Z_m, shown in Fig. 36.

Simulations with *Opera2D* show that the internal inductance of the electrodes, of a stripline kicker, is frequency dependent: at frequencies approaching DC the current is uniformly distributed in each electrode, and the magnetic field can pass through the electrode. Whereas at high frequencies the current distribution is determined by the skin effect and proximity effect: for the CLIC DR striplines the highest current density is on the electrode surface closest to the beam pipe. At high frequencies, the magnetic field hardly penetrates into the electrodes and thus the internal inductance of the electrodes is negligible. Hence, the characteristic impedance of the electrodes is frequency dependent (Fig. 37) [86].

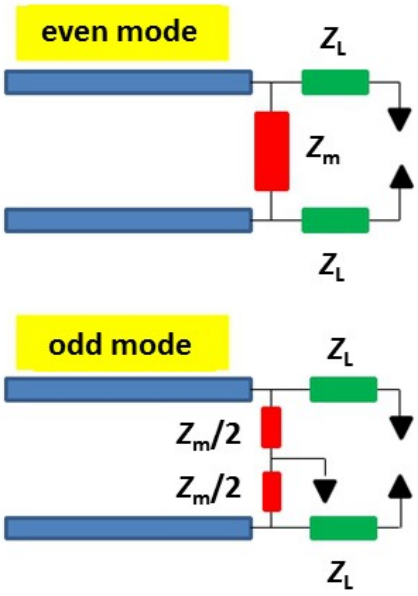


Fig. 36: Proposed method for matching the characteristic impedance for both odd and even modes: adding matching Z_m resistor allows for a tuning of the odd mode characteristic impedance while the even mode characteristic impedance is unchanged [117].

The electric field contribution to the deflection angle is constant at 0.75 mrad (taken from an electrostatic simulation), whereas the magnetic field contribution decreases from a value slightly larger than 0.76 mrad at 50 Hz, to a value of approximately 0.6 mrad, at higher frequencies (Fig. 38): the shape of the magnetic deflection curve, versus frequency, is similar to the shape of the curve shown in Fig. 37 [86].

The maximum field inhomogeneity allowed is ±0.01%, over 1 mm radius, although a radius of 0.5 mm has also been accepted from beam optics considerations [119]. The magnetic field uniformity for different frequencies, for the optimized stripline geometry [117], has been studied with *Opera2D*, and the results are shown in Fig. 39, for radii of 1 mm and 0.5 mm. When considering 1 mm radius, the field uniformity requirement is achieved for frequencies above ~500 kHz. If the good field region is reduced to 0.5 mm radius, the field uniformity requirement is achieved above 10 kHz [86].

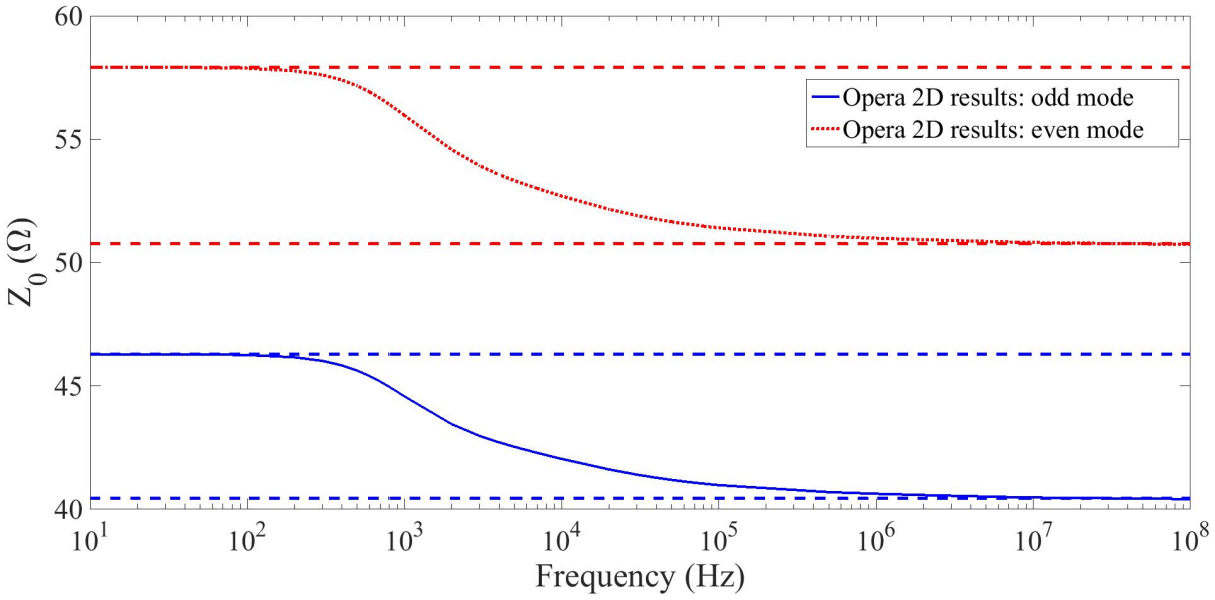


Fig. 37: Odd (blue) and even (red) mode characteristic impedances versus frequency. The dashed lines show the upper and lower values [86].

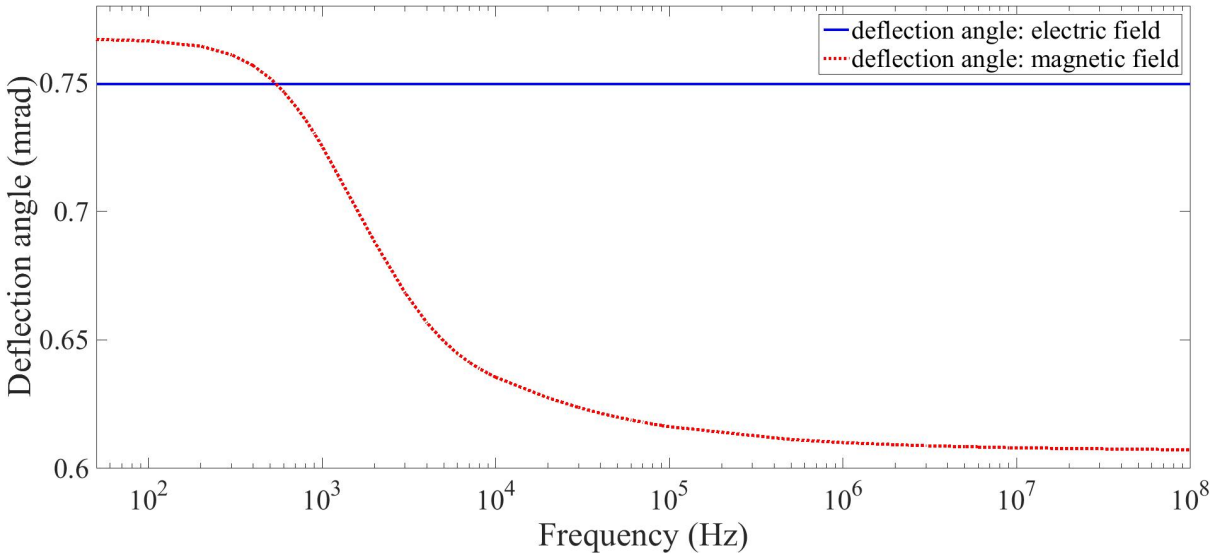


Fig. 38: Deflection angle resulting from the predicted electric and magnetic fields versus frequency, for the optimum (half-moon) geometry of the electrodes of the striplines (±12.5 kV, ±250 A) [86].

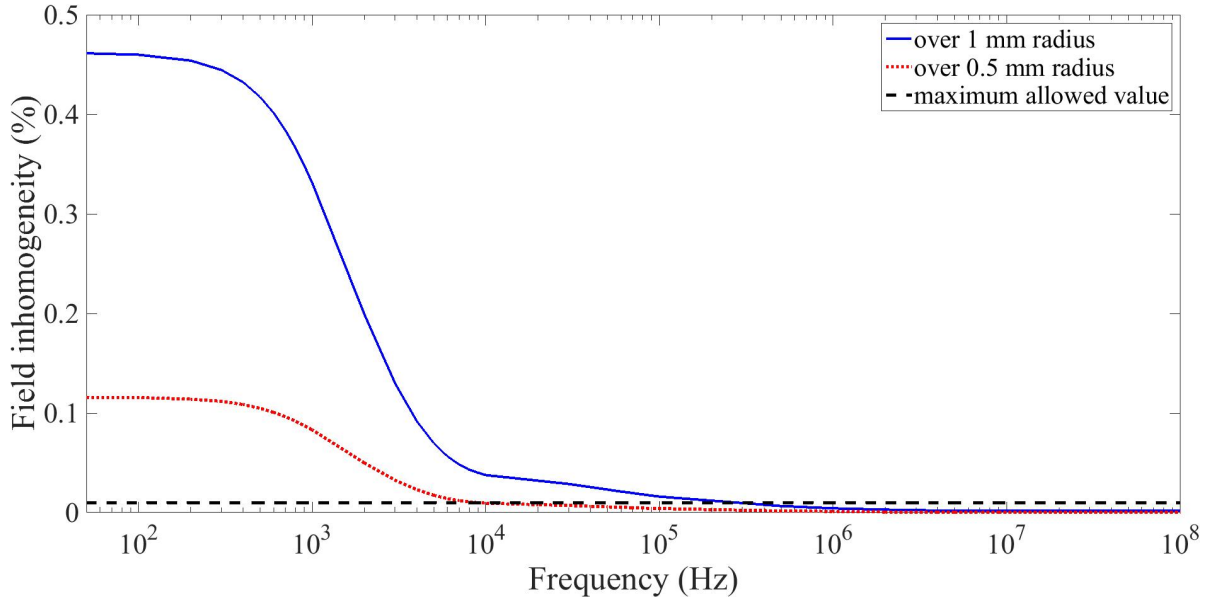


Fig. 39: Magnetic field inhomogeneity versus frequency, over 1 mm radius (blue) and 0.5 mm radius (red). The maximum allowed inhomogeneity, ±0.01%, is shown in black [86].

In order to achieve the required deflection angle of 0.75 mrad for the magnetic field, for the CLIC DR, given the relatively high-frequency content of a pulse, the odd mode characteristic impedance of the striplines must be matched by the termination resistance. It is possible to match the even mode impedance to 50 Ω and the odd mode impedance to 40.9 Ω, by connecting a resistance (Z_m) of 450 Ω between the output of the two electrodes, as shown in Fig. 36. However the output current to be supplied by the IA (see section 5.9.2.2.1) will be increased from 250 A to ~310 A**.**

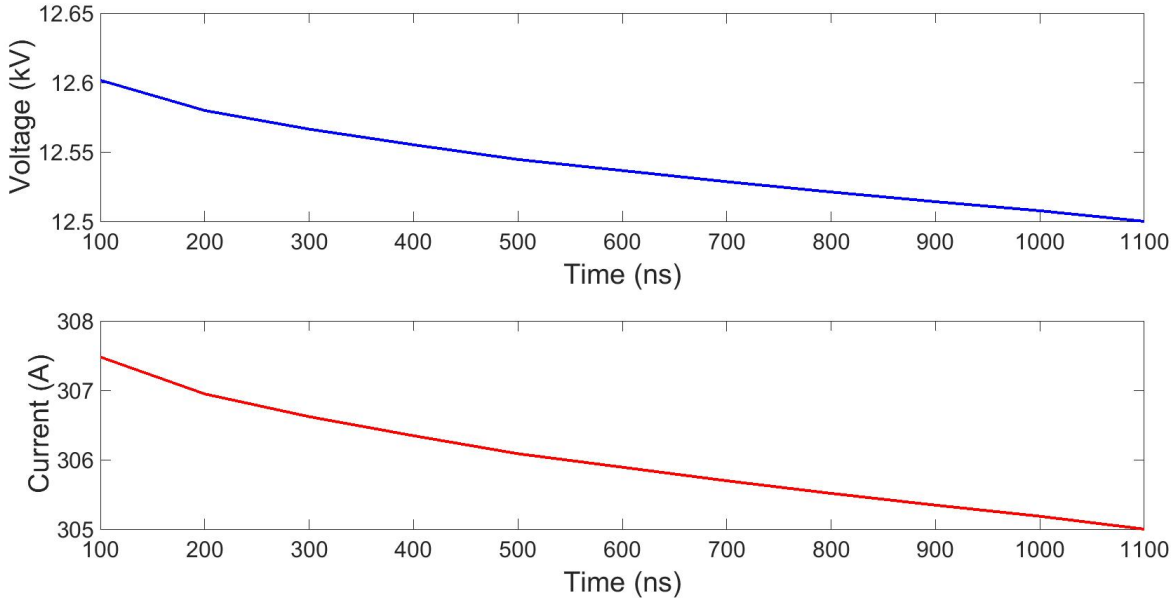


Fig. 40: Modulation of the flat-top voltage (top) and current (bottom), in order to obtain a 'constant' deflecting field (40.9 Ω termination).

As a result of the frequency dependence of the impedance and hence deflection angle of the striplines, and given the frequency content of the current and voltage waveform used to excite the striplines, in order to achieve a deflection angle flat to within ±0.02%, the 'flat top' of the required

excitation waveforms is not exactly flat, as shown in Fig. 40 [120]. It has been demonstrated that the prototype IA, developed for the CLIC DRs, can generate the required shape of the output waveform [92].

The prototype striplines have been installed at ALBA Cells, Spain, initially for beam impedance and DC field homogeneity measurements: these measurements will be carefully compared to predictions from simulations. Following the assembly and laboratory testing of the 12.5 kV prototype inductive adders, these may also be installed at ALBA for testing together with the striplines.

5.11 **Double kicker system**

A double kicker system (Fig. 41), consisting of two 'identical' ferrite loaded kicker magnets and a single power supply, has been developed at KEK [121, 122]. The first kicker extracts the beam from a damping ring and the second kicker, displaced from the first kicker by a suitable Betatron phase, results in anti-phase ripple to that of the first kicker (Fig. 42).

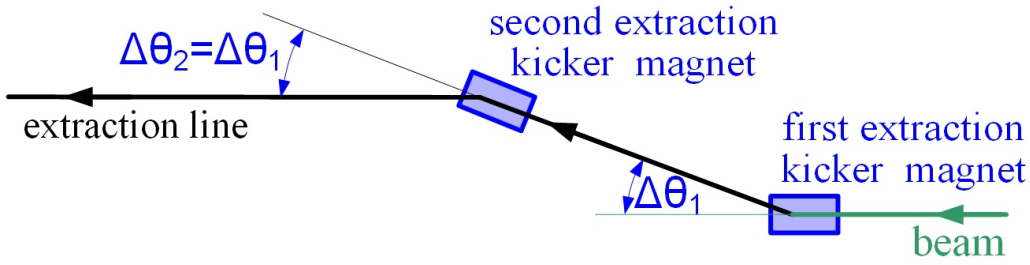


Fig. 41: First and second kickers separated by a betatron phase of $2n\pi$: for a betatron phase of $(2n-1)\pi$ the second kick would be towards the vertical direction.

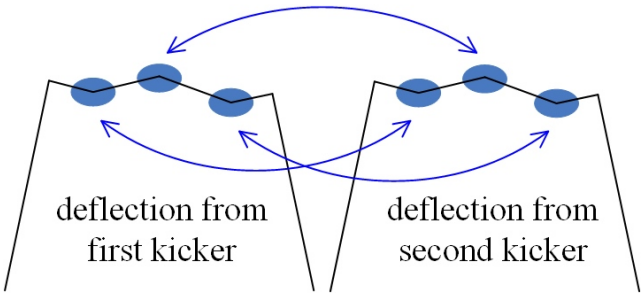


Fig. 42: Exactly the same flat top ripple from both kickers ideally results in ripple cancellation

Theoretically, using a double kicker system, the effect of ripple in the two kickers and small variations in the output of the HVDC supplies can completely cancel. In addition, theoretically, the double kicker can compensate for some field inhomogeneity in each kicker; this therefore permits the field uniformity requirement of an individual kicker to be relaxed, while still achieving excellent overall effective deflection uniformity. However, this places demanding requirements on the beam optics.

Measurements were carried out at the KEK/ATF with a double kicker system: the two kicker magnets were nominally identical, ferrite loaded, transmission-line magnets. However, to reduce beam-coupling impedance, the inside of the ceramic tube in the magnet apertures was coated with 1 µm thick titanium nitride (TiN): the thickness of the TiN coating is probably not uniform. This is evidenced by the fact that the second kicker had an apparent deflection angle of only 83% of the first kicker (for approximately the same current flow [121]): the deflection angle was determined from measurements of the beam orbit, shot by shot, in the extraction line [121]. The two KEK kickers each had 25 cells and were manufactured to high mechanical accuracy. In addition, the ferrite used (TDK-PE14) was from a single manufactured batch [123]: thus magnetically the two kickers should perform in a very similar manner.

The phase advance of the two kickers, although nominally π, was experimentally determined to minimise the jitter angle in the two kicker mode. Subsequently, to measure the equivalent jitter in single kicker mode, the second kicker was replaced with a dipole [122]: the strength of the dipole was adjusted to give the same average beam trajectory as for the double kicker. The dipole had a small aperture and a high stability power supply, thus its field uniformity and field stability was very good. All measurements were made with single bunches, i.e., at a particular time on the kick field waveform [122].

BPMs were used to determine the ripple of the double kicker system. The ATF damping ring had button BPMs with a positional resolution of between 20 μm and 30 μm. The extraction line used both Cavity BPMs (2 μm resolution) and stripline BPMs (20 μm to 30 μm resolution). The KEK double kicker achieved a factor of ~3.3 reduction in kick jitter angle, with respect to a single kicker [122]. The resolution of the BPMs limited the optimization of the phase advance and thus the jitter angle reduction.

5.12 **Ultra-fast kickers**

Ultra-fast kicker systems generally use either fast high-voltage MOSFETs [108, 124−127] or a Fast Ionization Dynistor (FID) [129–131] or a drift-step recovery diode [132]: the FID is also sometimes called a fast ionization device. Turn on of a thyristor by impact ionization is also an interesting means of achieving super-fast turn-on of a thyristor [133]. Where sub-nanosecond jitter is required, only semiconductor switches can be used. A case in point is the tail-clipper kicker for the CLIC Test Facility 3 (CTF3) [134] at CERN: the tail clipper must have a fast field rise time, of 5 ns or less, to minimize uncontrolled beam loss. For this application, there are eight identical pulse generators: each pulse generator is composed of a 50 Ω PFL, a fast semiconductor MOSFET switch, 50 Ω stripline electrodes (no magnetic material) and a matched terminating resistor [127, 128]. The deflection of the electron beam makes use of both the electric and magnetic field in the stripline electrodes: in order that the effects of the electric and magnetic field do not cancel, the striplines are fed from the beam exit end of the plates (Fig. 43).

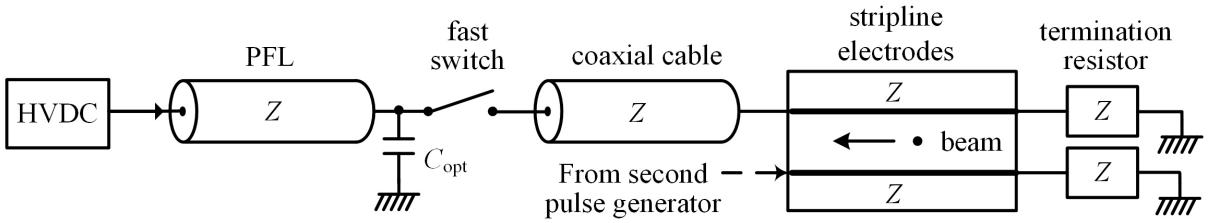


Fig. 43: Simplified schematic of CTF3 tail-clipper kicker

The overall length of the stripline assembly is mechanically sub-divided into four striplines of equal length of 380 mm each: a pair of stripline electrodes is energized a time delay of 1.27 ns (0.38 m/c) after the previous stripline electrodes to minimize the overall apparent rise time of the kick [127]. Each set of striplines is driven by two fast switches, one connected to a positively charged PFL and the other connected to a negatively charged PFL (Fig. 43). Figure 44 shows a pulse, measured using a fast current transformer, flowing in the load, and generated using a MOSFET switch: the 56 A of current has a 10% to 90% rise time of 2.5 ns. A novel design of gate driver results in an overall 3-sigma jitter of less than ±300 ps for the modulator system [127].

A FID was considered for use for the CTF3 tail clipper but a suitable device was not delivered in time.



Fig. 44: Measured trigger pulse (cyan trace – 1 V/div), gate driver output (green trace – 2 V/div), and load current (lilac trace – 10 A/div) for a CTF3 tail-clipper kicker system.

5.13 **Resistive terminators**

In order that the impedance of the resistive terminator is matched to the system impedance, over a wide range of frequencies, high-power resistor disks are generally used, which are housed in a coaxial structure to minimize inductance of the terminator. The coaxial housing is normally tapered, with the maximum internal diameter of housing at the input end [135]: this design permits the resistive terminator to withstand a high pulse voltage while minimizing the parasitic inductance of the terminator.

Current distribution within the resistor discs, during the pulse, is dependent upon the resistivity of the disks and the frequency content of the pulse. As a result of proximity effect, image current flows on the inside diameter of the coaxial housing. For cooling and insulation purposes, the terminator is generally filled with oil [18]. Long-term stability of the resistance value is linked to the ageing of the resistor discs: the ageing process is affected by the oil [68]. In order to accelerate the stabilization towards a final value of resistance, the discs are pre-impregnated under vacuum at high temperature. The short-term stability is linked to the temperature coefficient of the resistor material. Each current pulse will raise the temperature of the resistor stack: thus, where high stability is required, a suitable heat exchanger is required to maintain the oil temperature.

The resistor discs also exhibit a voltage coefficient of resistance, which is typically around −1.5%/kV/cm: thus the resistive terminator changes in value during the pulse. For high-precision applications, where a well-matched system is required, the voltage dependence of the terminator must be taken into account at the design stage.

5.14 **Acknowledgements**

There are many people who have given invaluable help and contributed to the preparation of the CAS lectures, and to whom I am very grateful. These include C. Belver Aguilar, A. Chmielinska, L. Ducimetière, T. Fowler, J. Holma, T. Kramer, L. Redondo, B. Salvant, V. Senaj, L. Sermeus, T. Stadlbauer, L. Vega Cid, V. Vlachodimitropoulos, and D. Woog. I thank all of these colleagues for their contributions, and apologize to anyone who I forgot to mention.

References

[1] G.H. Rees, in Proceedings of the CAS-CERN Accelerator School, 5th General Accelerator Physics Course, Jyväskylä, Finland, 7-18 September 1992, CERN 94-01 v 2, (CERN, Geneva, 1994), pp. 346-355, https://doi.org/10.5170/CERN-1994-001.

[2] C. Bovet, R. Gouiran, I. Gumowski and K.H. Reich, A selection of formulae and data useful for the design of A.G. synchrotrons, CERN/MPSSI/Int. DL/70/4.

[3] D. Fiander, K.D. Metzmacher and P.D. Pearce, Kickers and septa at the PS complex, CERN, Proc. KAON PDS Magnet Design Workshop, Vancouver, Canada, 3−5 October 1988, pp. 71–79.

[4] M.J. Barnes and G.D. Wait, Low power measurements on an AGS injection kicker magnet, Proc. 22nd Particle Accelerator Conference (PAC'07), Albuquerque, New Mexico, USA, 25−29 June 2007, pp. 2188–2190, https://doi.org/10.1109/PAC.2007.4441192.

[5] L. Ducimetière, Advances of transmission line kicker magnets, Proc. 21st Particle Accelerator Conference (PAC'05), Knoxville, USA, 16−20 May 2005, pp. 235–239, https://doi.org/10.1109/PAC.2005.1590404.

[6] J.N. Weaver *et al*., Design, analysis and measurement of very fast kicker magnets at SLAC, Proc. 13th Particle Accelerator Conference (PAC'89), Chicago, USA, 20−23 March 1989, pp. 411–413, https://doi.org/10.1109/PAC.1989.73062.

[7] W. Zhang *et al*., An overview of high voltage dielectric material for travelling wave kicker magnet application, Proc. 25th International Power Modulator Conference and High Voltage Workshop, Hollywood, California, USA, 30 June–3 July 2002, pp. 674–678.

[8] M.J. Barnes, G.D. Wait and I.M. Wilson, Comparison of field quality in lumped inductance versus transmission line kicker magnet, Proc. 4th European Particle Accelerator Conference (EPAC'94), London, UK, 27 June–1 July 1994, pp. 2547–2549.

[9] L. Ducimetière, N. Garrel, M.J. Barnes and G.D. Wait, The LHC injection kicker magnet, Proc. 20th Particle Accelerator Conference (PAC'03), Portland, Oregon, USA, 12−16 May 2003, pp. 1162–1164, https://doi.org/10.1109/PAC.2003.1289639.

[10] D.C. Fiander, K.D. Metzmacher and L. Sermeus, AA ejection kicker for AA Complex, CERN PS/BT/Note 87-22.

[11] K.D. Metzmacher, L. Sermeus, AA injection kicker for AA Complex, CERN PS/BT/Note 87-5.

[12] M.J. Barnes, F. Caspers, T. Kroyer, E. Metral, F. Roncarolo and B. Salvant, Measurement of longitudinal and transverse impedance of kicker magnets using the coaxial wire method, Proc. 23rd Particle Accelerator Conference (PAC'09), Vancouver, Canada, 4−8 May 2009, pp. 1726−1728.

[13] M.J. Barnes and G.D. Wait, Comparison of measured and predicted inductance per cell for a travelling wave kicker magnet, Proc. 5th European Particle Accelerator Conference (EPAC'96), Sitges, 10−14 Spain, June 1996, pp. 2588–2590.

[14] R.B. Armenta *et al*., Design concept for AGS injection kicker upgrade to 2 GeV, Proc. 21st Particle Accelerator Conference (PAC'05), Tennessee, USA, 16−20 May 2005, pp. 1380–1382, https://doi.org/10.1109/PAC.2005.1590768.

[15] R.B. Armenta *et al*., Electromagnetic modeling of the AGS A10 injection kicker magnet, *IEEE Trans. Appl. Supercond*.,**16** (2006), pp. 293–296, https://doi.org/10.1109/TASC.2005.864458.

[16] Cobham Technical Services, 24 Bankside Kidlington, Oxford OX5 1JE, UK, www.vectorfields.com, accessed 22 December 2017.

[17] M.J. Barnes *et al*., The beam screen for the LHC injection kicker magnets, Proc. 10th European Particle Accelerator Conference (EPAC'06), Edinburgh, Scotland, 26–30 June 2006, pp. 1508–1510.

[18] L. Ducimetière, U. Jansson, G.H. Schröder, E.B. Vossenberg, M.J. Barnes and G.D. Wait, Design of the injection kicker magnet system for CERN's 14TeV proton collider LHC, Proc. 10th International Pulsed Power Conference, Albuquerque, New Mexico, USA, 10–13 July 1995, pp. 1406–1411.

[19] M.J. Barnes, G.D. Wait and L. Ducimetière, Inductance calculations and measurements for the CERN LHC injection pulse forming network, Proc. 7th European Particle Accelerator Conference (EPAC'00), Vienna, Austria, 26–30 June 2000, pp. 2355–2357.

[20] Integrated Engineering Software, Winnipeg, Manitoba, Canada, www.integratedsoft.com, accessed 22 December 2017.

[21] ANSYS, http://www.ansys.com/, accessed 22 December 2017.

[22] CST - Computer Simulation Technology, http://www.cst.com, accessed 22 December 2017.

[23] T. Kroyer, F. Caspers and E. Gaxiola, Longitudinal and transverse wire measurements for the evaluation of impedance reduction measures on the MKE extraction kickers, CERN AB-Note-2007-028.

[24] F. Caspers *et al*., The fast extraction kicker system in SPS LSS6, Proc. 10th European Particle Accelerator Conference (EPAC'06), Edinburgh, Scotland, 26–30 June 2006, pp. 3125–3127.

[25] C. Zannini, Electromagnetic simulation of CERN accelerator components and experimental applications, PhD Thesis, École Polytechnique Fédérale de Lausanne, CERN-THESIS-2013-076 (available on-line).

[26] H. Day, PhD thesis, University of Manchester, 2019, CERN-THESIS-2013-083 (available on-line).

[27] M. Beck, Master thesis, Karlsruhe Institute of Technology, Germany, 2015, CERN-THESIS-2015-374 (available on-line).

[28] Y.H. Chin, K. Takata, T. Toyama, Y. Shobuda and J. Kamiya, Impedance and beam instability issues at J-PARC Rings, Proceedings of HB2008, Nashville, Tennessee, USA, 2008.

[29] J. Uythoven *et al*., Beam induced heating of the SPS fast pulsed magnets, Proc. 9th European Particle Accelerator Conference (EPAC'04), Lucerne, Switzerland, 5–9 July 2004, pp. 623–625.

[30] J. Rümmler, DESY, DESY III dump system with one fast kicker, Proc. 16th Particle Accelerator Conference (PAC'95), Dallas, Texas, USA, 1–5 May 1995, pp. 1933–1935, https://doi.org/10.1109/PAC.1995.505411.

[31] B. Salvant *et al*., Beam induced RF heating in LHC in 2015, Proc. 7th International Particle Accelerator Conf. (IPAC'16), Busan, Korea, 8−13 May 2016, pp. 602−605.

[32] M.J. Barnes *et al*., Beam induced ferrite heating of the LHC injection kickers and proposals for improved cooling, Proc. 4th International Particle Accelerator Conf. (IPAC'13), Shanghai, 12−17 May 2013, pp. 732−734.

[33] W. Zhang, L.A. Ahrens, J. Glenn, J. Sandberg and N. Tsoupas, Beam coupling phenomena in fast kicker systems, Proc. 19th Particle Accelerator Conference (PAC'01), Chicago, USA, 18–22 June 2001, pp. 3708–3710, https://doi.org/10.1109/PAC.2001.988227.

[34] H. Tsutsui, Some simplified models of ferrite kicker magnet for calculation of longitudinal coupling impedance, CERN-SL 2000-004 AP, 2000.

[35] G. Nassibian, CERN/PS 84-25 (BR), 1984; CERN/PS 85-68 (BR), 1986.

[36] Y. Shobuda, Y. Irie and T. Toyama, Analytical approach to evaluate coupling impedances of traveling kicker magnets, *Nuclear Instrum. Methods in Phys. Res. A* **691** (2012) 135–151, https://doi.org/10.1016/j.nima.2012.06.066.

[37] N. Biancacci *et al*., Impedance calculations for simple models of kickers in the non-ultrarelativistic regime, Proc. 2nd International Particle Accelerator Conf. (IPAC'11), San Sebastián, Spain, 4−9 September 2011, pp. 772-774.

[38] F. Caspers, Impedance Determination, in: Handbook of Accelerator Physics and Engineering (World Scientific, 1999).

[39] M.J. Barnes *et al*., Measurement and analysis of SPS kicker magnet heating and outgassing with different bunch spacing, Proc. 23rd Particle Accelerator Conference (PAC'09), Vancouver, Canada, 4−8 May 2009, pp. 4264−4266.

[40] M.J. Barnes, F. Caspers, L. Ducimetière, N. Garrel and T. Kroyer, An improved beam screen for the LHC injection kickers, Proc. 22nd Particle Accelerator Conference (PAC'07), Albuquerque, New Mexico, USA, 25−29 June 2007, pp. 1574–1576, https://doi.org/10.1109/PAC.2007.4440827.

[41] J. Rümmler, Kicker Types for in- and ejection in the damping rings of TESLA, DESY/MIN December 2000.

[42] F. Caspers, A. Mostacci and H. Tsutsui, Impedance evaluation of the SPS MKE kicker with transition pieces between tank and kicker module, CERNSL-2000-071 (AP).

[43] H. Hahn, A. Dunbar, C.I. Pai, R.T. Sanders, N. Tsoupas and J.E. Tuorzzolo, The RHIC beam abort kicker system, Proc. 18th Particle Accelerator Conference (PAC'99), New York, USA, 29 March–2 April 1999, pp. 1100–1102, https://doi.org/10.1109/PAC.1999.795461.

[44] H.J. Tran, M.J. Barnes, G.D. Wait and Y. Yan, Longitudinal impedance of a prototype kicker magnet system, Proc. 15th Particle Accelerator Conference (PAC'93), Washington, D.C, USA, 17−20 May 1993, pp. 3502−3404, https://doi.org/10.1109/PAC.1993.309664.

[45] P. He, H.C. Hseuh and R.J. Todd, Metallization of ceramic vacuum chambers for SNS ring injection kicker magnets, *Thin Solid Films* **420–421** (2002) 38–42, https://doi.org/10.1016/S0040-6090(02)00661-2.

[46] M. Pont, R. Nunez and E. Huttel, Septum and kicker magnets for the alba booster and storage ring, Proc. 2nd International Particle Accelerator Conf. (IPAC'11), San Sebastián, Spain, 4−9 September 2011, pp.4421−4423.

[47] A.D. Ghodke, D. Angal-Kalinin and G. Singh, Kicker chamber metallization induced effects on the beam in Indus-1, Proc. 2nd Asian Particle Accelerator Conf. (APAC'01), Beijing, China, 17−21 September 2001.

[48] M.J. Barnes, T. Fowler, M.G. Atanasov, T. Kramer and T. Stadlbauer, Effect of a metallized chamber upon the field response of a kicker magnet: simulation results and analytical calculations, Proc. 3rd International Particle Accelerator Conf. (IPAC'12), New Orleans, USA, 20−25 May 2012, pp. 3686−3688.

[49] A. Adraktas, M.J. Barnes, L. Ducimetière, Influence of conducting serigraphy upon field pulse shape of the SPS extraction kicker systems, Proc. 8th International Particle Accelerator Conf. (IPAC'17), Copenhagen, Denmark, 14−19 May 2017, pp. 3491−3494.

[50] M.J. Barnes *et al*., Studies of impedance-related improvements of the SPS injection kicker system, Proc. 7th International Particle Accelerator Conf. (IPAC'16), Busan, Korea, 8−13 May 2016, pp. 3611−3614.

[51] Y. Shobuda, Y. Irie, T. Toyama, J. Kamiya and M. Watanabe, Measurement scheme of kicker impedances via beam-induced voltages of coaxial cables, *Nucl. Instrum. Methods Phys. Res. A* **713** (2013) 52–70, https://doi.org/10.1016/j.nima.2013.02.037.

[52] V. Vlachodimitropoulos, M.J. Barnes, L. Ducimetière, L.V. Cid and W. Weterings, Predicted beam induced power deposition in the LHC injection kicker magnets for HL-LHC Type Beams,

Proc. 8th International Particle Accelerator Conf. (IPAC'17), Copenhagen, Denmark, 14−19 May 2017, pp. 3471−3474.

[53] H. Day *et al*., Evaluation of the beam coupling impedance of new beam screen designs for the LHC injection kicker magnets, Proc. 4th International Particle Accelerator Conference (IPAC13), Shanghai, China, 12−17 May 2013, pp. 1649−1651.

[54] M.J. Barnes *et al*., High voltage performance of the beam screen of the LHC injection kicker magnets, IPAC'14, Dresden, 15−20 June 2014, pp. 541−543.

[55] M.J. Barnes *et al*., Operational experience of the upgraded LHC injection Kicker Magnets, Proc. 7th International Particle Accelerator Conf. (IPAC'16), Busan, Korea, 8−13 May 2016, pp. 3623−3266.

[56] H. Day, M.J. Barnes, L. Feliciano, Impedance studies of the LHC injection kicker magnets for HL-LHC, Proc. 6th International Particle Accelerator Conf. (IPAC'15), Virginia, USA, May 3-8, 2015, MOPJE038, pp. 370-373.

[57] D. Alesini, S. Guiducci, F. Marcellini and P. Raimondi, Design, test, and operation of new tapered stripline injection kickers for the e^+ e^- collider DAΦNE, *Phys. Rev. ST Accel. Beams*. **13** (2010) 111002, https://doi.org/10.1103/PhysRevSTAB.13.111002.

[58] Z. Sobiech, Master thesis, UTP University of Science and Technology, Faculty of Mechanical Engineering, Bydgoszcz, Poland, 2014.

[59] L. Vega Cid, Master Thesis, Universidad del Pais Vasco, Bilbao, Spain, 2015 (in Spanish).

[60] M. Timmins, A. Bertarelli, J. Uythoven and E. Gaxiola, SPS extraction kicker magnet cooling design, AB-Note-2004-005 BT.

[61] L. Vega *et al*., Thermal analysis of the LHC injection kicker magnets, *J. Phys. Conf. Ser*., **874** (2017) 012100, https://doi.org/10.1088/1742-6596/874/1/012100.

[62] M.J. Barnes, L. Ducimetière, N. Garrel, B. Goddard, V. Mertens and W. Weterings, Analysis of ferrite heating of the LHC injection kickers and proposals for future reduction of temperature, Proc. 3rd International Particle Accelerator Conf. (IPAC'12), New Orleans, 21−26 May 21-26, pp. 2038−2040.

[63] M.J. Barnes *et al*., Operational experience of the upgraded LHC injection kicker magnets during Run 2 and future plans, *J. Phys. Conf. Ser*., **874** (2017) 012101, https://doi.org/10.1088/1742-6596/874/1/012101.

[64] T.S. Sudarshan and J.D. Cross, The effect of chromium oxide coatings on surface flashover of alumina spacers in vacuum, *IEEE Trans. Electr. Insul*., **EI-1I** (1976) 32.

[65] T. Shioiri, N. Asari, S. Saito, H. Nakamuta, M. Homma and K. Suzuki, Effect of chromium oxide coating on surface flashover characteristics of ceramic in vacuum, XXIInd Int. Symp. on Discharges and Electrical Insulation in Vacuum-Matsue-2006, pp. 140−143, https://doi.org/10.1109/DEIV.2006.357251.

[66] G. Arduini *et al*., Performance of the CERN SPS fast extraction for the CNGS facility, Proc. 21st Particle Accelerator Conference (PAC'05), Knoxville, USA, 16−20 May 2005, pp. 1757–1759.

[67] B. Goddard, M. Barnes, L. Ducimetière, W. Höfle and G. Kotzian, Emittance growth at LHC injection from SPS and LHC kicker ripple, Proc. 11th European Particle Accelerator Conference (EPAC'08), Genoa, Italy, 23−27 June 2008, pp. 3629–3631.

[68] M.J. Barnes, G.D. Wait and L. Ducimetière, Low voltage measurements on nine PFNs for the LHC injection kicker systems, Proc. 8th European Particle Accelerator Conference (EPAC'02), Paris, France, 3–8 June 2002, pp. 2520–2522.

[69] E2V Technologies, 'Hydrogen Thyratrons Preamble', https://www.tayloredge.com/reference/Electronics/VacuumTube/thyratron_preamble.pdf, accessed 22 December 2017.

[70] L. Ducimetiere and D.C. Fiander, Commutation losses of a multigap high voltage thyratron, IEEE Conference Record of the 1990 Nineteenth Power Modulator Symposium, San Diego, USA, 26−28 June 1990, pp. 248−253.

[71] R.L. Cassel and T.S. Mattison, Kicker prepulse canceller, IEEE Conference Record of the 1991 Particle Accelerator Conference, 6−9 May 1991, pp. 3162−3164.

[72] M.J. Barnes and G.D. Wait, A mathematical model of a three-gap thyratron simulating turn-on, Proc. 9th International Pulse Power Conf., Albuquerque, USA, 21−23 June 1993, pp. 293−296, https://doi.org/10.1109/PPC.1993.513335.

[73] G.D. Wait, M.J. Barnes, K.D. Metzmacher and L. Sermeus, The application of saturating inductors for improving the performance of the CERN PS kicker systems, Proc. 17th Particle Accelerator Conf. (PAC'97), Vancouver, Canada, 12−16 May 1997, pp. 1328−1330.

[74] M.J. Barnes, G.D. Wait, E. Carlier, L. Ducimetière, G.H. Schröder and E.B. Vossenberg, A fast 60 kV resonant charging power supply for the LHC inflectors, Proc. 17th Particle Accelerator Conference (PAC'97), Vancouver, Canada, 12–16 May 1997, pp. 1325–1327.

[75] M.J. Barnes, G.D. Wait, L. Ducimetière, G.H. Schröder and E.B. Vossenberg, Measurements on a fast 66 kV resonant charging power supply for the LHC inflectors, Proc. 6th European Particle Accelerator Conference (EPAC'98), Stockholm, Sweden, 22–26 June 1998, pp. 2278–2280.

[76] J.A. Uythoven, E.H.R. Gaxiola and M.A. Timmins, Upgrade of the SPS extraction kickers for LHC and CNGS operation, Proc. 8th European Particle Accelerator Conf. (EPAC'02), Paris, France, 3−8 June pp. 2517−2519.

[77] V. Senaj, N. Voumard, M.J. Barnes and L. Ducimetière, Optically isolated circuit for failure detection of a switch in an HV series connected stack, Proc. 17th IEEE International Pulsed Power Conference, Washington DC, USA, 29 June–2 July 2009, https://doi.org/10.1109/PPC.2009.5386389.

[78] C. Boatella Polo, SEE Single Event Effects, ESA-CERN-SCC Workshop, CERN, 09−10 May 2017, https://indico.cern.ch/event/635099/contributions/2570672/attachments/1456364/2249943/Single_Event_Effecs_Radiation_Course_May_2017_SEE_CBP.pdf

[79] H.R. Zeller, Cosmic ray induced failures in high power semiconductor devices, *Microelectron. Reliab.*, **37** (1997) 1711, https://doi.org/10.1016/S0026-2714(97)00146-7.

[80] R. Edwards and J. Woodhouse, Determination of high energy neutron voltage stress margins for high voltage IGBT and diode pairs from two manufacturers using energetic particle induced charge spectroscopy, EPICS, IEEE Xplore, Radiation Effects Data Workshop, 2006 IEEE, 17−21 July 2006, https://doi.org/10.1109/REDW.2006.295488.

[81] E. Normand, J.L. Wert, D.L. Oberg, P.P. Majewski, P. Voss and S.A. Wender, Neutron-induced single event burnout in high voltage electronics, *IEEE Trans. Nucl. Sci.*, **44** (1997) 2358, https://doi.org/10.1109/23.659062.

[82] J. Bonthond, J.H. Dieperink, L. Ducimetière, U. Jansson and E. Vossenberg, Dual branch high voltage pulse generator for the beam extraction of the Large Hadron Collider, Proc. 2002 Power Modulator Symposium, Hollywood, USA, 30 June–3 July 2002, pp. 114–117, https://doi.org/10.1109/MODSYM.2002.1189429.

[83] J. Uythoven *et al.*, Calibration measurements of the LHC beam dumping system extraction kicker magnets, Proc. 10th European Particle Accelerator Conference (EPAC'06), Edinburgh, Scotland, 26–30 Jun 2006, pp. 1520−1522.

[84] M.A. Fraser *et al.*, Considerations on an upgrade possibility of the LHC beam dump kicker system, Proc. 7th International Particle Accelerator Conf. (IPAC'16), Busan, Korea, 8−13 May 2016, pp. 3631−3633.

[85] M. Aicheler (ed.) *et al.*, A multi-TeV linear collider based on CLIC technology: CLIC Conceptual Design Report, CERN-2012-007, https://doi.org/10.5170/CERN-2012-007.

[86] C. Belver-Aguilar, M.J. Barnes and L. Ducimetière, Review on the effects of characteristic impedance mismatching in a stripline kicker, Proc. IPAC'16, Busan, Korea, 8−13 May 2016, pp. 3627−3630.

[87] Y. Papaphilippou *et al*., Conceptual design of the CLIC damping rings, Proc. IPAC'12, New Orleans, USA, May 2012, pp. 1368−1370.

[88] J. Holma *et al*., Preliminary design of the pulse generator for the CLIC DR extraction system, Proc. PPC'11, Chicago, USA, 19−23 June 2011, pp. 1353−1358, https://doi.org/10.1109/PPC.2011.6191614.

[89] J. Holma, Ph.D. Thesis, Aalto University publication series, Doctoral Dissertations 196/2015, Helsinki, Finland, 2015. CERN-THESIS-2014-359 (available on-line).

[90] K. Takayama (ed.) and Briggs, R.J. (ed.), Induction Accelerators. Springer, 2011. ISBN: 978-3-642-13916-1.

[91] E.G. Cook *et al*., Solid-state modulator R&D at LLNL, Int. Workshop on Recent Progress of Induction Accelerators, Tsukuba, Japan, 29−31 October 2002.

[92] J. Holma and M.J. Barnes, Prototype inductive adders with extremely flat-top output pulses for the Compact Linear Collider at CERN, paper submitted to a Special Issue of Transactions on Plasma Science on Pulsed Power Science and Technology, 2018.

[93] M. Kemp, High rep-rate CLIC combiner ring kicker, ALERT 2014 Workshop, IFIC-Valencia, Spain, 5−6 May 2014, https://indico.cern.ch/event/306551/contributions/705065/attachments/583693/803464/20_ALERT2014_Kemp.pdf

[94] D. Woog, M.J. Barnes, L. Ducimetière, J. Holma and T. Kramer, Design of an inductive adder for the FCC injection kicker pulse generator, Proc. 8th International Particle Accelerator Conf. (IPAC'17), Copenhagen, Denmark, 14−19 May 2017, WEPVA024, pp. 3312−3315, https://doi.org/10.1088/1742-6596/874/1/012096.

[95] A. Chmielinska, M.J. Barnes, W. Bartmann, F. Burkart and B. Goddard, Preliminary estimate of beam induced power deposition in a FCC-hh injection kicker magnet, Proc. 8th International Particle Accelerator Conf. (IPAC'17), Copenhagen, Denmark, 14−19 May 2017, pp. 3475−3478.

[96] K.-D. Metzmacher and L. Sermeus, The PS injection kicker KFA45 performance for LHC, PS/PO Note 2002-015, CERN, Geneva, 2002.

[97] M.J. Barnes, L. Ducimetière, J. Holma, T. Kramer and A. Fowler, Inductive adders for replacing thyratron based modulators at CERN, Published in: Power Electronics and Applications (EPE'15 ECCE-Europe), 2015 17th European Conference, 8−10 September 2015.

[98] W.L. Willis, in *Pulse-Voltage Circuits*, 1st ed. Eds. R.E. Dollinger and J.W. Sarjeant (Tab Books Inc., 1989), Chap. 3, ISBN 0-8306-9094-8.

[99] J.A. Casey, F.O. Arntz, M.P.J. Gaudreau and M. Kempes, Solid-State Marx bank modulator for the Next Linear Collider, Pulsed Power Conference, 15−18 June 2003, pp. 641−644, https://doi.org/10.1109/PPC.2003.1277791.

[100] A. Krasnykh, R. Akre, S. Gold and R. Koontz, A Solid-state Marx type modulator for driving a TWT, Power Modulator Symposium, 2000. Conference Record of the 2000 Twenty-Fourth International, Norfolk, VA, USA, 26−29 June 2000.

[101] R. Phillips, M.P.J. Gaudreau, M. Kempkes, K. Ostlund and J. Casey, Affordable, short pulse Marx modulator, IEEE International Vacuum Electronics Conference, 22−24 April 2014, Monterey, CA, pp. 455−456.

[102] S.C. Glidden and H.D. Sanders, Solid state Marx generator, Power Modulator Conference, Arlington, VA, USA, 14−18 May 2006, Arlington, https://doi.org/10.1109/MODSYM.2006.365246.

[103] T. Beukers, C. Burkhart, M. Kemp, R. Larsen, M. Nguyen J. Olsen and T. Tang., P1-Marx modulator for the ILC, IEEE Power Modulator and High Voltage Conference, 23-27 May 2010, pp. 21−22.

[104] M.A. Kemp, A. Benwell, C. Burkhart, D. MacNair and M. Nguyen, The SLAC P2 Marx, IEEE Power Modulator and High Voltage Conference, 2012, pp. 1721−1728.

[105] L.M. Redondo and J. Fernando Silva, *IEEE Trans. Plasma Sci.*, **37** (2009) 1632. https://doi.org/10.1109/TPS.2009.2023221.

[106] L.M. Redondo, A. Kandratsyeu, M.J. Barnes, S. Calatroni and W. Wuensch, Solid-state Marx generator for CLIC breakdown studies, to be published in Proc. 2016 IEEE Power Modulator and High Voltage Conference, San Francisco, CA, USA, 5−9 July 2016, https://doi.org/10.1109/IPMHVC.2016.8012824.

[107] L.M. Redondo, A. Kandratsyeu and M.J. Barnes, Marx generator prototype for kicker magnets based on SiC MOSFETs, IEEE Transactions on Plasma Science, 2018, DOI: 10.1109/TPS.2018.2808194.

[108] M.J. Barnes and G.D. Wait, A 25 kV, 75 kHz, Kicker for measurement of muon lifetime, IEEE Trans. Plasma Sci., **32** (2004) 1932, https://doi.org/10.1109/TPS.2004.835455.

[109] C. Figley and D.A. Stensrud, Stacked FET switch for use in a 20 kV klystron modulator. Digest of Tech. Papers PPC'91, San Diego, CA, USA, 16−19 June 1991, pp. 1001−1004, https://doi.org/10.1109/PPC.1991.733455.

[110] G.D. Wait and M.J. Barnes, Pulse width control at 10^6 Pulses/sec and 15 kV for the KAON factory beam chopper, Proc. 9th International Pulse Power Conf., Albuquerque, USA, 21−23 June 1993, pp. 841−844.

[111] M.J. Barnes and G.D. Wait, A FET based kicker for a charge booster for the TRIUMF ISAC Project, Proc. 13th IEEE International Pulsed Power Conf., Las Vegas, USA, 17−22 June 2001, pp. 1245−1248.

[112] R.B. Armenta, M.J. Barnes and G.D. Wait, Precision 25 kV pulse and RF suppression in a 75 kHz kicker, Proc. 15th IEEE International Pulsed Power Conf., Monterey, USA, 13−17 June 2005, pp. 676−679, https://doi.org/10.1109/PPC.2005.300750.

[113] W. Jiang *et al.*, MHz Pulsed Power Generator Using MOSFET. Conf. Rec. Power Modulator Symposium and High Voltage Workshop, Hollywood, CA, USA, 30 June 30−3 July 2002, pp. 599−601, https://doi.org/10.1109/MODSYM.2002.1189550.

[114] M.J. Barnes and G.D. Wait, Design and measurements of a damping ring kicker for the ILC, Proc. 22nd Particle Accelerator Conf. (PAC'07), 25−29 June 2007, pp. 846−848, https://doi.org/10.1109/PAC.2007.4440735.

[115] I. Rodríguez, Calculation methodology and fabrication procedures for particle accelerator strip-line kickers: application to the CTF3 combiner ring extraction kicker and TL2 tail clippers, Ph.D. thesis, Universidad Politécnica de Madrid, 2009.

[116] C. Belver-Aguilar, A. Faus-Golfe, F. Toral and M.J. Barnes, Stripline design for the extraction kicker of compact linear collider damping rings, *Phys. Rev. ST: Accel. Beams* **17** (2014) 071003, https://doi.org/10.1103/PhysRevSTAB.17.071003.

[117] C. Belver-Aguilar, PhD thesis, Universidad de Valencia, Spain, 2015. CERN-THESIS-2015-175 (available on-line).

[118] T. Naito, S. Araki, H. Hayano, K. Kubo, S. Kuroda, N. Terunuma and J. Urakawa, *Phys. Rev. ST Accel. Beams* **14** (2011) 051002, https://doi.org/10.1103/PhysRevSTAB.14.051002.

[119] R. Apsimon *et al.*, Optics and protection of the injection and extraction regions of the CLIC damping rings, Proceedings of the 4th International Particle Accelerator Conference, IPAC 2013, MOPWO025, pp. 939−941.

[120] C. Belver-Aguilar and M.J. Barnes, Transient studies of the stripline kicker for beam extraction from CLIC damping rings, International Beam Instrumentation Conference (IBIC), Barcelona, Spain, 11−15 September 2016.

[121] T. Imai *et al*, Double Kicker system in ATF, XX International Linac Conference, Monterey, California, LINAC2000, pp. 77−79.

[122] T. Imai, Thesis, Development of Double Kicker System for Stable Beam Extraction, Dept. of Physics, Tokyo University of Science, September 2002.

[123] T. Naito, KEK, private communication.

[124] M.J. Barnes *et al*., A high frequency MOSFET driver for the TITAN facility at TRIUMF, Proc. 15th IEEE International Pulsed Power Conference, Monterey, USA, 13−17 June 2005, pp. 178−181, https://doi.org/10.1109/PPC.2005.300554.

[125] F. Arntz, M. Gaudreau, M. Kempkes, A. Krasnykh and A. Kardo-Sysoev, A kicker driver for the international linear collider, Proc. 22nd Particle Accelerator Conference (PAC'07), Albuquerque, New Mexico, USA, 25–29 June 2007, pp. 2972–2974, https://doi.org/10.1109/PAC.2007.4440638.

[126] F. Arntz, J. Casey, M. Kempkes, N. Butler and M. Gaudreau, Solid-state high voltage pulse power in the 10-100 nanosecond regime, Proc. 10th European Particle Accelerator Conference (EPAC'06), Edinburgh, Scotland, 26–30 June 2006, pp. 3134–3136.

[127] M.J. Barnes, T. Fowler and G. Ravida, Design, Testing and operation of the modulator for the CTF3 tail clipper kicker, Proc. 23rd Particle Accelerator Conference (PAC'09), Vancouver, Canada, 4−8 May 2009, pp. 1723−1725.

[128] M.J. Barnes, T. Fowler, G. Ravida and A. Ueda, Design of the modulator for the CTF3 tail clipper kicker, Proc. 22nd Particle Accelerator Conference (PAC'07), Albuquerque, New Mexico, USA, 25−29 June 2007, pp. 2185–2187.

[129] A. Wolski, Critical R&D issues for the ILC damping rings and new test facilities, Proc. 22nd Particle Accelerator Conference (PAC'07), Albuquerque, New Mexico, USA, 25–29 June 2007, pp. 1945–1949.

[130] T. Naito *et al*., Development of strip-line kicker system for ILC damping ring, Proc. 22nd Particle Accelerator Conference (PAC'07), Albuquerque, New Mexico, USA, 25–29 June 2007, pp. 2772–2774.

[131] D. Alesini, S. Guiducci, F. Marcellini and P. Raimondi, Fast injection kickers for Daphne collider and ILC damping rings, DAPHNE Technical Note, INFN - LNF, Accelerator Division. Note I-17, 6 June 2006.

[132] A. Krasnykh, Development of a Fast High-Power Pulser and ILC DR Injection/Extraction Kicker, SLAC-WP-077, https://inis.iaea.org/search/search.aspx?orig_q=RN:39104667.

[133] A. Gusev, S. Lyubutin, A. Ponomarev, S. Rukin, B. Slovikovsky, S. Tsyranov, High current and current rise rate thyristor based switches, Proc. 21st International Pulsed Power (PPC), Brighton, England, 18-22 June 2017, DOI: 10.1109/PPC.2017.8291162.

[134] G. Geschonke and A. Ghigo, CTF3 design report, CERN/PS 2002-008 (RF).

[135] M.J. Barnes and G.D. Wait, Optimization of speed-up network component values for the 30 Ω resistively terminated prototype kicker magnet, Proc. 15th Particle Accelerator Conference (PAC'93), Washington DC, USA, 17–20 May 1993, pp. 1330–1332.

Bibliography

Eds. O. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole and P. Proudlock, *LHC Design Report*, volume I, The LHC main ring, Chap. 16, Injection system, http://ab-div.web.cern.ch/ab-div/Publications/LHC-DesignReport.html, CERN, Geneva, 2004, CERN-2004-003.

Eds. O. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole and P. Proudlock, LHC Design Report, volume I, The LHC main ring, Chap. 17, Beam dumping system, http://ab-div.web.cern.ch/ab-div/Publications/LHC-DesignReport.html, CERN, Geneva, 2004, CERN-2004-003.

Eds. M. Benedikt, P. Collier, V. Mertens, J. Poole and K. Schindl, LHC Design Report, volume III, The LHC injector chain, http://ab-div.web.cern.ch/ab-div/Publications/LHC-DesignReport.html, CERN, Geneva, 2004, CERN-2004-003.

P. Smith, *Transient Electronics: Pulsed Circuit Technology* (John Wiley & Sons, 2002), ISBN: 047197773X.

L. Ducimetière and D. Fiander, Commutation losses of a multigap high voltage thyratron, Proc. Power Modulator Symposium, San Diego, USA, 26–28 June 1990, pp. 248–253.

K.D. Metzmacher and L. Sermeus, The PSB injection and recombination kicker systems for LHC, CERNPS/CA/Note 2000-004.

K.D. Metzmacher and L. Sermeus, The PS injection kicker KFA45 performance for LHC, CERNPS/PO/Note 2002-015 (Tech.).

L.I. Donley, J.C. Dooling and G.E. McMichael, Using ferrite as a fast switch for improving rise time of IPNS extraction kicker, Proc. 20th Particle Accelerator Conference (PAC'03), Portland, Oregon, USA, 12–16 May 2003, pp. 1159–1161.

W. Zhang *et al*., Design, development, and construction of SNS extraction fast kicker system, Proc. 20th Particle Accelerator Conference (PAC'03), Portland, Oregon, USA, 12–16 May 2003, pp. 550–552.

G. Nassibian, Travelling Wave kicker magnets with sharp rise and less overshoot, Proc. 8th Particle Accelerator Conference (PAC'79)*, IEEE Trans. Nucl. Sci.,* **NS-26** pp. (1979) 4018.

M.J. Barnes, G.D. Wait and R. Barnes, Estimation of negative mutual coupling between adjacent cells of a magnetic kicker, TRIUMF Design Note TRI-DN-89-K87.

C. Jensen, B, Hanna and R. Reilly, A fast injection kicker magnet for the Tevatron, Proc. 19th Particle Accelerator Conference (PAC'01), Chicago, USA, 18–22 June 2001, pp. 3720–3722.

T. Kawakubo, E. Nakamura and S. Murasugi, Low beam-coupling impedance kicker magnet system generating a high magnetic field with a fast rise time, Proc. 7th European Particle Accelerator Conference (EPAC'00), Vienna, Austria, 26–30 june 2000, pp. 2246–2248.

P.E. Faugeras, E. Frick, C.G. Harrison, H. Kuhn, V. Rodel, G.H. Schröder, J.P. Zanasco, The SPS fast pulsed magnet systems, Proc. 12th Modulator Symposium, New York, USA, Feb. 4–5, 1976 (CERN-SPS-BT-76-1).

J. Wang, Design and test results of kicker units for the positron accumulator ring at the APS, Proc. 16th Particle Accelerator Conference (PAC'95), Dallas, Texas, USA, 1–5 May 1995, pp. 1245–1247.

M. Giovannozzi *et al*., Design and tests of a low-loss multi-turn ejection for the CERN PS, 39th ICFA Advanced Beam Dynamics Workshop: High Intensity High Brightness Hadron Beams (HB2006), Tsukuba, Japan, 29 May–2 June 2006, pp. 192–196.

M.J. Barnes, E.W. Blackmore, G.D. Wait, J. Lemire-Elmore (UBC), B. Rablah, G. Leyh, M. Nguyen and C. Pappas, Analysis of high power IGBT short circuit failures, *IEEE Trans. Plasma Sci*., **33** (2005) 1252.

T.S. Mattison, R.L. Cassel, A.R. Donaldson and G.Gross, Fast and reliable kicker magnets for the SLC damping rings, Proc. 16th Particle Accelerator Conference (PAC'95), Dallas, Texas, USA, 1–5 May 1995, pp. 1915–1917.

D. Qunell, C. Jensen, D. Tinsley, Kicker system for 8 GeV proton injection, Proc. 17th Particle Accelerator Conference (PAC'97), Vancouver, Canada, May 12–16, 1997, pp1287–1289.

A. Ueda, T. Ushiku and T. Mitsuhashi, Construction of travelling wave kicker magnet and pulse power supply for the KEK-photon factory storage ring, Proc. 19th Particle Accelerator Conference (PAC'01), Chicago, USA, 18–22 June 2001, pp. 4050–4052.

H. Hahn, N. Tsoupas and J.E. Tuozzolo, The RHIC injection kicker, Proc. 17th Particle Accelerator Conference (PAC'97), Vancouver, Canada, 12–16 May 1997, pp. 213–215.

H. Hahn, W. Fischer, V.I. Ptitsyn and J.E. Tuozzolo, All-ferrite RHIC injection kicker, Proc. 19th Particle Accelerator Conference (PAC'01), Chicago, USA, 18–22 June 2001, pp. 3705–3707.

H. Hahn, W. Fischer, Y.K. Semertzidis and D.S. Warburton, Up-graded RHIC injection kicker system, Proc. 20th Particle Accelerator Conference (PAC'03), Portland, Oregon, USA, 12–16 May 2003, pp. 1625–1627.

C.C. Jensen and G.E. Krafczyk, NuMi Proton Kicker Extraction System, Proc. 21st Particle Accelerator Conference (PAC'05), Knoxville, Tennessee, USA, 16–20 May 2005, pp. 692–694.

M.J. Barnes, Optimize SPICE models to accurately simulate frequency-dependent impedances, *Personal Engineering & Instrumentation News*, December 1996.

M.J. Barnes and G.D. Wait, Application of ELEKTRA to the prediction of inductance distribution for a travelling wave kicker magnet, Vector Fields North American User Group Meeting, San Jose, California, 17 October 1996.

M.J. Barnes and G.D. Wait, Solid state switch alternatives for klystron modulators, Proc. 1995 Modulator/Klystron Workshop, Stanford, California, USA, 9–11 October 1995.

T. Oki, The bridged-T network lumped kicker: A novel fast magnetic kicker system for a compact synchrotron, *Nucl. Instr. Meth. Phys. Res. A: Accel., Spectr., Det. Ass. Equip.,* **607** (2009) 489.

W. Jiang, Fast High-Voltage Switching Using Stacked MOSFETs, *IEEE Trans. Dielec. Elec. Ins.*, **14** (2007) 947.

T. Adachi and T. Kawakubo, *Phys. Rev. ST Accel. Beams* **16** (2013) 053501.

A. Krasnykh, Mini Kicker Workshop in ANL, Report No. SLAC-WP-096, 2013.

Proceedings of the CAS–CERN Accelerator School: Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018)

Fast Extraction: Single and Multiturn

M.A. Fraser
CERN, Geneva, Switzerland

Abstract
The basic principles and design considerations for fast extraction systems are introduced and reviewed before examples of CERN systems are given to illustrate different techniques. Single-turn fast extraction is used to introduce the main concepts before an overview of multiturn fast extraction techniques, including mechanical and magnetic beam splitting, is presented.

Keywords
Fast extraction; continuous transfer; CT; multiturn extraction; MTE.

1 Introduction

In many ways, extraction is simply the reverse process of injection; however, extraction is usually made after acceleration and therefore at higher energy, requiring stronger elements. As a result, many kicker and septum modules may be required; space–charge effects are usually less of a concern but beam losses, induced radio-activation, and machine protection considerations are typically more important. Different extraction techniques from synchrotrons exist, depending on the given requirements:

1. fast single-turn extraction ($\leq$1 turn): e.g., transfer between machines in an accelerator complex, transfer to experimental (production) targets or to safely dump the circulating beam (fast abort);
2. fast multiturn extraction (a few turns): to uniformly fill a synchrotron with a larger circumference or to vary the spill length to experimental targets over a few turns;
3. slow resonant multiturn extraction (many thousands of turns): providing experimental target, or patient, with long uniform spills;
4. other exotic extraction types, such as with bent crystals or charge-exchange extraction.

In this contribution, only items 1 and 2 will be discussed in detail. More details on items 3 and 4 can be found in the dedicated contributions made elsewhere in these proceedings.

Nowadays, extraction systems are usually considered from the conception of an accelerator, but this was not always the case in the past because lower stored beam energies and intensities meant that machine protection and radio-protection considerations were less of a concern. Consideration of the extraction system is a necessity during the conception of high-momentum machines, such as the Super Proton Synchrotron (SPS) or Large Hadron Collider (LHC) at CERN, where the layout, performance, and protection of the synchrotron might be significantly influenced by the design of the extraction system itself. A relevant example at present is the ongoing design work for the Future Circular Collider (FCC), where the extraction region is one of the most critical systems at the heart of the layout and lattice considerations [1]. The main considerations for a general extraction system can be summarized as follows:

1. destination or user: e.g., precision of beam delivery, tolerated beam loss, or emittance blow-up;
2. beam optics, integration, aperture, interference, and interplay with other essential subsystems: e.g., insertion regions may be required to meet specific requirements;
3. failure scenarios and their mitigation: e.g., at high energy or intensity, the extraction (beam-dump) system becomes an integral part of the machine protection system.

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-005.285

All of these considerations affect the choice of hardware employed for any given application, leading to an iterative design process involving many subsystems of the accelerator. To satisfy all of these constraints, one might require the extraction system to be placed in a dedicated insertion region that breaks the regular lattice structure of the synchrotron. The archetypal extraction system where machine protection concerns overshadow the design would be the LHC beam-dump system [2].

2 Single-turn fast extraction

A classical fast extraction system consists of a fast-pulsed kicker (magnetic or electric) that deflects the circulating beam over the septum blade and into the high-field region of a septum magnet located at a suitable phase advance downstream, as described in Fig. 1. A local, closed-orbit bump is used to reduce the kicker strength required to jump the septum blade. As shown in the inset in Fig. 1, the field pulse of the kicker rises synchronously with a particle-free gap in the circulating beam to avoid sweeping the beam over the septum blade. For extraction over a single turn, the length of the kicker pulse should exceed the length of the circulating bunch train, or the revolution period of the synchrotron.

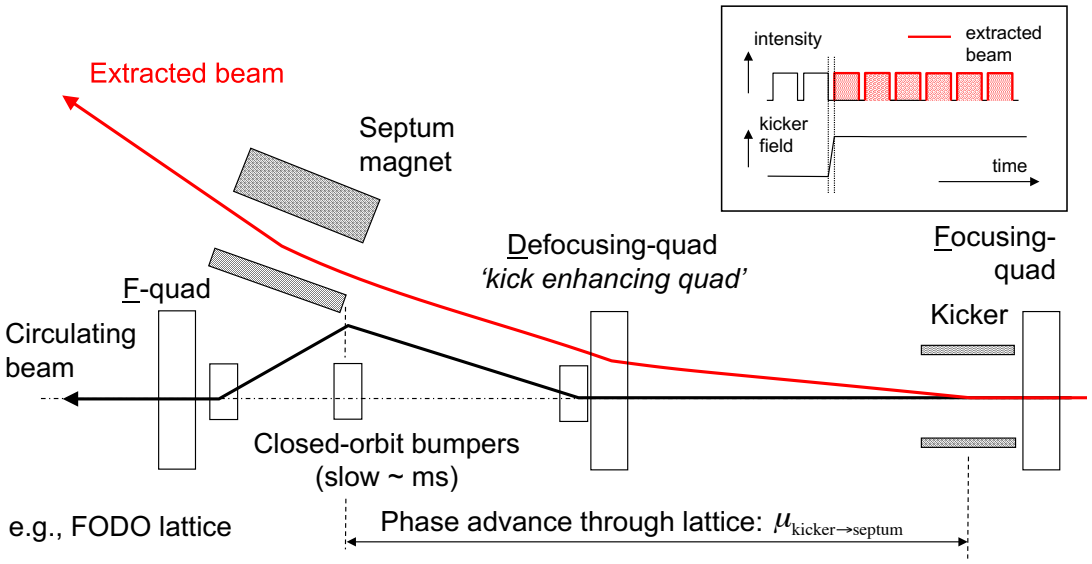


Fig. 1: Classical fast extraction system

2.1 Kick dynamics

The defocusing quadrupole placed between kicker and septum, often termed a kick enhancement quadrupole, can be seen as helping to deliver the phase advance and enhancing the kick imparted by the kicker, as described by the commonly used expression

$$\Delta x_{\text{septum}} = \Delta x'_{\text{kicker}} \sqrt{\beta_{\text{kicker}} \beta_{\text{septum}}} \sin\left(\mu_{\text{kicker}\to\text{septum}}\right) . \tag{1}$$

It is common to describe the kick strength in terms of the maximum number of beam widths of spatial separation developed downstream, which is why a normalized phase-space description becomes particularly useful. The beam width is often parametrized by a statistical quantity describing the distribution of the beam density, e.g., using the root-mean-square termed σ in the case of a Gaussian distribution. For a normalized deflection of $n\sigma$ at the kicker, one can write

$$\Delta x'_{\text{kicker}} = n\sqrt{\frac{\epsilon}{\beta_{\text{kicker}}}} \qquad \text{and} \qquad \Delta \bar{X}'_{\text{kicker}} = n\sqrt{\epsilon} \, , \tag{2}$$

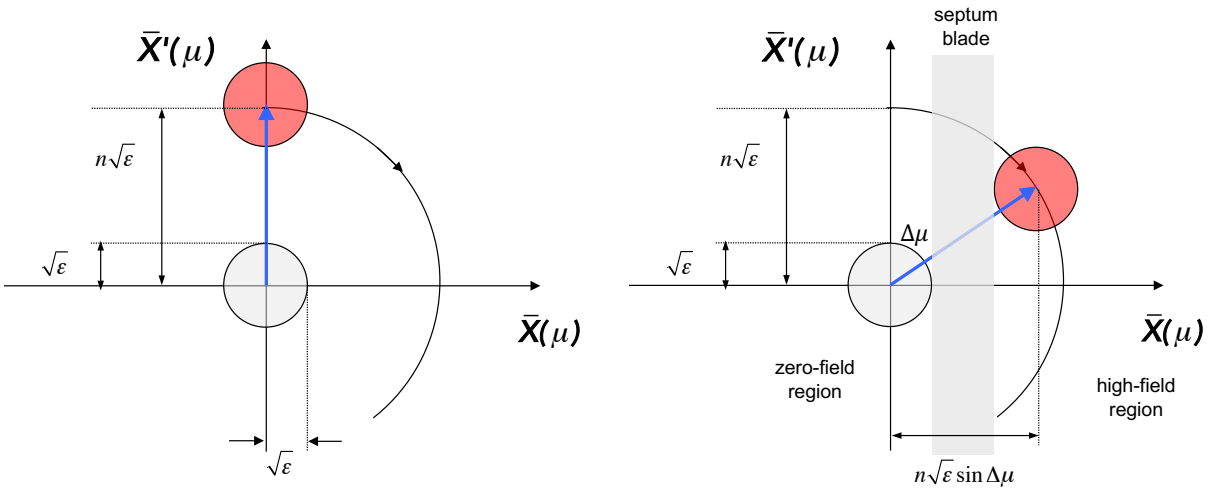
such that the spatial separation at the septum located at a phase advance $\Delta\mu$ downstream can be expressed as

$$\Delta x_{\text{septum}} = n\sqrt{\epsilon\beta_{\text{septum}}}\sin\Delta\mu \qquad \text{and} \qquad \Delta\bar{X}_{\text{septum}} = n\sqrt{\epsilon}\sin\Delta\mu\ , \tag{3}$$

where the 1σ beam emittance is defined by

$$\epsilon = \frac{\sigma^2}{\beta}. \tag{4}$$

The separation in normalized phase-space created by the kicker, at both its location and the downstream septum, is shown schematically in Fig. 2, where the particular normalization $(x, x') \rightarrow (\bar{X}, \bar{X}')$ is the same as that presented in the 'Introduction to beam transfer' paper of these proceedings.



(a) At the kicker

(b) At a septum located at a phase advance $\Delta\mu$ downstream

Fig. 2: Extraction kick dynamics in normalized phase-space developing an opening of $n\sigma$ between the extracted (red) and circulating (grey) beams, where $\epsilon = \sigma^2/\beta$.

There is no strong constraint on the distance between kicker and septum, as long as the phase advance is chosen so that spatial separation is achieved. As a result, it is not uncommon to see kickers and septa separated by large distances around a synchrotron or to find the kicker deflecting in the opposite direction to the septum, i.e., kicking to the inside of the ring to extract via a septum on the outside. For septa located on the outside of the ring, i.e., for $x > 0$ and maximal separation, one can write:

$$\Delta x_{\text{septum}} = \underbrace{(-1)^{m-1}|\Delta x'_{\text{kicker}}|}_{\Delta x'_{\text{kicker}}}\sqrt{\beta_{\text{kicker}}\beta_{\text{septum}}}\sin\left(\mu_{\text{kicker}\rightarrow\text{septum}}\right)\ , \tag{5}$$

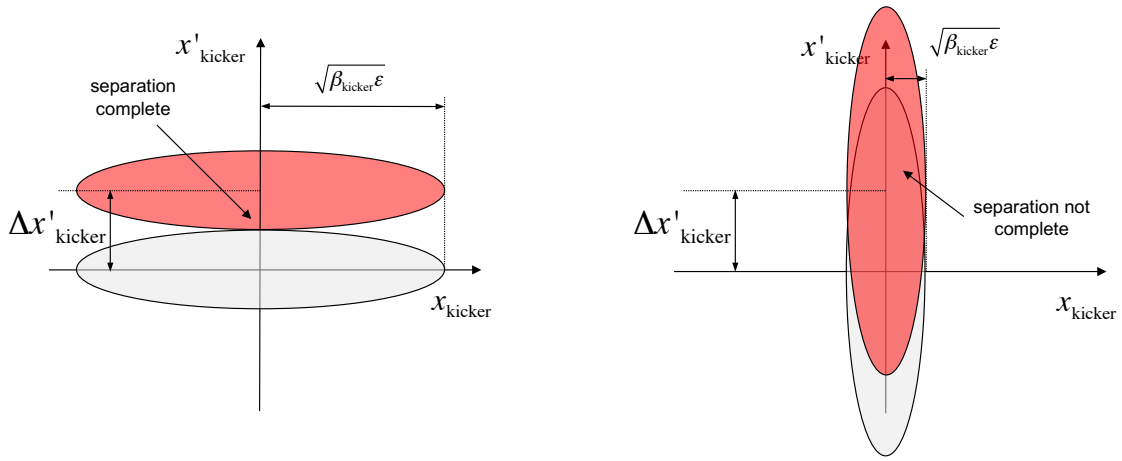
where

$$\mu_{\text{kicker}\rightarrow\text{septum}} = \frac{2m-1}{2}\pi \quad \text{for} \quad m = 1, 2, 3, \ldots \tag{6}$$

A single kicker system can service a number of extraction channels, with septa located at different positions in the ring. The desired extraction channel can be selected by setting the kicker polarity and energizing the relevant extraction bump. Operation of a single kicker system can reduce the cost and maintenance of running several extraction beam lines and reduce the impedance that multiple systems would present to the beam. However, a reduced acceptance and stability of the extracted beam can be expected. The layout of the fast extraction systems and beam lines installed in the CERN Proton Synchrotron (PS) over the years was designed with such considerations in mind [3]. The effect of non-local

fast extraction on beam quality in a large high-energy synchrotron is detailed in Ref. [4]. An extreme case of non-local extraction can be found in Ref. [5], where a crystal was used to deflect protons over the septum and into the extraction channel after a journey of 1.5 turns in the synchrotron, making almost 41 betatron oscillations.

From the simple relationship in Eq. 1, it is obvious that, to minimize the kick strength required, a designer would aim at maximizing the β-functions at the locations of the kicker and septum, and optimize the phase advance between those locations. The importance of a large β-function at the kicker is shown intuitively in Fig. 3, where the separation between the extracted and circulating beams is given for two extreme cases but the same kick angle. When the beam divergence is small, less angular deflection is needed to 'jump' outside the circulating beam. It is interesting to note that the large-β condition is exactly the worst case for emittance growth arising from imperfections in the uniformity of the kicker pulse, which is, in many ways, analogous to emittance growth in thin foils (or scatterers).



(a) Large β_{kicker}, assuming $\alpha_{\text{kicker}} = 0$

(b) Small β_{kicker}, assuming $\alpha_{\text{kicker}} = 0$

Fig. 3: Intuitive schematic example, demonstrating the advantage of a large β-function at the kicker

A similar argument can be made to explain why the β-function should also be large at the septum. One finds that, for a given kick angle, the separation, and therefore the maximum septum blade thickness that can be used, is proportional to $\sqrt{\beta_{\text{septum}}}$. Typically, the thicker the blade, the easier the engineering of the septum.

2.1.1 Kicker design parameters

The main considerations are typically rise and fall time, flat-top ripple, or field uniformity along the pulse and post-pulse ripple, depending on the particular application. The rise and fall times are usually defined between given limits of the nominal field, e.g., 2–98%, and the ripple in units relative to the nominal field. The typical figures of merit used to set tolerances for these design parameters are the acceptable beam emittance growth and beam loss. Three examples are given in Fig. 4 for the following applications:

- transfer between injectors in an accelerator complex for a high-energy collider: emittance conservation of extracted beam is of maximum importance, low flat-top ripple and fast rise times are needed;
- transfer for fixed-target physics in multiburst mode: perturbing the circulating beam might induce blow-up and losses of the circulating beam, fast rise or fall times and low post-pulse ripple are needed;

– fast beam abort to external absorber: safe, low-loss extraction of the circulating beam is important, flat-top ripple is relevant only for aperture reasons as emittance growth is not a consideration.

The expected perturbation imparted on the beam in the examples shown in Figs. 4(a) and 4(b) can be computed by taking the time-dependent waveform, applying the kick along a simulated bunch distribution and assuming complete filamentation in the receiving machine, see e.g., Ref. [6]. A transverse damper and feedback system in the receiving machine may be able to correct the time-correlated angular distortions introduced by the extraction kicker and reduce the emittance growth. In the case of the beam abort system shown in Fig. 4(c), flat-top ripple may be introduced deliberately to dilute the beam energy density on the dump absorber to an acceptable level that also guarantees transmission through the extraction channel and external beam line. Further details of beam dilution by fast-pulsed kickers for the SPS and LHC fast beam abort systems can be found in Refs. [7, 8]. Another important parameter driving the kicker specification is the aperture or stay-clear distance required for the circulating beam.

2.1.2 Septum design parameters

Owing to the relative strength of septa, which are typically 10 times stronger than kickers, field homogeneity is a more important design consideration. Both field homogeneity and shot-to-shot jitter can affect the transfer and beam quality. Along with aperture constraints, the stray field seen by the circulating beam must also be taken into account.

2.1.3 Aperture considerations

The main concerns for the extraction aperture are beam-loss-induced heating (impact on cooling requirements) and radio-activation (maintenance or damage) of the septum. The aperture is usually written in terms of the betatron beam size, where, in the horizontal plane,

$$\sigma_x = \sqrt{K_\beta \beta_x \epsilon_x}\,, \tag{7}$$

where the symbols have their usual meaning and K_β is an optics safety factor, generally taken as ≈ 1.2. The bumped, circulating beam aperture can be written as

$$n_{\sigma_x,\,\text{bump}} = \frac{A_{\text{septum}} - x_{\text{bump}} - \delta_{\text{CO}} - \delta_{\text{align}} - (|\delta_p| + |\delta_{\Delta p}|)|D_x|K_\beta}{\sigma_x}\,, \tag{8}$$

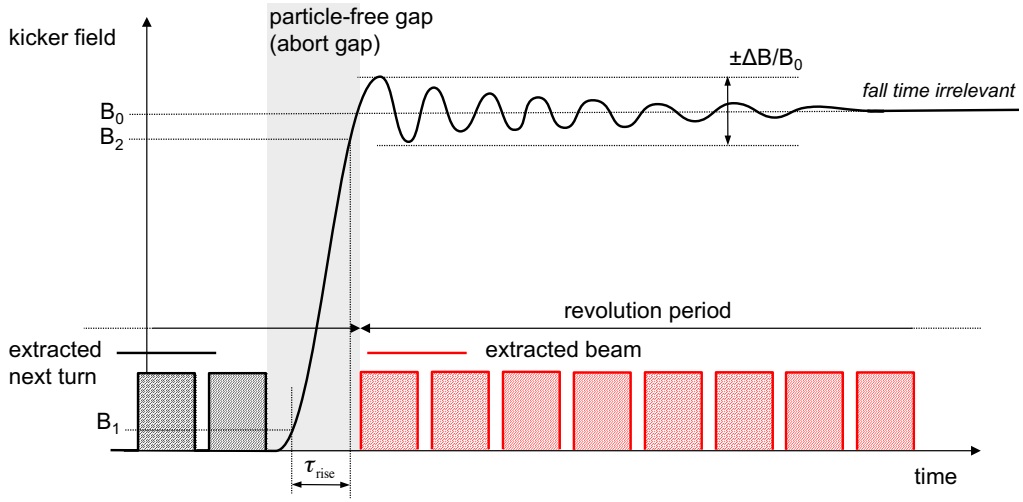
where the variables are defined in Figs. 5 and 6, δ_{CO} is the maximum residual on the closed orbit, δ_{align} is the mechanical alignment tolerance of the septum position, and δ_p and $\delta_{\Delta p}$ are the maximum momentum error of the beam and its spread, respectively. Similarly, the extracted beam aperture can be written

$$n_{\sigma_x,\,\text{ext}} = \frac{\Delta x_{\text{septum}} + x_{\text{bump}} - (\Delta x_{\text{blade}} + A_{\text{septum}}) - \delta_{\text{CO}} - \delta_{\text{align}} - (|\delta_p| + |\delta_{\Delta p}|)|D_x|K_\beta}{\sigma_x}\,, \tag{9}$$

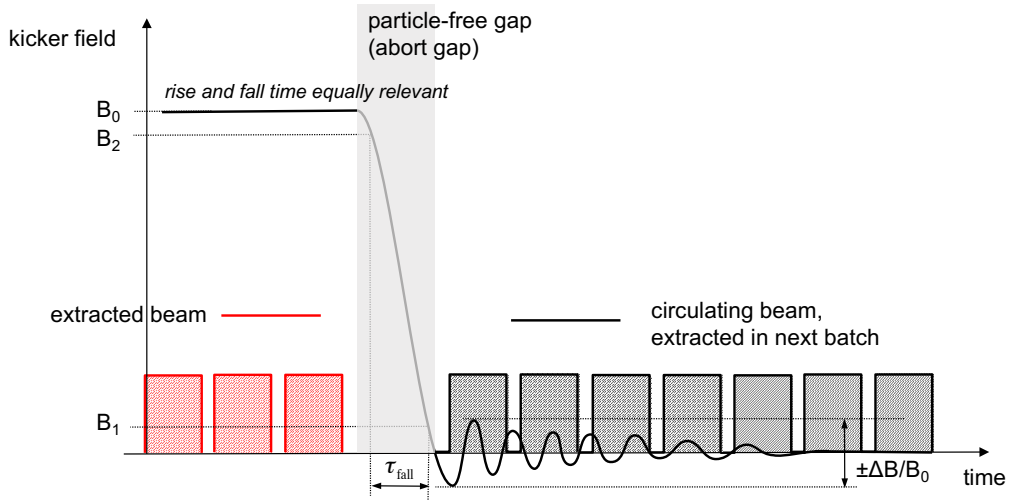
which is dependent on the strength of the kicker and the septum blade thickness. The vertical aperture for the extracted beam is usually critical because of the narrow septum gap height,

$$n_{\sigma_y} = \frac{y_{\text{gap}}/2 - \delta_{\text{CO}} - \delta_{\text{align}}}{\sigma_y}\,. \tag{10}$$

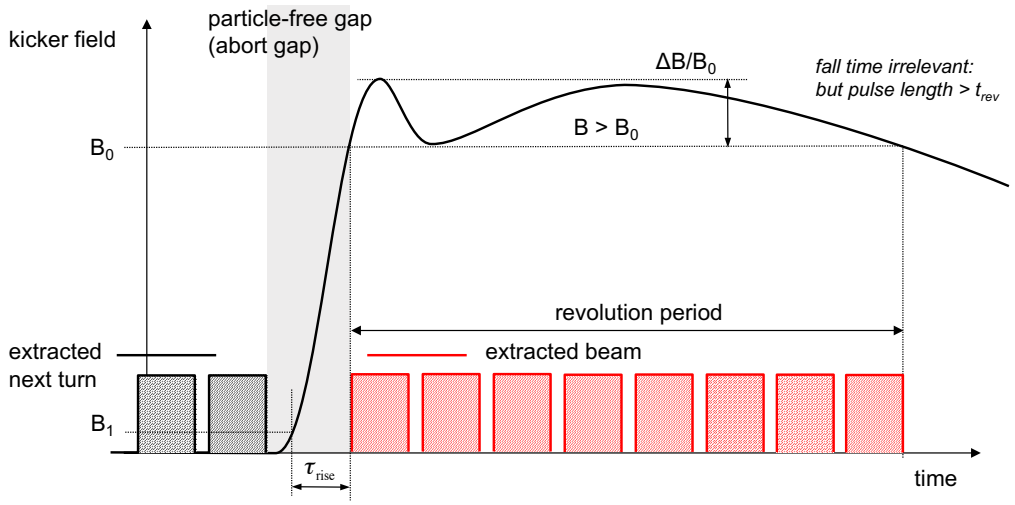
The extraction kicker is usually installed on the circulating beam trajectory; therefore, its gap height (vertical aperture) is constrained by the beam size at injection for a horizontal kicker, as shown in Fig. 5. The proximity of the extraction septum to the circulating beam is typically limited by the size of the beam, and therefore the machine acceptance, at injection, as shown in Fig. 6. Most synchrotrons serve a number of beam types and users; to avoid mechanically actuating the septum position to maximize the acceptance on a beam-by-beam basis, a closed-orbit bump is exploited. The septum is left mechanically



(a) Single-turn extraction between injectors in accelerator complex for high-energy collider



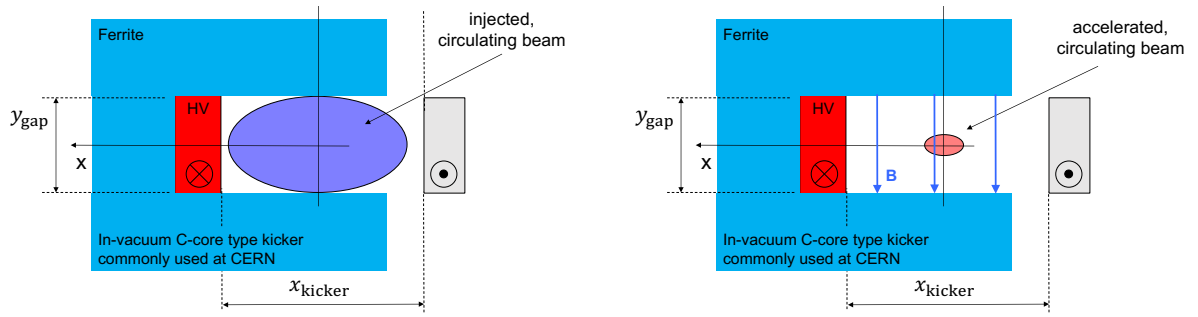
(b) Transfer for fixed-target physics in multiburst mode



(c) Fast beam abort to external absorber

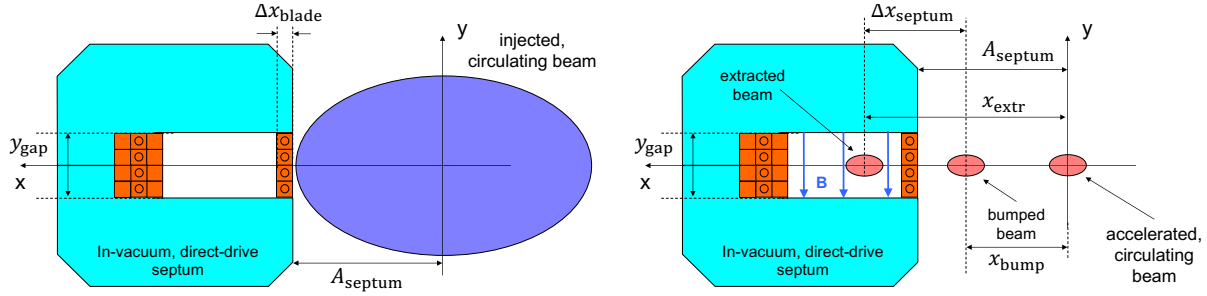
Fig. 4: Kicker waveforms, showing typical applications for fast extraction systems and relevant parameters

fixed outside of the machine acceptance and the accelerated beam is moved closer to the septum by the bump immediately before extraction. The aperture that can be gained is related to the adiabatic damping of the emittance during acceleration and the ratio of the relativistic factor, $\sqrt{\beta_{\mathrm{inj}}\gamma_{\mathrm{inj}}/\beta_{\mathrm{ext}}\gamma_{\mathrm{ext}}}$. Similar ideas have also been considered to reduce the required aperture of extraction kickers by bumping the beam into an open C-shaped geometry after acceleration, where the return conductor is hidden on the rear side of the ferrite yoke [9]. In this way, the gap height is reduced and the available kick strength for a given voltage is increased. Interestingly, the first kicker systems used in the PS had to be actuated into the beam after acceleration for exactly this reason [10].



(a) At injection energy (b) At extraction energy, after acceleration

Fig. 5: Aperture considerations at the kicker for a fast extraction system



(a) At injection energy (b) At extraction energy, after acceleration

Fig. 6: Aperture considerations at the septum for a fast extraction system

2.1.4 Closed-orbit bumps

Closed-orbit bumps are regularly used during extraction to bring the circulating beam close to the septum (slow bump), to reduce the kicker strength, and to control multiturn extraction turn by turn (fast bump). Closed-orbit bumps are also commonly used for injection for similar purposes. Dipole 'bumper' magnets are used to steer the closed orbit away from the nominal trajectory in a localized part of the synchrotron. The following standard bump configurations exist and will be briefly discussed:

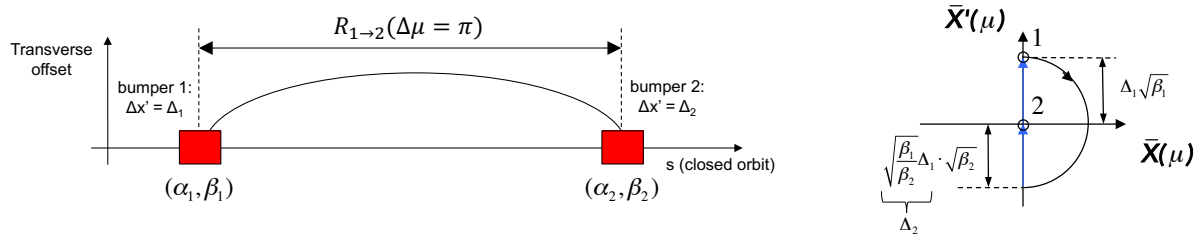
- π-bump;
- three- and four-magnet bumps.

The most important results needed for the implementation and design of closed-orbit bumps are quoted next. For further details and derivations, see Refs. [11, 12]. In reality, the analytical results presented next are rarely applicable to practical scenarios because there are usually far more constraints to consider. For such purposes, dedicated optics codes, e.g., MAD-X [13], are used to match bumps with numerical optimization routines.

2.1.4.1 π-bump

The simplest closed bump, the π-bump, is constrained by a phase advance of π or 180° between two dipole bumper magnets. As shown in Fig. 7, the first magnet opens the bump and the second closes it after half a betatron oscillation, where R is the linear transfer map describing the dynamics from one location to another along the accelerator. The relative deflection angle of the two bumpers is given by the lattice functions at the bumpers, as

$$\frac{\Delta_2}{\Delta_1} = \sqrt{\frac{\beta_1}{\beta_2}} \,. \tag{11}$$



(a) Real space (x, s) along the reference trajectory

(b) Normalized phase-space $(\bar{X}, \bar{X}')$

Fig. 7: π-bump

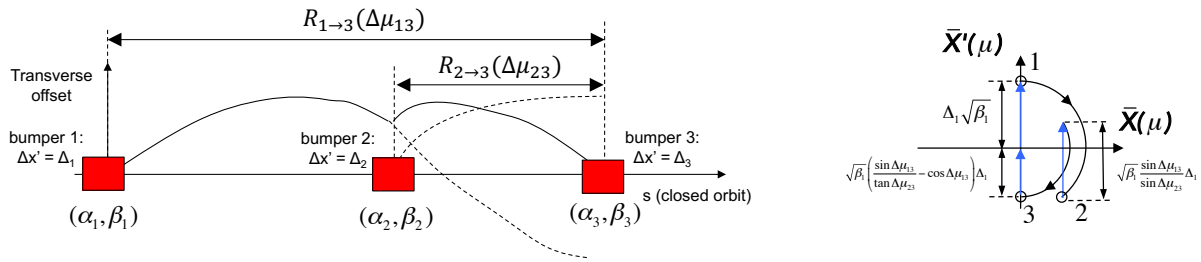
2.1.4.2 *Three-magnet bump*

In a real accelerator, more degrees of freedom are often needed, for practical reasons. A third magnet can be added to close the bump for (almost) any value of phase advance between the first and last bumper. In this case, either the position or the angle can be matched at a location inside the bump. As shown in Fig. 8, the first and second bumpers put the beam back on the axis of the reference trajectory at the third bumper, which compensates for any remaining angle. The relative deflection angles of the bumpers can be written in terms of the lattice functions at the bumpers:

$$\frac{\Delta_2}{\Delta_1} = -\sqrt{\frac{\beta_1}{\beta_2}}\sqrt{\frac{\sin\Delta\mu_{13}}{\sin\Delta\mu_{23}}} \quad \text{and} \quad \frac{\Delta_3}{\Delta_1} = \sqrt{\frac{\beta_1}{\beta_3}}\left(\frac{\sin\Delta\mu_{13}}{\tan\Delta\mu_{23}} - \cos\Delta\mu_{13}\right) \,. \tag{12}$$

2.1.4.3 *Four-magnet bump*

To control both the position and angle at a given point S inside the bump, a fourth magnet is needed, see Fig. 9. The first two bumpers select the position and angle at a given location downstream of the



(a) Real space (x, s) along the reference trajectory

(b) Normalized phase-space $(\bar{X}, \bar{X}')$

Fig. 8: Three-magnet bump

second bumper, e.g., at the extraction septum, before two more bumpers return the beam to the reference trajectory. The relative deflection angles of the bumpers can be written in terms of the lattice functions at the bumpers and the matching point, as

$$\Delta_1 = -\frac{1}{\sqrt{\beta_1\beta_S}}\frac{\cos\Delta\mu_{2S} - \alpha_S \sin\Delta\mu_{2S}}{\sin\Delta\mu_{12}}x_S - \sqrt{\frac{\beta_S}{\beta_1}}\frac{\sin\Delta\mu_{2S}}{\sin\Delta\mu_{12}}x'_S \tag{13}$$

and

$$\Delta_2 = -\frac{1}{\sqrt{\beta_2\beta_S}}\frac{\cos\Delta\mu_{1S} - \alpha_S \sin\Delta\mu_{1S}}{\sin\Delta\mu_{12}}x_S + \sqrt{\frac{\beta_S}{\beta_2}}\frac{\sin\Delta\mu_{1S}}{\sin\Delta\mu_{12}}x'_S\,. \tag{14}$$

One can derive by symmetry the strengths of Δ_3 and Δ_4 by applying the following transformations to Eqs. (13) and (14):

$$\beta_1 \to \beta_4, \quad \beta_2 \to \beta_3, \quad \alpha_S \to -\alpha_S, \quad x_S \to -x_S, \quad \Delta\mu_{1S} \to \Delta\mu_{S4}, \quad \Delta\mu_{2S} \to \Delta\mu_{S3}\,. \tag{15}$$

The derivation of these relations is detailed at length in Ref. [11]. A useful application of the four-magnet bump is the construction of 'orthogonal bumps' to allow the independent variation of either the position or angle of the beam at a given location, whilst the other variable is held constant.

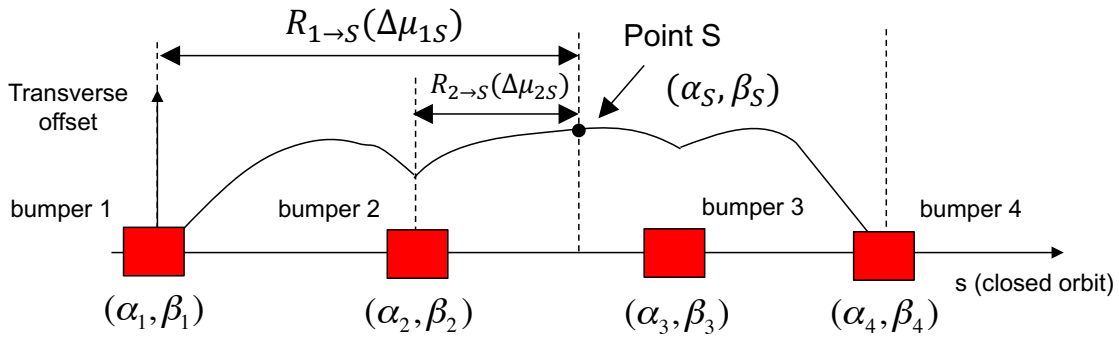


Fig. 9: Four-magnet bump in real space (x, s) along the reference trajectory

2.2 Example: fast extraction from SPS to LHC

The SPS fills the LHC with batches of high-brightness proton beams at a momentum of 450 GeV/c ($B\rho \approx 1500$ m) fast-extract over a single turn. The SPS has two fast extraction regions in Long Straight Sections (LSSs) 4 and 6 to fill the two counter-rotating LHC beams, as shown in Fig. 10.

We will consider LSS4, shown in the simplified layout of Fig. 11, in more detail, to explore an example of an operational fast extraction system. A full description of the system can be found in Ref. [15]. The optics in LSS4 are an extension of the SPS's periodic FODO structure with widened aperture quadrupoles (QA) and a closely symmetrical four-magnet (MP) bump. The extraction kicker (MKE) is located about 70° in phase advance upstream of the extraction septum (MSE) and both deflect towards the outside of the ring in the horizontal plane. The basic optics parameters describing the extraction system are summarized in Table 1. The bump amplitude is matched with a maximum displacement at the MSE. It is interesting to note the deflection of the closed orbit in the quadrupoles and how the trajectories contrast in the real and normalized co-ordinates shown in Fig. 11(a). The bump amplitude is non-zero in the MKE. The upstream ends of both the MKE and MSE systems are located close to focusing quadrupoles (QF), where the β-function is at its largest. The length of the MSE, consisting of six separate tanks, is impressive and consumes most of the half-period in which it is installed. The large-aperture quadrupoles have a window in the coil through which the extracted beam can pass unimpeded

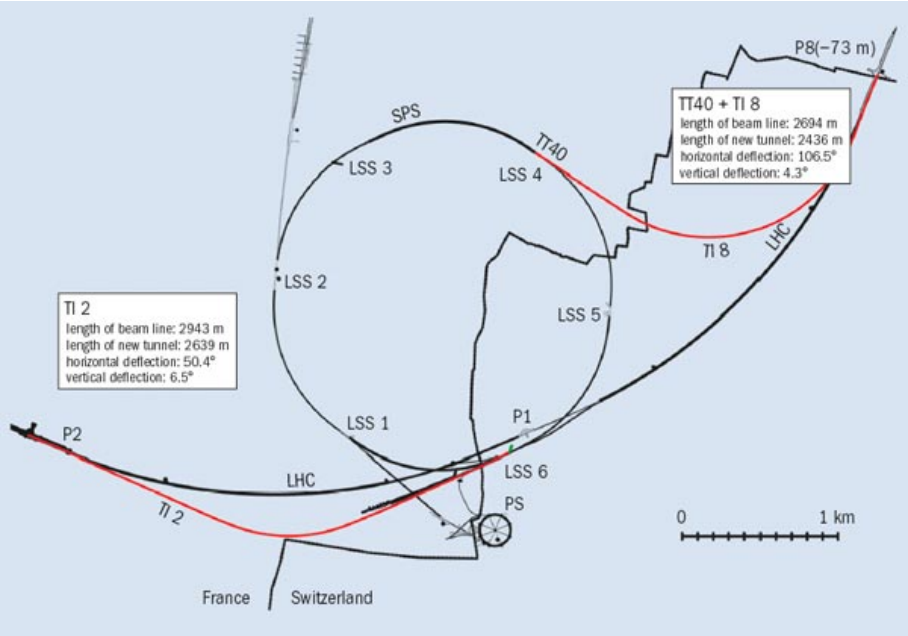
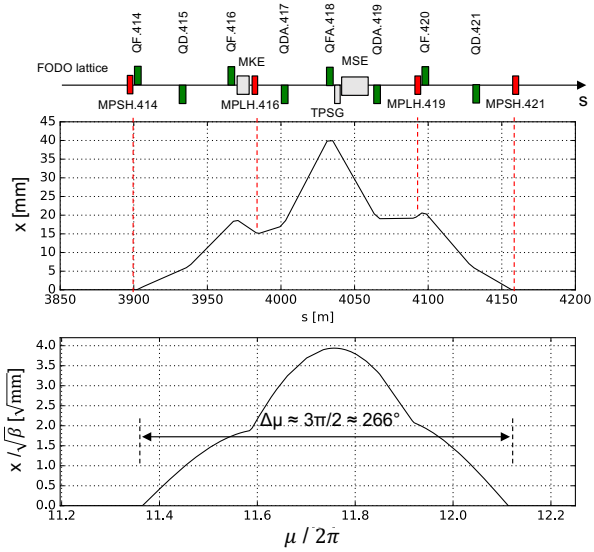
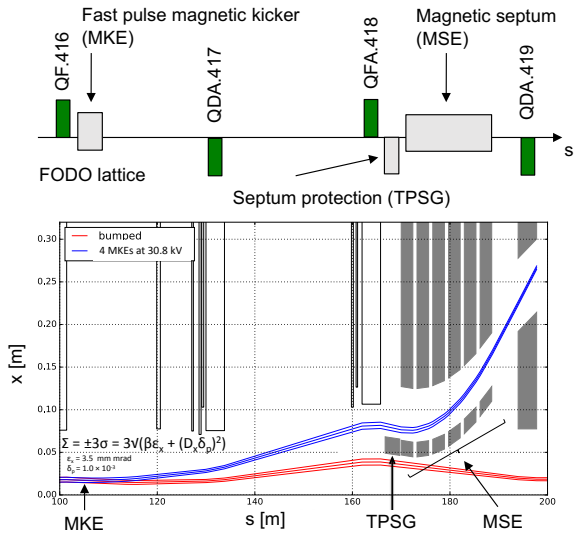


Fig. 10: SPS layout and LHC transfer lines: LSS4 (TI8) and LSS6 (TI2) [14]

to integrate an extraction beam line into the LSS. The passage of the extracted beam through the coil window of QDA.419 can be seen in Fig. 11(b). The position and angle of the MSE, the bump shape, and the bump shape amplitude are optimized with the aperture of the beam in mind. Approximately 10σ is available after acceleration for both the bumped and extracted beams.



(a) Closed four-magnet bump trajectory



(b) Extraction channel aperture

Fig. 11: LSS4 fast extraction system for transfer from SPS to LHC

The stored energy of the beam at extraction amounts to close to 2.5 MJ and its brightness is high enough to damage the machine if mis-steered onto the aperture. To reduce the risk of failure and damage to the machine, a sophisticated interlock system is in place. As another level of protection against fast failure scenarios, which are difficult to detect and react to, such as a kicker flashing over as it pulses

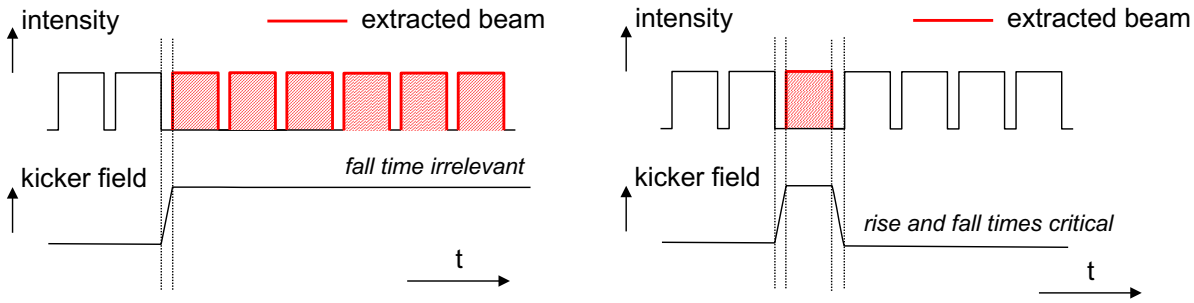
Table 1: LSS4 optics parameters (given at upstream end of element)

Optics parameter	MKE	TPST	MSE	QF (QD)
β-function, β_x [m]	97.0	100.2	88.4	105 (32)
Phase advance, $\Delta\mu_x$ [degrees]	0	64.8	66.8	32
Deflection angle, $\Delta x'$ [mrad]	0.42	—	≈12	—

(potentially causing direct beam impact) or a kicker triggering erratically and asynchronously (causing the beam to sweep over the septum blade), a dedicated absorber or protection device (TPSG) is installed directly in front of the MSE to protect it. The TPSG is designed to survive a direct beam impact and is built from a sandwich of different materials to dilute the beam energy density sufficiently to avoid damage to the MSE [16].

3 Multiburst fast extraction

Multiburst fast extraction is demonstrated schematically in Fig. 12. It is essentially identical to single-turn fast extraction except that fast-pulsed kickers with short rise and fall times are used to extract part of the circulating beam by cutting-out 'bursts'. As discussed for the context of single-turn fast extraction, particle-free gaps must be present in the circulating beam, within which the extraction kicker system can pulse cleanly.



(a) Single-turn extraction (executed once)

(b) Multiburst extraction (repeated many times)

Fig. 12: Single-turn vs. multiburst extraction (time on horizontal axis)

An example of the multiburst fast extraction technique was the extraction of two bursts, separated in time by 50 ms (≈2000 turns), of approximately 2.5×10^{13} protons at 400 GeV from the SPS through LSS4 to the CERN Neutrino to Gran Sasso (CNGS) neutrino production target [17, 18]. In this case, the entire circumference of the SPS (23 µs) was filled by two batches, each 10.5 µs in length, injected from the PS and separated by approximately 1 µs. The circulating beam intensity around approximately half of the SPS ring is shown in Fig. 13, alongside the waveform of the kicker's current pulse. After acceleration and on the flat-top, a single fast extraction kicker system, recharged between bursts, was used to extract the two bursts into the extraction channel and towards the neutrino production target. In this case, two capacitor banks were used to resonantly and rapidly recharge the pulse-forming network. In reality, the particle-free gaps are not completely empty and a small amount of beam is lost during the extraction process, which, in this case, was estimated as a few tenths of one percent.

The total intensity that can be delivered by the synchrotron is limited by the presence of particle-free gaps and is dependent on the performance of the relevant kicker systems. Unless several extraction kicker systems are used and triggered independently, the repetition rate of the 'burst-mode' is typically limited by the need to recharge the kicker's power supply. An interesting application of burst-mode extraction is for the filling of a future high-energy collider, e.g., the FCC, where the amount of beam that can be transferred at once is limited by machine protection considerations, owing to the high beam

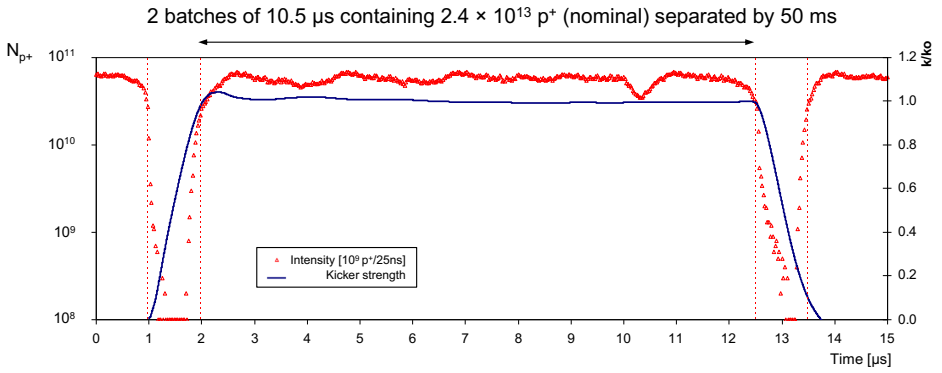


Fig. 13: Circulating CNGS beam intensity around approximately half of the SPS circumference, compared with the measured extraction kicker's current waveform [18].

energy and brightness [19]. To keep filling times reasonable, the filling pattern in the injector synchrotron and the design of its extraction system will have to be compatible with rapid pulsing in a burst-mode to empty the injector batch-by-batch safely and fill the collider efficiently.

Even by exploiting resonant charging from pre-charged capacitor banks, the time taken to recharge a pulsed-power supply, e.g., a pulse-forming network or pulse-forming line, is, in most cases, many times the revolution period. Installing more than one kicker system is challenging, owing to physical space constraints, or unfavourable, owing to the large beam impedance presented by kicker magnets. As a consequence, extracting in burst-mode very quickly over sequential turns, or at timescales close to the revolution period, becomes rather impractical. In applications where a continuous spill from the synchrotron is required over a few sequential turns, splitting of the beam must typically be carried out in the transverse plane; this is discussed next.

4 Multiturn fast extraction

To extend the spill length from a single turn to several turns of a synchrotron, one must consider separating or splitting the circulating beam in transverse phase-space whilst applying impulses on a turn by turn basis to extract the beam stepwise. When the energy and intensity of the circulating beam is high, a main challenge is the induced radio-activation of the accelerator and its extraction equipment, resulting from the loss of particles during the splitting and extraction processes. The fast and dynamic nature of the extraction process and resulting beam losses does not make it any easier to control. The impact of beam-induced radio-activation is particularly relevant at CERN, where fixed-target experimental facilities request upwards of 1×10^{19} protons on target per year at energies ranging from approximately 20 to 400 GeV. This has motivated significant developments for lower-loss multiturn extraction techniques at CERN over recent decades [20].

4.1 Non-resonant multiturn extraction: mechanical splitting

Non-resonant multiturn fast extraction by transverse shaving is a brute-force method employed to lengthen the spill from a synchrotron, which is useful for uniformly filling a larger downstream synchrotron and reducing the filling time, or for providing experiments with spills over a few turns. The concept is outlined in Fig. 14 and an example with a fractional tune of $q_{x,\mathrm{frac}} \approx 0.125$ is shown to illustrate the principle in Fig. 15. It can be instructive to see the process as reverse multiturn injection [21]. The circulating beam is shaved turn by turn on a thin electrostatic septum using a fast, programmable closed bump that forces the beam over the septum blade. The spatial separation required to extract the

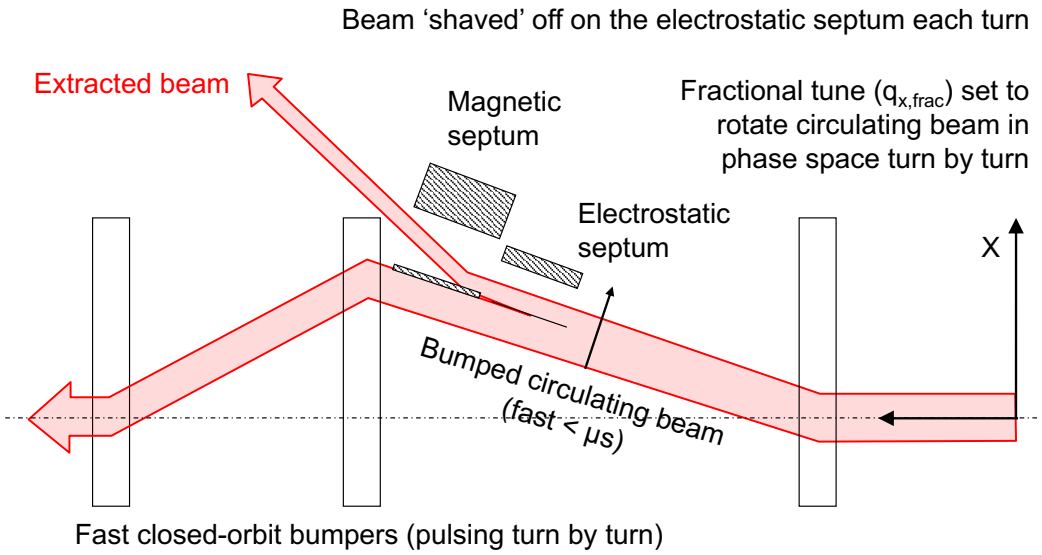


Fig. 14: Non-resonant multiturn extraction

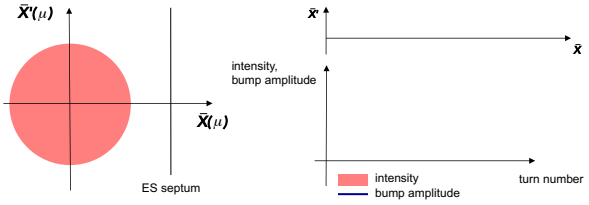
beam from the synchrotron is created by cutting it on the thin septum blade, the deflection of the septum's electric field, and a non-zero phase advance to another, usually thicker, magnetic septum placed downstream. The tune of the synchrotron rotates the beam into the electrostatic septum to continuously shave the beam, turn by turn. If the circulating beam is debunched, the extracted spill will have a continuous time structure over the length of the spill, convoluted with the transverse beam density being shaved. The fractional part of the tune is adjusted, together with the height of the closed bump, to deliver the spill length desired.

As demonstrated in Fig. 15, the extraction process typically results in smaller beam emittances because the circulating beam is mechanically sliced in the plane of the extraction. This results in a spill with optical parameters that vary turn by turn and, as a consequence, the emittance, centroid position, and trajectory vary along the spill. It can be shown that, for realistic transverse beam distributions, spills with both uniform intensity and emittance are not possible [22].

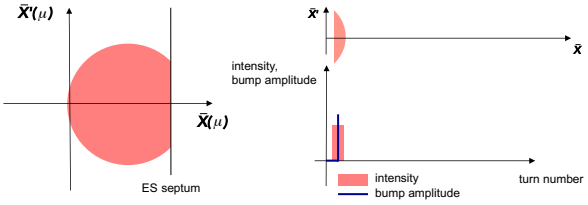
The process is intrinsically lossy, as the septum blade directly intercepts the beam and, as a result, the accelerator is radio-activated by the extraction process. The extraction efficiency depends strongly on the number of turns extracted; therefore, it is seldom used to provide spills longer than ∼10–20 turns. This has limited its application at laboratories worldwide, especially for beam delivery to fixed-target experiments, although recently published beam tests at the U-70 synchrotron at IHEP describe the process and its challenges well [23]. Instead, resonant slow extraction can provide a more stable and efficient extraction mechanism, as described elsewhere in these proceedings, and can even be effective for relatively short spills of only ∼100 turns where the resonance is pulsed rapidly [24].

4.1.1 Continuous transfer at the CERN Proton Synchrotron

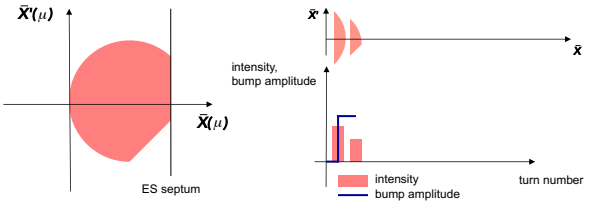
The term 'continuous transfer' (CT), as coined at the CERN PS [25], will be used synonymously with the expression 'non-resonant multiturn extraction' in these proceedings. Continuous transfer was developed to fill the larger SPS efficiently and uniformly for its fixed-target experimental programme, where the circumference of the SPS is approximately 11 times that of the PS. It should be noted that, in this case, uniform filling is of particular relevance for avoiding the onset of longitudinal instabilities in the SPS. The first versions of CT experimented with single-batch fills, with extractions over 10 and 11 turns from the



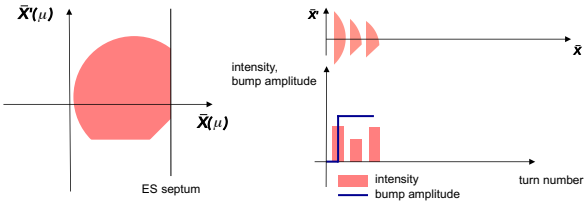
(a) turn 0



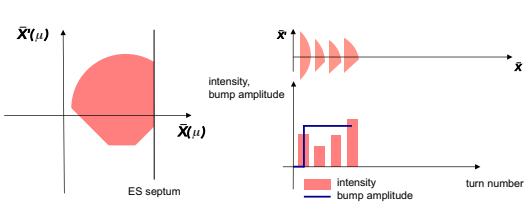
(b) turn 1



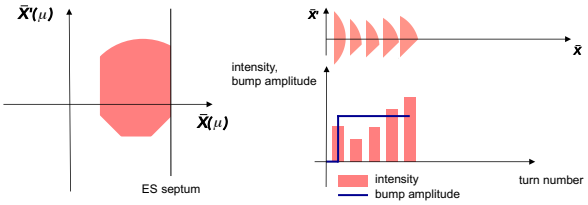
(c) turn 2



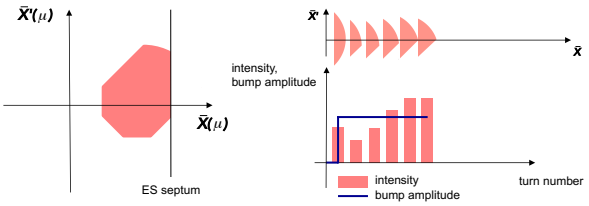
(d) turn 3



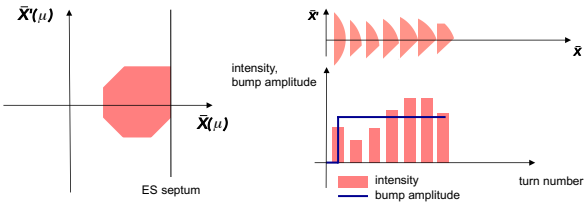
(e) turn 4



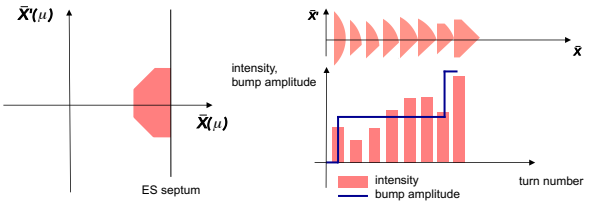
(f) turn 5



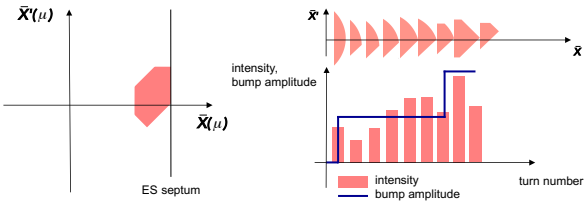
(g) turn 6



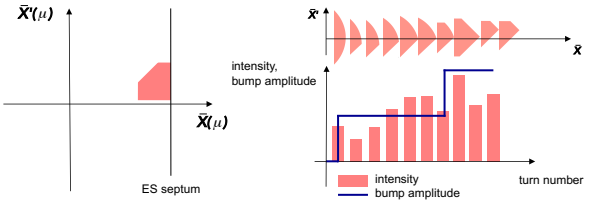
(h) turn 7



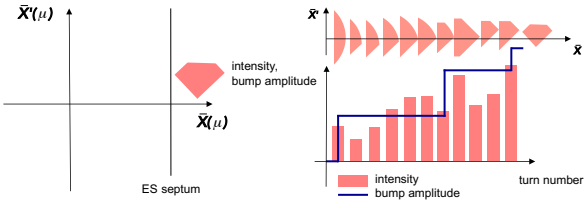
(i) turn 8



(j) turn 9



(k) turn 10



(l) turn 11

Fig. 15: Example of non-resonant multiturn extraction with fractional tune $q_{x,\text{frac}} \approx 0.125$.

PS, before the present-day double-batch scenario was developed (Fig. 16), which consists of the injection of two batches, each with a length of five PS turns. After acceleration and on the flat-top at 14 GeV/c the tune is brought close to $4Q_x = 25$. Slow, closed bumpers (BSW) bring the circulating beam close to the electrostatic (SEH31) and magnetic (SMH16) extraction septa before a fast-pulsed closed bump (BFA) around the electrostatic septum fires to commence the extraction. The beam is forced over the SEH31 blade, initiating the shaving. The deflection imparted by the SEH31 and the phase advance to SMH16 are such that the shaved slice is extracted into the transfer line as it passes the magnetic septum ∼500 m downstream. The tune rotates the beam over the subsequent four turns, continuously shaving the circulating beam before the fast bump amplitude increases further to push the remaining beam core out of the accelerator in the final turn. To reduce the density of the beam impinging the septum blade and enhance the deflection of the kickers and septa, a local, closed optics perturbation is applied between kick enhancement quadrupoles (QKE) during the extraction to increase the horizontal β-function at the SEH31. The closed QKE perturbation is also used to reduce the dispersion at SEH31 to ensure that the extraction is less sensitive to the momentum spread in the beam. Further details of the optics during the CT extraction can be found in Ref. [26]. To improve the RF capture in the SPS, the beam is debunched and partially recaptured at the SPS RF frequency of 200 MHz before extraction. A complete description of the CT extraction system at the PS can be found in Ref. [10].

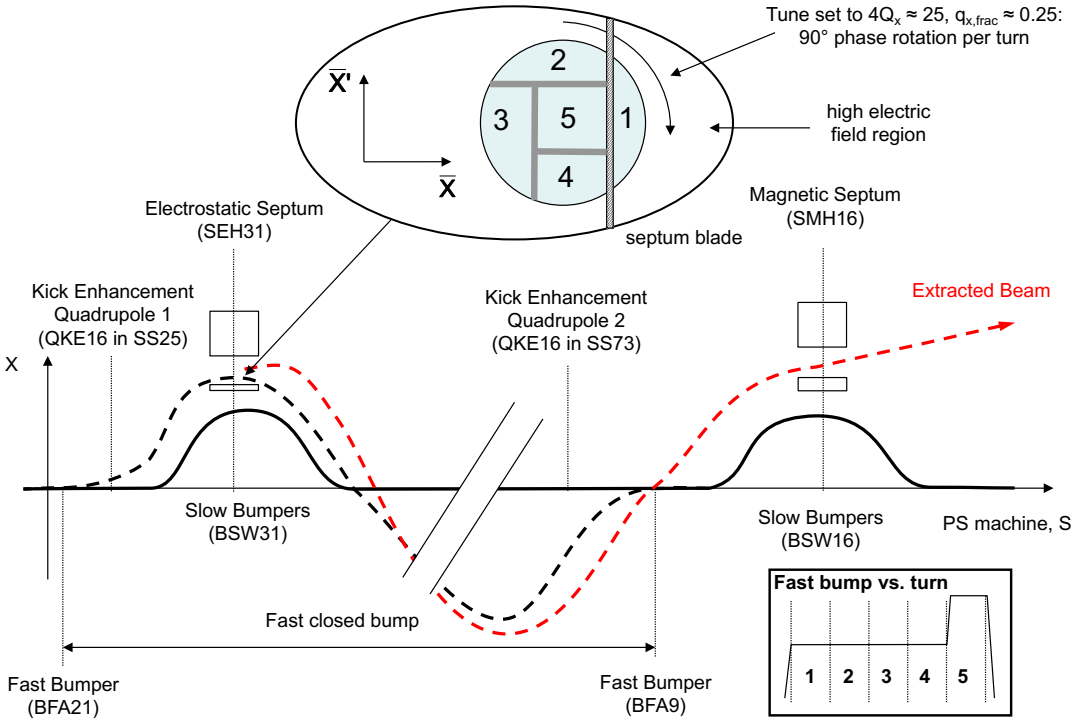
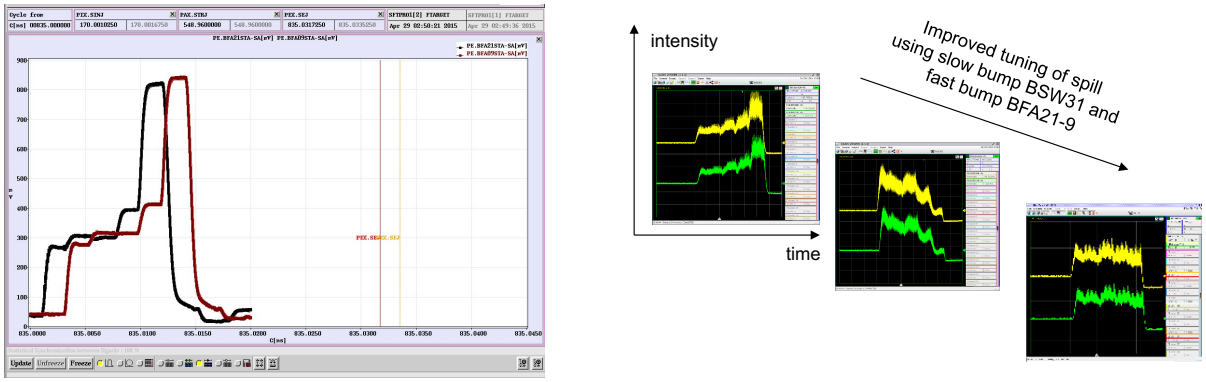


Fig. 16: Overview of CT extraction scheme at CERN PS: the numbers indicate the turn on which each slice is extracted.

A simplified waveform of the fast bump amplitude is shown in the bottom-right panel of Fig. 16, where in reality some turn by turn adjustment is available to help tune the spill by changing the presentation of the circulating beam to the electrostatic septum. In Fig. 17, an operational example of the BFA9 and 21 kicker waveforms used to control the fast bump around the SEH31 is presented, where the time-of-flight delay between the kickers located at different positions around the ring is evident. The relatively large number of turns makes the kicker system complex; more details of the kicker hardware used to produce such a turn by turn, variable amplitude closed bump can be found in Refs. [27, 28].



(a) BFA9 and 21 waveforms

(b) Beam intensity extracted over five PS turns (10.5 µs)

Fig. 17: Turn by turn tuning of fast bumpers around SEH31 to achieve a spill with uniform intensity [29]

4.1.1.1 Turn by turn mismatch and trajectory error

A turn by turn variation in the position and emittance of each slice is observed in the extraction beam line and receiving machine as a result of tuning for constant intensity over the spill. This is illustrated at the electrostatic septum in Fig. 18. The resulting trajectory oscillations can be compensated by fast-pulsed, turn by turn kickers installed in the extraction beam line; however, the turn by turn optical mismatch is far more challenging to correct and is left uncorrected [30]. The turn by turn mismatch between the PS and SPS has been studied analytically for Gaussian and quasi-parabolic transverse distributions and for two cases: tuning the fast bump to maintain (i) constant intensity and (ii) constant emittance throughout the spill [22]. The results of this study are summarized in Fig. 19 for a Gaussian transverse distribution. The blow-up factor from mismatch for each turn is typically very high (30–50%) and smaller when holding the extracted intensity constant. In both cases, the average emittance, even after filamentation in the SPS, is considerably reduced compared with the emittance before CT extraction in the PS. This effect, resulting from the slicing of the phase-space into smaller pieces, has been exploited at CERN to overcome the vertical aperture limitation of the SPS by exchanging the horizontal and vertical emittances in the transfer line upstream of the SPS injection point. Such a phase-space manipulation is achieved using an insertion region containing normal and skew quadrupoles [31].

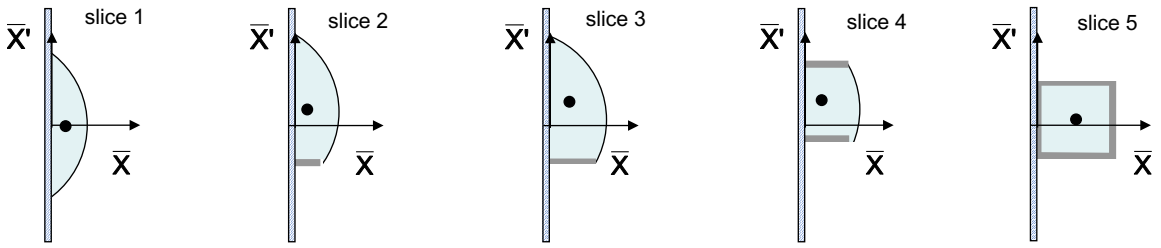
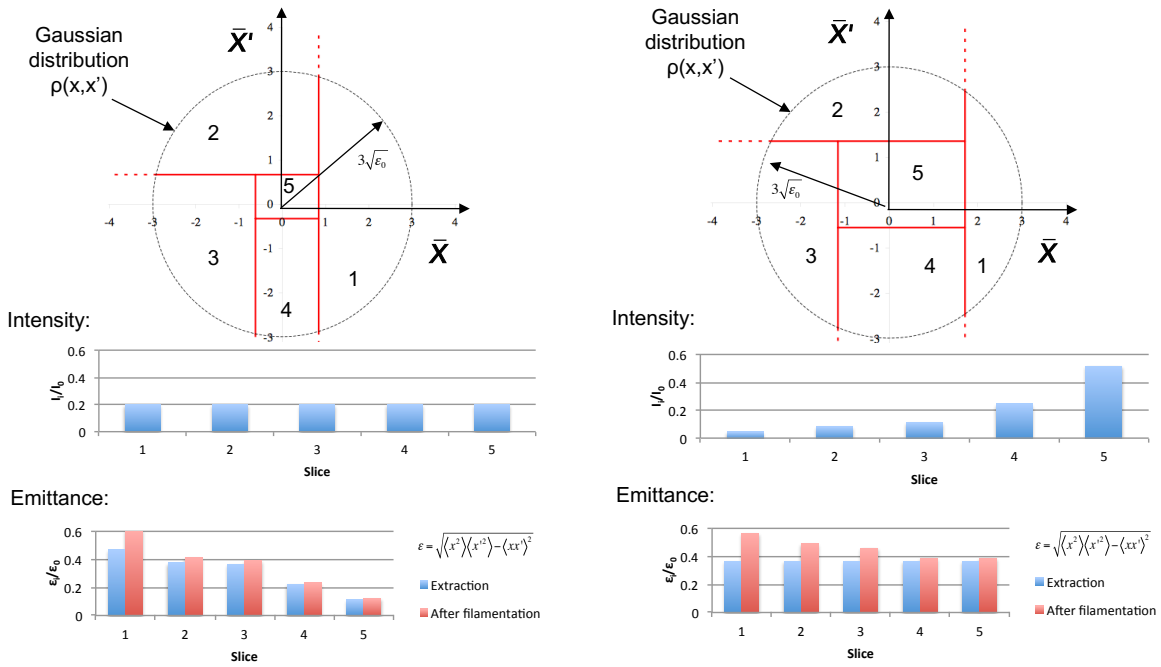


Fig. 18: Turn by turn centroid variation of slices extracted by CT

4.1.1.2 Beam loss during extraction and mitigation techniques

Depending on how well the extraction has been set up and optimized, approximately 6% of the beam is lost in the machine during the CT extraction process as it impinges and scatters from the electrostatic septum blade. Typically, one strives for a large β-function at the septum with as thin a blade as possible to reduce the amount of beam intercepted. The 1.8 m long SEH31 blade, composed of a molybdenum



(a) Constant intensity per slice $= I_0/5$

(b) Constant emittance per slice $\neq \epsilon_0/5$

Fig. 19: Turn by turn intensity and emittance at extraction and after filamentation from mismatch in SPS [22]

foil, has a nominal thickness of 0.1 mm but warping of the foil caused by beam-induced heating typically increases its effective thickness over time. The magnitude of the β-function at the septum is limited by the gap width for the extracted beam over which the available high voltage must be applied to provide the electric deflection needed to jump the magnetic septum. The most critical parameter determining the extraction efficiency for CT is the relative alignment of the beam and septum, followed by the horizontal beam size (emittance) and the thickness of the septum blade; a complete explanation of the optimization of beam loss for CT can be found in Refs. [32, 33].

The loss distribution around the ring depends strongly on the optics and aperture of the machine. During high-intensity CNGS operation, these losses amounted to more than 1×10^{18} protons per year, which, at 14 GeV/c, induced significant radio-activation of the machine. This is especially problematic when localized on the extraction septa and makes hands-on maintenance difficult. The losses in the ring have also proven problematic where they arise at aperture restrictions located in regions of the PS with less shielding. The loss distribution around the ring for CT is presented in Fig. 25(b). For this reason, the loss distribution in the machine induced by primary beam scattering on the SEH31 has been studied in detail [26]. During the CNGS run, the optics of the machine had to be changed by moving the position of the QKE to displace the losses induced by the extraction into better-shielded parts of the ring.

The rise time of the kickers creating the fast bump and sweeping the beam over the electrostatic septum is not particularly important for the extraction efficiency, even though there is no particle-free gap in the circulating beam, because the losses are dominated by the transverse shaving process. A 30 cm passive array of tungsten wires is installed directly upstream of the SEH31 to reduce the beam loss on the electrostatic septum by inducing scattering and reducing the beam density on the downstream septum foil. This technique is commonly applied; the device is typically called a diffuser or scatterer [34–38].

For ambitious future beam intensities [39], the induced radioactivity poses a potential limit for the total intensity throughput of the CERN accelerator complex, potentially limiting the beam availability for the LHC. If the induced radioactivity is too high, lengthy cool-down periods would be imposed

before personnel could safely work on and repair the machine, impacting the availability in the event of inevitable failures. It was such concerns that motivated the development of magnetic splitting to boost the extraction efficiency and reduce the activation of the machine.

4.2 Resonant multiturn extraction: magnetic splitting

A beam can be split in transverse phase-space by capturing particles around separate, stable fixed points, or islands, created by non-linear magnetic fields applied as the betatron tune is moved slowly through a resonance. Particles captured around the stable fixed points are drawn away from the core of the beam by careful manipulation of the tune to physically split the beam. The concept of magnetic splitting removes the need to mechanically shave the beam on an electrostatic septum and can reduce the beam lost in the extraction process. In addition, the optical parameters of the extracted island are constant, turn by turn, at the extraction point and, as a consequence, the matching to the downstream machine can be better controlled with respect to shaving [22]. The extraction technique has been pioneered at CERN over the last two decades and has recently become the operational technique to extract proton beams from the PS destined for the SPS fixed-target physics programme [20, 40, 41]. Many splitting experiments have taken place in the PS over the years for different-order resonance crossings, including the third-order unstable resonance, as well as the fourth- and fifth-order stable resonances, splitting the beam into three, five, and six beamlets, respectively [42]. The fourth-order configuration was identified as operationally relevant as the splitting mechanism to replace CT and pursued further at CERN under the name 'multiturn extraction' (MTE). Readers may be particularly interested by the videos of the different splitting and capture processes collected on the CERN MTE team's website [43].

4.2.1 Non-linear dynamics: splitting and capture

A detailed treatment of non-linear dynamics is beyond the intended scope of this contribution, nevertheless, a very brief overview will be given to help introduce the extraction concept. For more information on the subject, the reader is invited to read further, e.g., Refs. [44, 45]. The past proceedings of the advanced accelerator physics courses of the CERN Accelerator School provide a particularly relevant resource on the subject, e.g., Ref. [46]. Non-linear dynamics is essentially concerned with the solution of the following equation of motion:

$$\frac{\mathrm{d}^2\bar{X}}{\mathrm{d}\mu^2} + Q^2\bar{X} = -Q^2\beta^{3/2}\frac{\Delta B(\bar{X},\mu)}{B\rho}\,, \tag{16}$$

where $(\bar{X}, \mu)$ are the canonical phase-space variables of normalized displacement and phase advance of a charged particle along the reference trajectory of the synchrotron, β is the betatron function, Q the tune, and $B\rho$ its magnetic rigidity. Many mathematical tools exist to help solve this differential equation, such as the Hamiltonian, Taylor maps, and Lie algebra, along with perturbation theory, normal form analysis, etc. This well-known equation of motion represents driven simple harmonic motion, where the source of the driving term on the right-hand side consists of linear (dipole and quadrupole) imperfections, as well as non-linear imperfections (sextupole, octupole, and so on), which can be expanded in terms of the magnetic multipole coefficients $b_n(\mu)$ present in the synchrotron:

$$\frac{\Delta B(\bar{X},\mu)}{B\rho} = \frac{B_0}{B\rho}\Big[b_0(\mu) + \beta^{1/2}b_1(\mu)\bar{X} + \beta b_2(\mu)\bar{X}^2 + \beta^{3/2}b_3(\mu)\bar{X}^3 + \ldots\Big]\,. \tag{17}$$

It is these driving terms that perturb the circular particle trajectories in normalized phase-space; they become quickly significant to the beam stability when the frequency of the multipole terms drive a resonant condition with the tune of the synchrotron.

Solving the equation of motion numerically with a few lines of computer code will suffice here for the level of insight required. This can be achieved by constructing a Hénon-like map to model the

transverse dynamics, as is detailed at length in Ref. [47]. One can track a particle in the horizontal plane from turn n to turn $n+1$ by approximating the motion of each turn by a tune-dependent linear one-turn map followed by a thin lens multipole kick,

$$\begin{pmatrix} \bar{X} \\ \bar{X}' \end{pmatrix}_{n+1} = R_{1-\text{turn}}(2\pi Q) \begin{pmatrix} \bar{X} \\ \bar{X}' + K_2\bar{X}^2 + K_3\bar{X}^3 + \ldots \end{pmatrix}_n , \tag{18}$$

where K_n are integrated normalized multipole strengths and $K_0 = K_1 = 0$ in this case. Of course, for MTE, the imperfections are voluntarily applied to drive the splitting. Different normalizations are commonly applied in accelerator physics; one will see special adimensional normalized co-ordinates applied in most of the MTE literature to better elucidate the ratio of the sextupole and octupole strengths.

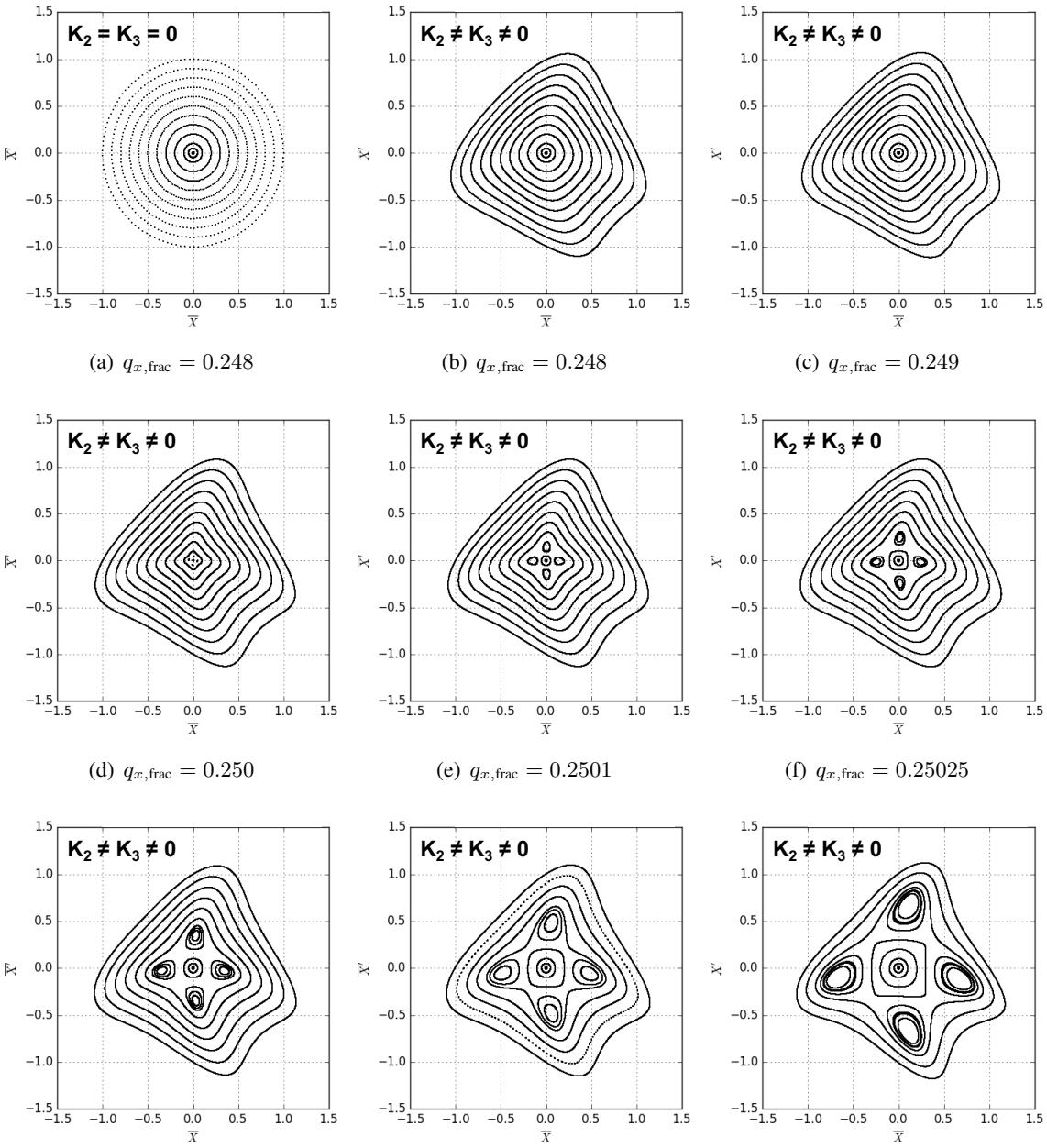
The concept employed at the PS is demonstrated in Fig. 20 for a horizontal fourth-order resonance crossing in the presence of sextupole and octupole magnetic fields, where 12 particles are initialized and tracked over 1000 turns. In this example, the dynamics is linear at small amplitudes before the tune is swept up through the resonance and four stable fixed points are created in addition to the one at the origin. The fixed points move away from the core as the tune distance above the resonance is increased, until two distinct regions in phase-space are formed: the core and four islands. The phase-space plots in Fig. 20 were created by iterating the map at different values of Q, K_2, and K_3. Islands are formed at the fourth-order resonance crossing even without octupole fields, i.e., with $K_3 = 0$; however, the introduction of octupoles provides another degree of freedom to better adjust the splitting and capture process, allowing better control of the tune and size of the islands, along with many other important parameters. The differences in the size, shape, and position of the islands with $K_3 = 0$ are demonstrated in Fig. 21. A Hamiltonian description of the process can be found in Ref. [48], where it is used to describe analytically how the position and size of the stable fixed points change as a function of the multipole strengths, as well as give an understanding of the non-linear contributions to the tune. Further details of the numerical simulations carried out and benchmarked to design and commission the MTE at the CERN PS can be found in Refs. [49–53].

The uniformity of the extracted spill depends directly on how evenly the beam intensity, which is initially located in the beam core, is shared and re-distributed between the core and the islands during the splitting process. The splitting efficiency is used as a figure-of-merit to parametrize the performance of MTE and is defined as

$$\eta_{\text{MTE}} = \frac{\langle I_{\text{islands}} \rangle}{I_{\text{total}}} , \tag{19}$$

where $\langle I_{\text{islands}} \rangle$ and I_{total} stand for the average intensity in the islands and the total beam intensity, respectively. The requirements of the SPS specify that the spill must be uniform to $\pm 1\%$ of the total intensity per turn, i.e., $\eta_{\text{MTE}} = (20 \pm 1)\%$ for a splitting over five turns. Unlike CT, the uniformity of the spill is predetermined by the splitting efficiency before the actual fast extraction process is initiated. The size of the beam core (horizontal emittance) before splitting has a significant impact on the splitting efficiency [20]. Excitation from the transverse damper is applied to the beam to increase the emittance of the core during resonance crossing and help improve the intensity distribution among the beamlets formed during the splitting and capture process. The role played by transverse excitation in the splitting and capture process is a topic of active research [53].

After acceleration and on the flat-top at 14 GeV/c, the non-linear sextupole and octupole fields are applied before the tune is swept adiabatically (slowly) through the fourth-order resonance at $4Q_x = 25$ to start the splitting and capture process. To minimize the perturbation to the tune from the longitudinal plane, the chromaticity is held close to zero and the momentum spread of the beam reduced by significantly reducing the RF voltage. The splitting efficiency is significantly degraded without reducing the momentum spread in the beam; in this way, most particles cross resonance together. Careful control of the machine is needed throughout the splitting process and especially in the presence of the non-linear fields applied. The stability and reproducibility of the machine was one of the main challenges

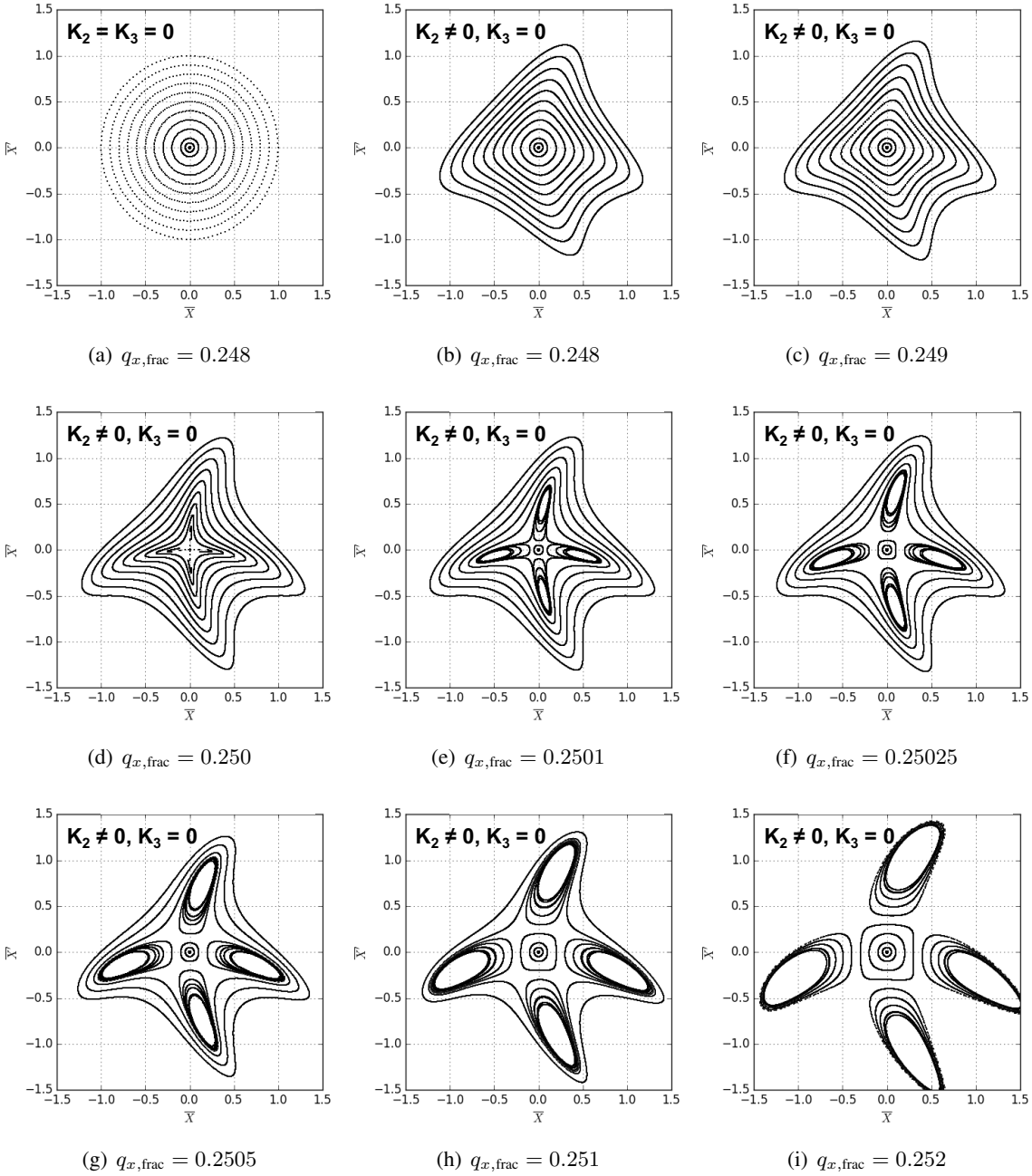


(a) $q_{x,\text{frac}} = 0.248$ (b) $q_{x,\text{frac}} = 0.248$ (c) $q_{x,\text{frac}} = 0.249$

(d) $q_{x,\text{frac}} = 0.250$ (e) $q_{x,\text{frac}} = 0.2501$ (f) $q_{x,\text{frac}} = 0.25025$

(g) $q_{x,\text{frac}} = 0.2505$ (h) $q_{x,\text{frac}} = 0.251$ (i) $q_{x,\text{frac}} = 0.252$

Fig. 20: Phase-space portraits of 12 particles tracked over 1000 turns by a Hénon-like mapping in the presence of a thin lens sextupole and octupole ($K_2 \neq K_3 \neq 0$) as the tune is swept across the horizontal fourth-integer resonance from $q_{x,\text{frac}} = 0.248$ to 0.252. The multipole is initially turned off (a) before being turned on and held constant as the tune is swept from (b) to (i).



(a) $q_{x,\text{frac}} = 0.248$ (b) $q_{x,\text{frac}} = 0.248$ (c) $q_{x,\text{frac}} = 0.249$

(d) $q_{x,\text{frac}} = 0.250$ (e) $q_{x,\text{frac}} = 0.2501$ (f) $q_{x,\text{frac}} = 0.25025$

(g) $q_{x,\text{frac}} = 0.2505$ (h) $q_{x,\text{frac}} = 0.251$ (i) $q_{x,\text{frac}} = 0.252$

Fig. 21: Phase-space portraits of 12 particles tracked over 1000 turns by a Hénon-like mapping in the presence of a thin lens sextupole ($K_2 \neq 0$, $K_3 = 0$) as the tune is swept across the horizontal fourth-integer resonance from $q_{x,\text{frac}} = 0.248$ to 0.252. The multipole is initially turned off (a) before being turned on and held constant as the tune is swept from (b) to (i).

that the MTE team faced in realizing the novel extraction scheme. Once the islands are formed, they are rotated in phase-space by manipulation of the multipole strengths to optimize the physical separation presented to the extraction septum [52]. Before extraction, the beam is debunched and partially recaptured at 200 MHz, as previously done for CT.

4.2.2 Extraction process

A detailed overview of the implementation of the MTE equipment installed in the PS, including the dedicated multipole magnets and extraction equipment, can be found in Ref. [20]. During the splitting process, the beam is separated into two distinct regions in horizontal phase-space: the core and the islands, which circulate on separate closed orbits in the machine. The core of the beam sits on the reference closed orbit, as it did before the splitting was initiated, and the islands circulate on another closed orbit, which loops around the machine four times before closing on itself. The beam captured in the islands represents a continuous entity (if the beam is debunched) that also snakes around the machine, as depicted by the well-known plot in Fig. 22. One would observe a test particle trapped in the islands at a given location in the ring jump from one island to the next as it returns and is rotated by the fractional tune $q_{x,\text{frac}} \approx 0.25$ each turn.

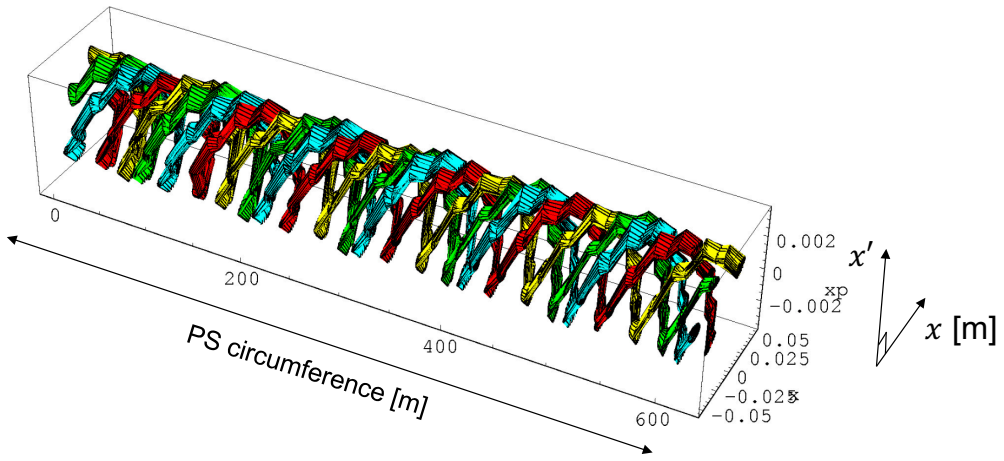
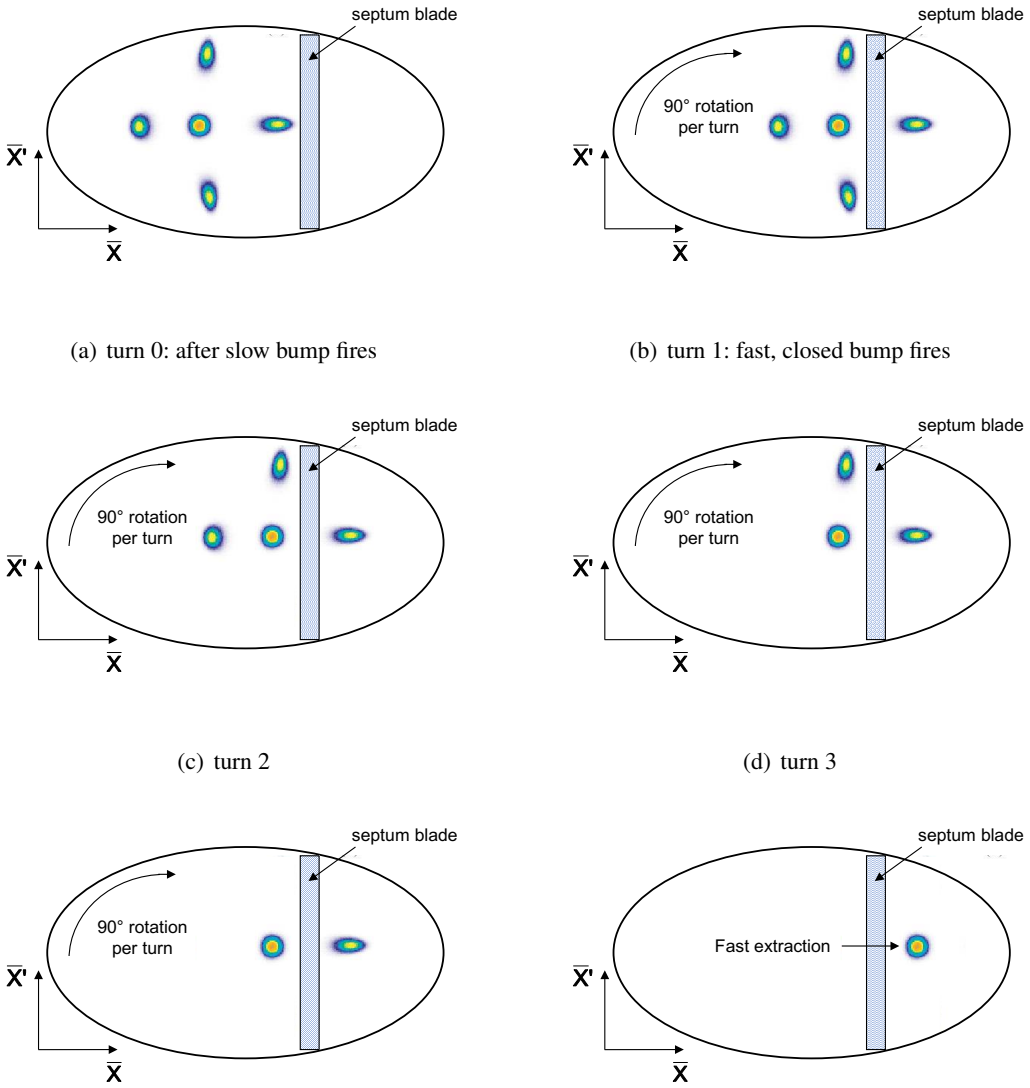


Fig. 22: Islands looping and rotating with the phase advance around the circumference of the PS four times, connected continuously into one single entity, where each island, or turn, is represented by a different colour [20].

After splitting and capture, a slow bump brings the circulating beams close to the septum before the extraction is initiated by a fast, closed bump pushing one island over the septum blade, as shown in Fig. 23. Most notably, an electrostatic septum is not required and a single magnetic extraction septum can be used, which in this case is the SMH16 mentioned previously. The tune rotates the rest of the beam, snaking around the machine into the septum aperture during the subsequent four turns, which parallels the CT extraction process described previously. A classic fast extraction is employed to extract the core on the fifth turn. Figure 23 also demonstrates clearly how the optical parameters of the island pushed over the septum blade stay constant throughout the extraction. The extraction kicker scheme used for MTE is summarized in Fig. 24.



(a) turn 0: after slow bump fires

(b) turn 1: fast, closed bump fires

(c) turn 2

(d) turn 3

(e) turn 4

(f) turn 5: fast extraction

Fig. 23: Turn by turn phase-space diagram of the MTE extraction process at the location of the extraction septum, SMH16.

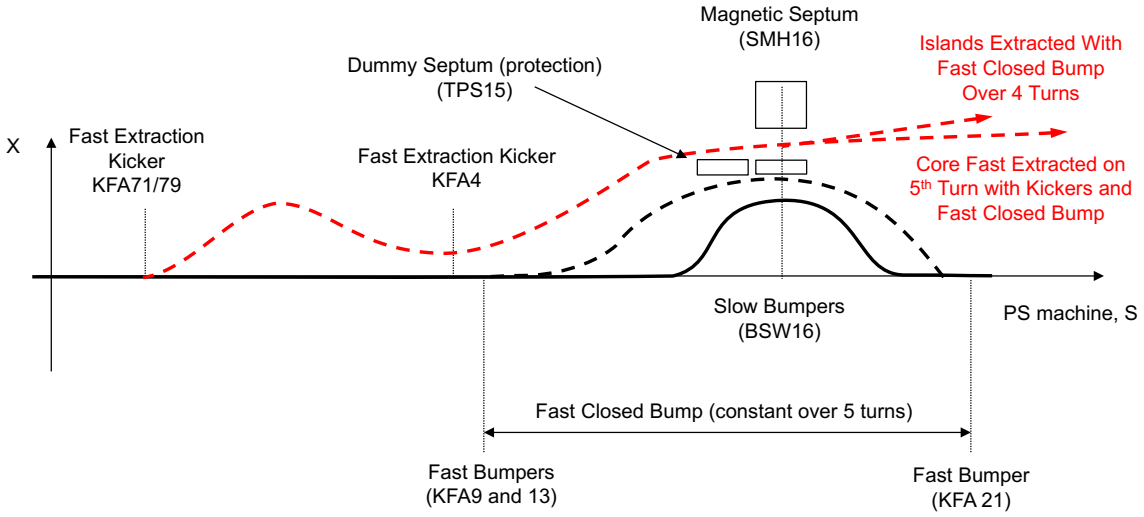


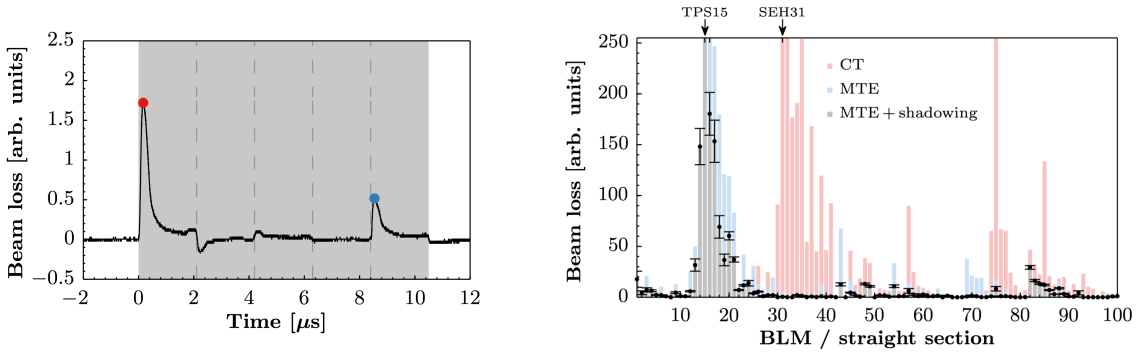
Fig. 24: Overview of the MTE extraction scheme at the CERN PS. The second fast kicker (KFA4) is used to correct the extraction trajectory of the core and minimize the difference with respect to the islands.

4.2.3 Beam loss

The MTE extraction process would be essentially loss-free in the presence of a longitudinal particle-free gap in the circulating beam because the transverse splitting and capture losses can be kept very low. Unfortunately, the beam must be debunched around the circumference of the PS to satisfy the requirements of the SPS and, as a result, beam is lost on the extraction septum as the beam is swept across its blade during the rise of the fields of the fast-pulsed kicker systems, both when the fast, closed bump is fired to extract the islands and when the core is fast-extracted. In contrast with CT, the rise times of the kicker systems involved are critical in minimizing the beam loss. To protect the SMH16, a water-cooled copper absorber was installed in the upstream straight section to absorb the majority of the beam energy that would otherwise have directly impacted the septum blade. The absorber, which resembles a passive septum blade and is commonly known as the 'dummy' septum or the TPS15, is the location in the ring at which most of the losses occur during the MTE process. Further details of the TPS15 can be found in Ref. [54]. The typical beam loss signal measured close to the extraction region is shown in Fig. 25(a). The difference in the rise times of the fast, closed bump (islands) and fast extraction (core) systems is evident in the magnitude of the loss signals.

The integrated beam loss monitor signals recorded during extraction around the PS ring for CT and MTE are compared in Fig 25(b). The MTE losses are far more localized close to the extraction region and, in particular, the TPS15, which allows the beam-induced activation of the machine to be more effectively shielded. The losses on aperture bottlenecks in the ring created during CT by beam scattering from the SEH31 largely disappear. Radiation survey measurements made approximately 30 hours after operation of CT or MTE reflect the change in the prompt extraction beam loss monitor profile, as presented in Fig. 26.

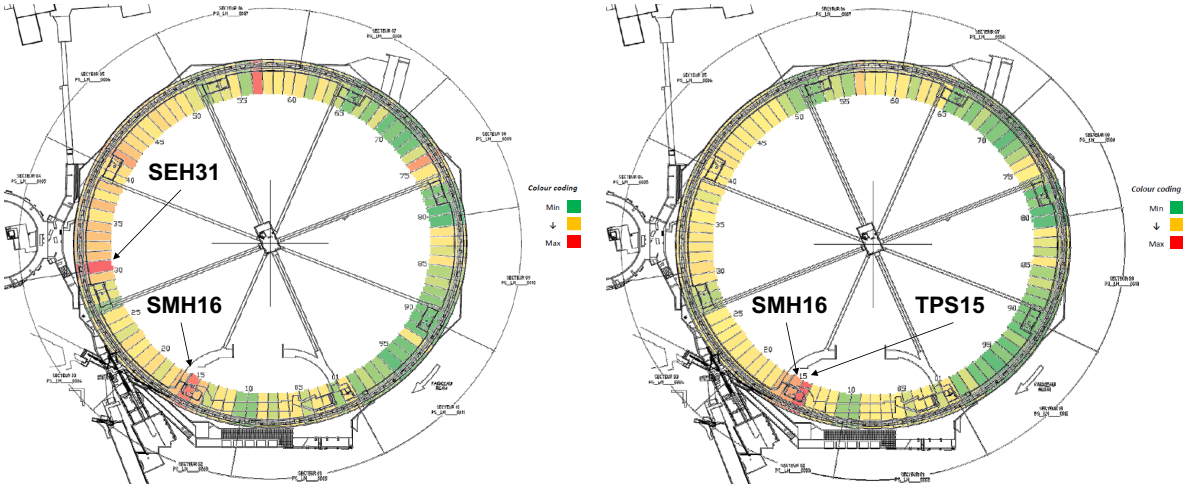
The implementation of MTE at the PS has improved the extraction losses by about a factor of three compared with CT, reducing from approximately 6% to less than 2%, despite the beam being debunched. The extraction efficiency for CT is compared with that for MTE during 2015 in Fig. 27, where the extraction efficiency is shown to be relatively insensitive to the beam intensity. Ways to reduce the fast, closed bump rise time and to implement a particle-free gap, perhaps via the barrier bucket technique [56], are being considered to further reduce the extraction losses.



(a) At TPS15 for MTE: red, islands; blue, core

(b) At extraction around the PS ring and compared with CT

Fig. 25: Beam loss monitor (BLM) signals during extraction for MTE: the beam is swept across the septum blade as the fast-pulsed magnets fire [52].



(a) After CT operation: end of 2014

(b) After MTE operation: end of 2015

Fig. 26: Radioactive dose distribution around the PS ring with MTE measured approximately 30 hours after operation [55].

4.2.4 Operational challenges

The MTE splitting process is resonant and sensitive to even very small changes in the magnetic reproducibility of the machine. An example of a bad spill is shown in Fig. 28, where the intensity over the five turns is shown alongside a beam profile measurement made using a wire-scanner device. The poor splitting efficiency in this case is clearly observed by the overpopulation of the core shown in both the spill and the profile measurements.

An important step in making MTE operational at the PS was to understand the different factors affecting the stability of the machine. In particular, periodic fluctuations in the splitting efficiency could be correlated to low frequency noise (at 5 kHz) of unsynchronized power converters driving the magnetic circuits used to control the tune and chromaticity of the PS. The resulting tune modulation caused oscillations in the splitting efficiency with a time period of the order of tens of minutes. With the power converters synchronized and the ripple reduced, the machine reproducibility and splitting efficiency was significantly improved. Detailed beam dynamics simulations were also carried out to understand the phenomenon and the frequency dependence of the tune ripple. Further details of this work can be found

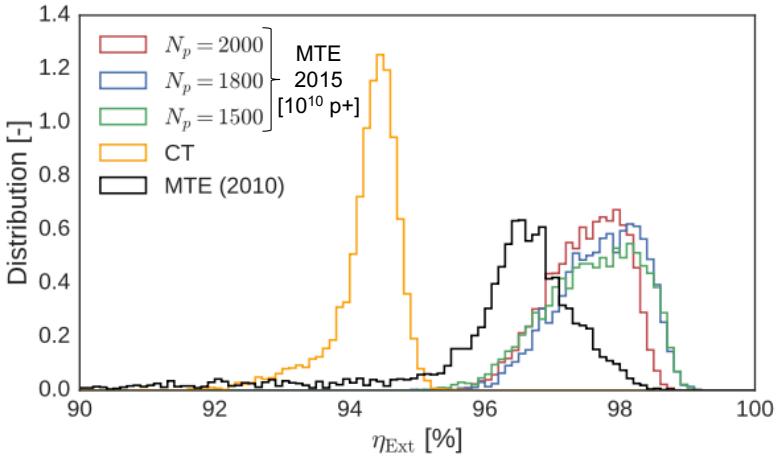
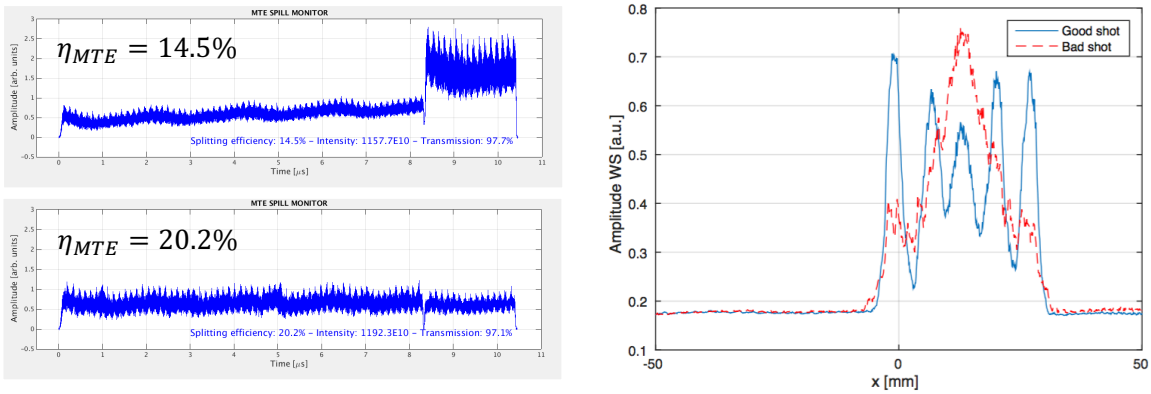


Fig. 27: MTE extraction efficiency compared with CT [40]



(a) Spill intensity as function of time: upper, bad shot; lower, good shot.

(b) Wire-scanner profile comparison for the same two shots.

Fig. 28: Example of a good spill compared with a bad spill following a change in machine reproducibility [40]

in Ref. [53]. It should be noted that the application of transverse damper excitation during splitting is imperative to increase the capture probability during island formation and reduce its sensitivity to the initial horizontal emittance. One major drawback of MTE at CERN is that the beam emittance in the plane of extraction is not significantly reduced compared with the initial emittance, as it is in the case of the shaving employed for CT. Without the possibility of reducing the vertical emittance through phase-space exchange in the transfer line between the PS and SPS, the larger vertical emittance seen by the SPS with MTE results in poorer transmission. Work is actively ongoing to create and preserve smaller vertical emittance throughout the CERN accelerator chain and throughout the MTE cycle in the PS. The future charge-exchange injection scheme foreseen in the PS booster will help reduce the vertical emittance in the longer term. There are many other operational issues too detailed for this paper that also required particular attention to realize the MTE scheme, e.g., mechanical aperture, operation of the PS in the presence of a dummy septum, impact of the slow bump on the tune, and turn by turn trajectory differences, to mention just a few. Further details can be read in the MTE Design Report [20].

Acknowledgement

The vast majority of the material in this contribution was provided by colleagues at CERN, in particular those from the ABP, ABT, and OP groups and especially the MTE team, with D. Cotte, B. Goddard, M. Giovannozzi, S. Gilardoni, A. Huschauer, and G. Sterbini requiring notable acknowledgement. A special, personal, thanks goes to A. Huschauer and F. Velotti for taking the time to review these proceedings.

References

[1] M. Benedikt and F. Zimmerman, Status of the future circular collider study, 25th Russian Particle Accelerator Conf., Saint Petersburg, 2016, CERN-ACC-2016-0331, p. TUYMH01.

[2] O. Brüning *et al.* (eds.), LHC design report v. 1: the LHC Main Ring, CERN-2004-003-V-1 (CERN, Geneva, 2004), https://doi.org/10.5170/CERN-2004-003-V-1.

[3] R. Bossart *et al.*, *IEEE Trans. Nucl. Sci.* **16** (1969) 286, https://doi.org/10.1109/TNS.1969.4325235.

[4] F.M. Velotti, Ph.D. thesis, Ecole Polytechnique Lausanne, CERN-THESIS-2017-041, 2017.

[5] M.A. Fraser *et al.*, Experimental results of crystal-assisted slow extraction at the SPS, 8th International Particle Accelerator Conf., Copenhagen, 2017, p. MOPIK048.

[6] V. Forte *et al.*, Beam-based kicker waveform measurements using long bunches, 8th International Particle Accelerator Conf., Copenhagen, 2017, p. MOPIK042.

[7] F. Velotti *et al.*, Performance studies of the SPS beam dump system for HL-LHC beams, CERN 5th International Particle Accelerator Conf., Dresden, 2014, CERN-ACC-2014-0239, p. THPME069.

[8] C. Wiesner *et al.*, LHC beam dump performance in view of the high luminosity upgrade, 8th International Particle Accelerator Conf., Copenhagen, 2017, p. WEPIK033.

[9] M.J. Barnes and B. Goddard, Considerations on a new fast extraction kicker concept for SPS, CERN-sLHC-Project-Note-0018, (CERN, Geneva, 2010).

[10] S. Gilardoni and D. Manglunki (eds.), Fifty years of the CERN Proton Synchrotron: vol. 1, CERN-2011-004 (CERN, Geneva, 2011), https://doi.org/10.5170/CERN-2011-004.

[11] K. Wille, *The Physics of Particle Accelerators: An Introduction* (Oxford University Press, Oxford, 2000).

[12] H. Wiedemann, *Particle Accelerator Physics*, 3rd ed. (Springer, Berlin, 2007).

[13] `http://mad.web.cern.ch/mad`, last accessed January 18th 2018.

[14] V. Mertens, Protons on the doorstep of the LHC, *CERN Courier*, 1 March 2005. `http://cerncourier.com/cws/article/cern/29275`.

[15] M. Benedikt *et al.* (eds.), LHC design report v. 3: the LHC injector chain, CERN-2004-003-V-3 (CERN, Geneva, 2004), https://doi.org/10.5170/CERN-2004-003-V-3.

[16] J. Borburgh *et al.*, Upgrade of the CERN SPS extraction protection elements TPS, 6th International Particle Accelerator Conf., Richmond, Va., 2015, CERN-ACC-2017-0020, p. WEPMN068.

[17] B. Goddard, Fast extraction from SPS LSS4 for the LHC and NGS projects, SL-Note-98-066-SLI, (CERN, Geneva, 1998).

[18] V. Kain *et al.*, High intensity commissioning of SPS LSS4 extraction for CNGS, 22nd Particle Accelerator Conf., Albuquerque, N.M., 2007, CERN-AB-2007-048, p. TUPAN096, https://doi.org/10.1109/PAC.2007.4440837.

[19] W. Bartmann *et al.*, Beam transfer to the FCC-hh collider from a 3.3 TeV booster in the LHC tunnel, IPAC2015, Richmond, Va., CERN-ACC-2015-012, p. THPF089.

[20] M. Giovannozzi (ed.), The CERN PS multi-turn extraction based on beam splitting in stable islands of transverse phase space: design report, CERN-2006-011 (CERN, Geneva, 2006), https://doi.org/10.5170/CERN-2006-011.

[21] K.H. Reich, Decaturn beam extraction for synchrotrons, accelerator department internal report BNL-AADD-100, BNL, N.Y., 1966.

[22] R. Cappi and M. Giovannozzi, Computation of betatron mismatch and emittance blow-up for multi-turn extraction, CERN/PS 2002-083 (AE) (CERN, Geneva, 2002).

[23] V.D. Rudko *et al.*, *Instrum. Exp. Tech.* **59** (2016) 325, https://doi.org/10.1134/S0020441216020330.

[24] M. Gyr and E.B. Vossenberg, Half-integer fast resonant extraction with quasi rectangular spill, 3rd European Particle Accelerator Conf., Berlin, 1992, p.1501.

[25] C. Bovet *et al.*, The fast shaving ejection for beam transfer from the CPS to the CERN 300 GeV machine, 5th IEEE Particle Accelerator Conf., San Francisco, 1973 [IEEE Trans. Nucl. Sci. **20** (1973) 438], https://doi.org/10.1109/TNS.1973.4327144.

[26] J. Barranco Garcia and S. Gilardoni, *Phys. Rev. ST Accel. Beams* **14** (2011) 030101. https://doi.org/10.1103/PhysRevSTAB.14.030101.

[27] D. C. Fiander *et al.*, A modulated fast bump for the CPS continuous transfer, Accelerator Conf., Chicago, 1977 [IEEE Trans. Nucl. Sci. **24** (1977) 1340], https://doi.org/10.1109/TNS.1977.4328938.

[28] L. Sermeus, MTE/CT ABT systems: past and present, ABT Engineering Forum, CERN, Geneva, 2018, `https://indico.cern.ch/event/685066/`.

[29] D. Cotte, MTE at PS: considerations from OP, MTE Internal Review, CERN, Geneva, 2015.

[30] A. Franchi *et al.*, Trajectory correction in the transfer line TT2-TT10 for the continuous transfer (CT), CERN-AB-Note-2008-005, (CERN, Geneva, 2007).

[31] L.R. Evans, A phase plane exchange section for the SPS antiproton injection beam-line, CERN-SPS-DI-MST-80-2 (CERN, Geneva, 1980).

[32] M. Giovannozzi, Comparison of extraction losses for present CT and novel multi-turn extraction, APC Meeting, 2005.

[33] M. Giovannozzi, Follow-up of comparison of extraction losses for present CT and novel multi-turn extraction, APC Meeting, 2005.

[34] C. Germain *et al.*, Technical developments of the CERN electrostatic program, 2nd International Symposium on Insulation of High Voltages in Vacuum, Boston, Mass., 1966, p. 279.

[35] M. Thivent, Developpements lies a la construction des deflecteurs electrostatiques, PS/PSR/Note 83-8, (CERN, Geneva, Switzerland), 1983.

[36] A. Durand, Efficacite d'un reseau de fils place devant la plaque du septum du transfert continue à 12 GeV/c, University de Rennes, France, 1974.

[37] A. Durand, Sur la diffusion de protons de $p = 400$ GeV/c par un reseau de fils place devant la plaue d'un septum electrostatique, University de Rennes, France, 1974.

[38] B. Goddard *et al.*, The use of a passive scatterer for SPS slow extraction beam loss reduction, Proceedings IPAC'17, Copenhagen, 2017, p. MOPIK044.

[39] A. Golutvin *et al.*, A facility to search for hidden particles (SHiP) at the CERN SPS, CERN-SPSC-2015-016 (SPSC-P-350), (CERN, Geneva, 2015).

[40] S. Abernethy *et al.*, *Phys. Rev. Accel. Beams* **20** (2017) 014001. https://doi.org/10.1103/PhysRevAccelBeams.20.014001.

[41] A. Huschauer *et al.*, *Phys. Rev. Accel. Beams* **20** (2017) 061001. https://doi.org/10.1103/PhysRevAccelBeams.20.061001.

[42] A. Franchi *et al.*, *Phys. Rev. ST Accel. Beams* **12** (2009) 014001. https://doi.org/10.1103/PhysRevSTAB.12.014001.

[43] `https://ab-project-mte.web.cern.ch`, last accessed January 18th 2018.

[44] D.A. Edwards and M.J. Syphers, *An Introduction to the Physics of High Energy Accelerators*, (Wiley, New York, 2003).

[45] T.W.B. Kibble and F.H. Berkshire, *Classical Mechanics*, 5th ed. (Imperial College Press, London, 2004).

[46] W. Herr (ed.), CAS–CERN Accelerator School: Advanced Accelerator Physics Course, Trondheim, Norway, 18–29 August 2013, edited by W. Herr, CERN-2014-009 (CERN, Geneva, 2014), https://doi.org/10.5170/CERN-2014-009.

[47] A. Bazzani *et al.*, A normal form approach to the theory of nonlinear betatronic motion, CERN-94-02 (CERN, Geneva, 1994), https://doi.org/10.5170/CERN-1994-002.

[48] M. Giovannozzi and P. Scaramuzzi, Nonlinear dynamics studies at the CERN Proton Synchrotron: precise measurements of islands parameters for the novel multi-turn extraction, EPAC'04, Lucerne, 2004, p. 1861.

[49] R. Cappi and M. Giovannozzi, *Phys. Rev. Lett.* **88** (2002) 104801. https://doi.org/10.1103/PhysRevLett.88.104801.

[50] R. Cappi and M. Giovannozzi, *Nucl. Instrum. Meth. A* **519** (2004) 442. https://doi.org/10.1103/PhysRevLett.88.104801.

[51] S.Gilardoni *et al.*, *Nucl. Instrum. Meth. A* **561** (2006) 249. https://doi.org/10.1016/j.nima.2006.01.021.

[52] A. Huschauer, Ph.D. thesis, Technische Universität Wien, 2016, CERN-THESIS-2016-061.

[53] A. Huschauer *el al.*, Transverse beam splitting made operational: key features of the multiturn extraction at the CERN Proton Synchrotron, 57th ICFA Advanced Beam Dynamics Workshop on High-Intensity and High-Brightness Hadron Beams, 2016 [Phys. Rev. Accel. Beams **20** (2017) 061001].

[54] C. Bertone *et al.*, Studies and implementation of the PS dummy septum to mitigate irradiation of magnetic septum in straight section 16, CERN-ACC-2014-0043 (CERN, Geneva, 2008).

[55] G. Dumont, RP survey of PSB, PS and SPS, MSWG Meeting 2016 #2, (CERN, Geneva, 2016).

[56] H. Damerau, Ph.D. thesis, Technische Universität Damstadt, 2005, CERN-THESIS-2005-048.

Proceedings of the CAS–CERN Accelerator School: Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018)

Resonant Extraction

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

Abstract
Resonant extraction has its roots in the early cyclotrons where the emphasis was on increasing inter-turn separation. High energy physics (HEP) synchrotrons followed in the 1960s, but this time the emphasis was on extending the spill time. This lecture will concentrate on the 1/3rd integer resonance extraction for synchrotrons that is widely used in medical machines.

Keywords
Resonant extraction; slow extraction; Kobayashi Hamiltonian.

1 Historical note

Resonant extraction has its roots in the early cyclotrons. These machines used scattering from an internal target to obtain an external proton beam, but the consequent activation and loss of intensity were strong incentives for change. In 1951, James Tuck and Lee Teng (Chicago cyclotron) suggested a resonant scheme that increased the amplitude of radial oscillations sufficiently in a single turn to clear the mouth of a magnetic extraction septum. A quantitative analysis that confirmed the principle of what was then known as the 'Peeler-Regenerative Beam Extraction Method' was published by Kenneth Le Couteur (Liverpool cyclotron) in 1953 and the first successful extraction by this technique was demonstrated in the Liverpool cyclotron in 1954. This was reported by Albert Crewe and Le Couteur in 1955. At this stage, the emphasis was on increasing inter-turn separation.

High energy physics (HEP) synchrotrons in the 1950s were equipped with full-aperture, fast kickers for extraction. Single-turn extraction was efficient, but for fixed-target experiments it overloaded the physics counters leading to dead time and lost data. In 1961, Hugh Hereward addressed this problem [1] and his proposal cites Tuck, Teng, and Le Couteur. At essentially the same time, C. L. Hammer and Lawrence Jackson Laslett published a similar paper [2]. In contrast to cyclotrons, the emphasis was now on extending the spill time and a new name 'slow extraction' came into general usage.

2 Overview

For extraction, cyclotrons and synchro-cyclotrons work with the integer or half-integer resonance. This is consistent with a need for a large inter-turn separation and a fast extraction. To complete the historical picture, it is noted that in the 1980s, this situation changed for high-intensity proton cyclotrons with the introduction of H minus extraction[I].

In contrast, the extraction from synchrotrons works with the half-integer or 1/3rd integer resonance. The half-integer resonance is usually associated with fixed-target physics and spills of a few milliseconds while the 1/3rd integer resonance is weaker and is used widely in medical machines for spills of a few seconds (i.e., a few 10^6 turns). Higher-order resonances are not used because the resonances are weaker, the transit times in the resonances are longer, and the extraction separatrices

I The introduction of H minus extraction was something of a revolution that made the IBA Cyclone 30 a *de facto* world standard for isotope production.

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-005.315

become too crowded making it difficult to ensure sufficient free aperture for the extraction equipment. Figure 1 shows some phase-space sketches to illustrate this last point. The present lecture will concentrate on the 1/3rd integer resonance.

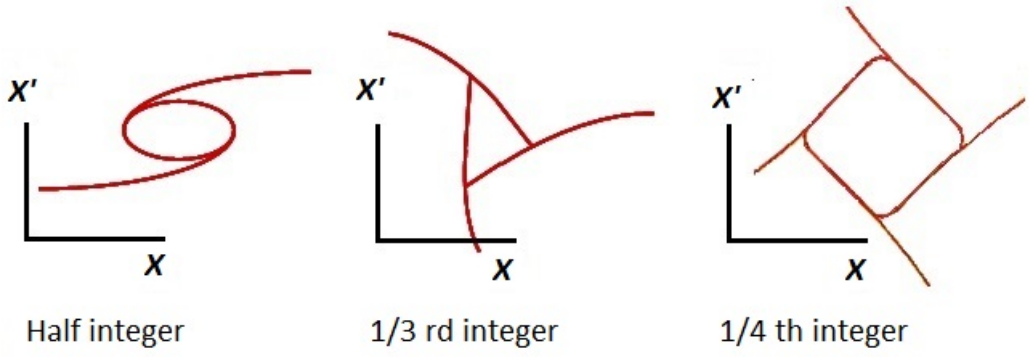


Fig. 1: Phase-space sketches of the half, 1/3rd, and 1/4th integer resonances

3 Phase space

The full beauty of resonant extraction is only visible in phase space. Figure 2 shows a 1/3rd integer extraction as designed for a medical machine. This diagram was generated by tracking single particles with progressively larger betatron amplitudes in the horizontal plane, but all with the same momentum (equivalent to the same horizontal position and tune). For the case shown, the relative momentum deviation with respect to the position of the resonance was about 0.001.

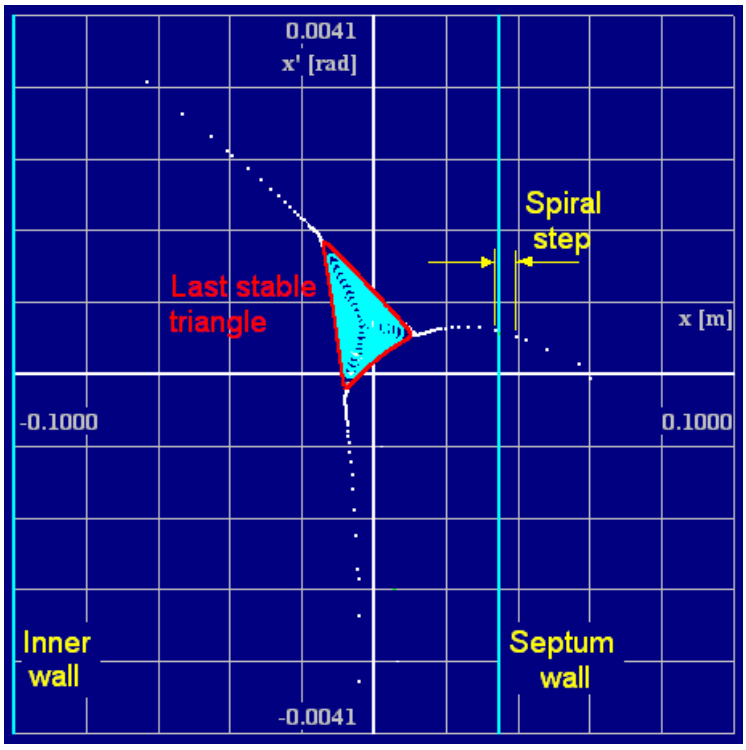


Fig. 2: Horizontal phase-space diagram of a 1/3rd integer extraction (annotated screen shot, WinAgile)

In the central region of Fig. 2, the beam has small betatron amplitudes and is stable. At the very centre, the invariants of the motion are unperturbed ellipses. At larger amplitudes, but still inside the

stable central region, the ellipses take on a triangular shape. In the outer region, the beam is unstable. The unstable part of the diagram is generated by one ion that 'locks' onto the 3 separatrices and gains amplitude turn after turn, while moving from one separatrix to the next.

The phase-space presentation has the advantage that the spiral step (inter-turn separation) and other geometric details can be determined to a high precision to facilitate the design of the extraction elements. By plotting this diagram for a range of momenta close to the resonance, it is possible to see how the 'last stable triangle' and the separatrices develop as a function of the tune separation from the resonance. In most cases, it is sufficient to consider just two limiting cases that define the beam, or the part of the beam that is being extracted. Before continuing to the quantitative analysis, it is useful to look at some other descriptive diagrams that describe slow extraction.

4 Other descriptive views of extraction

4.1 A 'strip' spill

The Steinbach diagram (attributed to C. Steinbach, CERN) has the advantages of being more holistic and showing the inclined threshold between stability and instability. The diagram plots the 'waiting' beam, the resonance and the spill using the radial position versus the effective, normalized amplitude of the betatron motion, $A_x = \sqrt{(E_x/\pi)}$, where E_x is the ion's motion invariant sometimes referred to as the single-particle emittance.

In the first example shown in Fig. 3, a wide momentum stack is being accelerated by a betatron core into the resonance. The tune of the resonance is usually close to the centre of the chamber and the chromaticity is finite and positive. Figure 2 corresponds to a snapshot of the phase space at an arbitrary point on the inclined stability boundary between the points 'A' and 'B' which define the part of the beam that is being extracted and correspond to the limiting cases referred to the last section. In this case, the phase-space snapshot would change significantly with position between 'A' and 'B'.

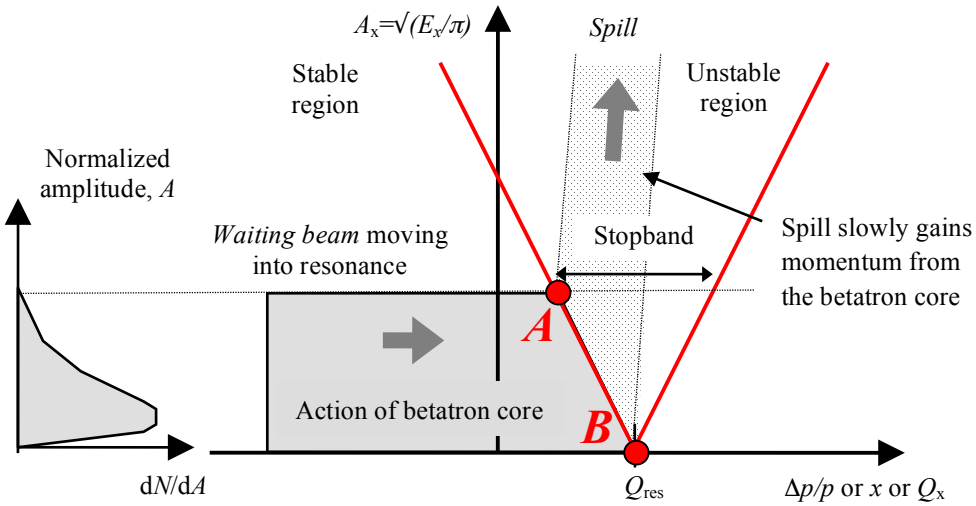


Fig. 3: An example of a Steinbach diagram for a momentum-driven slow extraction

In the second example shown in Fig. 4, a narrow momentum stack is being driven into the unstable region of the resonance by RF excitation at the revolution frequency causing amplitude growth. The limiting points denoted by 'A' and 'B' now define the top and bottom of the waiting beam and hence also the limits of the spill being extracted. As before, Fig. 4 corresponds to a snapshot of the phase space

somewhere between ‘A’ and ‘B’. In this case, the phase-space snapshot would be quasi-constant across the space from ‘A’ to ‘B’.

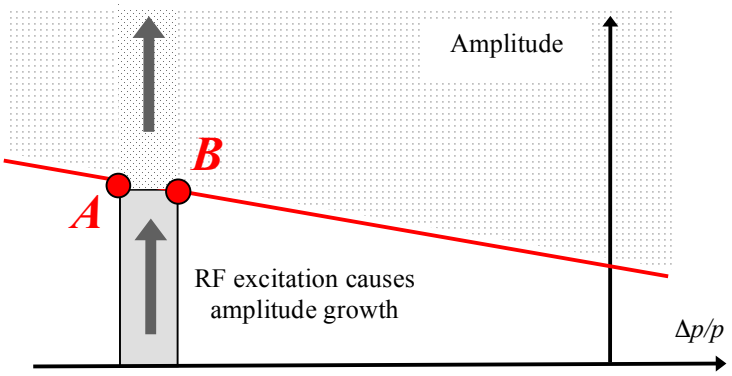


Fig. 4: Steinbach diagram of an amplitude-driven slow extraction

4.2 Near real-world view

The pictorial presentation in Fig. 5 is closer to the real-world situation of the momentum-driven extraction shown in Fig. 3. The waiting beam on the left is created by multi-turn injection in the inner half of the aperture well away from the 1/3rd integer resonance. After trapping and acceleration to the extraction energy, a betatron core is used to smoothly accelerate the waiting beam into the resonance, which is now positioned at the centre of the chamber. The particles ‘locked’ in the resonance grow in radial amplitude in steps of 3 turns. Starting with small steps, they extend outwards typically over several hundred turns before reaching the electrostatic extraction septum (ES). The beam enters the ES with a final step of ~10 mm and is kicked onto the extraction trajectory.

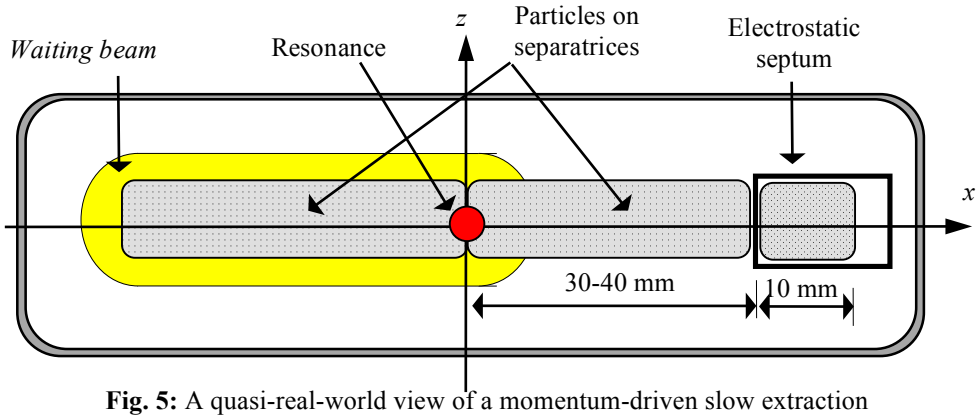


Fig. 5: A quasi-real-world view of a momentum-driven slow extraction

5 Sextupole fields

The radial 1/3rd integer or third-order resonance is driven by normal sextupole fields and can be treated as a perturbation to the linear machine. The transverse fields in a normal sextupole magnet are well known and can be written as,

$$B_x(x,z) = -6B_3xz \quad \text{and} \quad B_z(x,z) = -3B_3(x^2 - z^2)\ , \tag{1}$$

where

$$B_3 = -\frac{1}{6}\left(\frac{\mathrm{d}^2 B_z}{\mathrm{d}x^2}\right)_0 . \tag{2}$$

Thus, for a positively charged ion in an anticlockwise ring, the kicks given by a sextupole will be,

$$\Delta x' = \frac{B_z \ell_S}{|B\rho|} = \frac{1}{2}\frac{\ell_S}{|B\rho|}\left(\frac{\mathrm{d}^2 B_z}{\mathrm{d}x^2}\right)_0 \left(x^2 - z^2\right) = \frac{1}{2}\ell_S k'\left(x^2 - z^2\right), \tag{3}$$

$$\Delta z' = -\frac{\ell_S}{|B\rho|}\left(\frac{\mathrm{d}^2 B_z}{\mathrm{d}x^2}\right)_0 xz = -\ell_S k' xz \ , \tag{4}$$

where k' is the 'normalized sextupole gradient', $k' = \frac{1}{|B\rho|}\left(\frac{\mathrm{d}^2 B_z}{\mathrm{d}x^2}\right)_0$. (5)

6 Normalized coordinates

It is usual at this stage to introduce 'normalized coordinates'. Starting from the real-world position and angle of a test ion expressed with the usual Twiss parameters,

$$y(s) = C\beta^{1/2}(s)\cos\{\mu(s) + D\}, \tag{6}$$

$$y'(s) = -C\beta^{-1/2}(s)\left[\alpha(s)\cos\{\mu(s) + D\} + \sin\{\mu(s) + D\}\right], \tag{7}$$

where C and D are constants and y can be either of the transverse coordinates x or z. The phase terms can be isolated and used to define new coordinates $Y(\mu)$ and $Y'(\mu)$ that are known as normalized coordinates.

$$Y(\mu) = C\cos(\mu + D) = y(s)\beta^{-1/2} \tag{8}$$

$$Y'(\mu) = -C\sin(\mu + D) = y(s)\alpha\beta^{-1/2} + y'(s)\beta^{1/2} \ . \tag{9}$$

Note that the independent variable for real-world coordinates is s while for normalized coordinates it is μ. In line with general usage, normalized coordinates have upper case letters and real-world coordinates have lower case letters.

The object now is to express the motion of a test ion in normalized coordinates while crossing a sextupole. To simplify the analysis only a 'thin' sextupole (i.e., a sextupole with zero thickness) will be considered. This approximation is well justified in most cases. Under this assumption, the increments excited by the sextupole will be,

$$\Delta Y = 0, \quad \Delta y = 0 \quad \text{and} \quad \Delta Y' = \beta^{1/2}\,\Delta y'. \tag{10}$$

Assuming the *de facto* standard configuration of a horizontal ring and a normal sextupole, the kicks received in the horizontal and vertical planes are given by Eqs. (3) and (4) and substitution into Eq. (10) gives,

$$\Delta X' = \beta_x^{1/2}\Delta x' = \frac{1}{2}\beta_x^{1/2}\ell_S k'\left(x^2 - z^2\right)$$

$$\Delta Z' = \beta_z^{1/2}\Delta z' = -\beta_z^{1/2}\ell_S k' xz \ .$$

Using Eq. (8), these kicks become,

$$\Delta X' = \frac{1}{2}\beta_x{}^{3/2}\ell_{\mathrm{S}}k'\left(X^2 - Z^2\frac{\beta_z}{\beta_x}\right)$$

$$\Delta Z' = -2\left(\frac{1}{2}\beta_x{}^{3/2}\ell_{\mathrm{S}}k'\right)\frac{\beta_z}{\beta_x}XZ \ .$$

During the extraction process, the amplitude x increases to almost the full half-aperture, while z remains small. On this basis, the motion in z is neglected to first order and the final formulation of the sextupole kicks to be used for the extraction analysis is

Thin sextupole kicks:

$$\Delta X = \Delta Z = 0, \qquad \Delta X' = SX^2 \quad \text{and} \quad \Delta Z' = 0 \ . \tag{11}$$

where

$$S = \frac{1}{2}\beta_x{}^{3/2}\ell_{\mathrm{S}}k' \ . \tag{12}$$

7 Basic theory of the resonance

The horizontal transfer matrix M_{n} for normalized coordinates, describing n turns in the machine is given by:

$$\boldsymbol{M}_{\mathrm{n}} = \begin{pmatrix} \cos 2\pi(nQ_{\mathrm{x}}) & \sin 2\pi(nQ_{\mathrm{x}}) \\ -\sin 2\pi(nQ_{\mathrm{x}}) & \cos 2\pi(nQ_{\mathrm{x}}) \end{pmatrix} . \tag{13}$$

Consider an ion with a horizontal betatron tune close to a third-integer resonance, so that $Q_x = m \pm 1/3 + \delta Q$, where m is integer and $|\delta Q| \ll 1/3$). The tune increment δQ is defined as the tune distance of the ion from the resonance,

$$\delta Q = Q_{\text{particle}} - Q_{\text{resonance}} \ . \tag{14}$$

The explicit transfer matrix for n turns in the unperturbed machine can then be written as

$$\boldsymbol{M}_n = \begin{pmatrix} \cos\left[2n\pi\left(m \pm 1/3 + \delta Q\right)\right] & \sin\left[2n\pi\left(m \pm 1/3 + \delta Q\right)\right] \\ -\sin\left[2n\pi\left(m \pm 1/3 + \delta Q\right)\right] & \cos\left[2n\pi\left(m \pm 1/3 + \delta Q\right)\right] \end{pmatrix} . \tag{15}$$

Thus, to a good approximation the coordinates of the ion after one, two and three turns become:

1st turn neglecting δQ:

$$\begin{pmatrix} X \\ X' \end{pmatrix}_1 \cong \begin{pmatrix} -1/2 & \pm\sqrt{3}/2 \\ \mp\sqrt{3}/2 & -1/2 \end{pmatrix}\begin{pmatrix} X \\ X' \end{pmatrix}_0 \tag{16}$$

2nd turn neglecting δQ:

$$\begin{pmatrix} X \\ X' \end{pmatrix}_2 \cong \begin{pmatrix} -1/2 & \mp\sqrt{3}/2 \\ \pm\sqrt{3}/2 & -1/2 \end{pmatrix}\begin{pmatrix} X \\ X' \end{pmatrix}_0 \tag{17}$$

3rd turn with δQ:

$$\begin{pmatrix} X \\ X' \end{pmatrix}_3 \cong \begin{pmatrix} 1 & \varepsilon \\ -\varepsilon & 1 \end{pmatrix}\begin{pmatrix} X \\ X' \end{pmatrix}_0 \ , \tag{18}$$

where ε replaces $6\pi\delta Q$ for brevity . The small quantity ε is called the 'modified tune distance'. It can be seen that an ion with exactly the resonant tune (i.e., $\varepsilon = 0$) will return to its initial position every three turns.

The effect of the sextupole during three turns in the machine is now calculated as a perturbation by the linear addition of 3 instances using Eqs. (16), (17), and (18):

(A) The effect of 3 turns with a sextupole placed after the 3rd turn,
M3 + Sextupole;

(B) The effect of 3 turns with a sextupole placed after the 2nd turn,
M2 + Sextupole + M1;

(C) The effect of 3 turns with a sextupole placed after the 1st turn
M1 + Sextupole + M2.

In (A), (B), and (C), the sextupole is represented by the simplified, thin-lens kicks from Eqs. (11) and (12). Filling out the steps (A), (B), and (C) explicitly yields:

(A) $$\left.\begin{aligned} X_3 &= \left(X_0 + \varepsilon X_0'\right) \\ X_3' &= -\varepsilon X_0 + X_0' + S\left(X_0 + \varepsilon X'\right)^2 \end{aligned}\right\}.$$

(B) $$\left\{\begin{aligned} X_3 &= -\frac{1}{2}\left(-\frac{1}{2}X_0 \mp \frac{\sqrt{3}}{2}X'\right) \pm \frac{\sqrt{3}}{2}\left[\pm\frac{\sqrt{3}}{2}X_0 - \frac{1}{2}X_0' + S\left(-\frac{1}{2}X_0 \mp \frac{\sqrt{3}}{2}X_0'\right)^2\right] \\ X_3' &= \mp\frac{\sqrt{3}}{2}\left(-\frac{1}{2}X_0 \mp \frac{\sqrt{3}}{2}X'\right) - \frac{1}{2}\left[\pm\frac{\sqrt{3}}{2}X_0 - \frac{1}{2}X_0' + S\left(-\frac{1}{2}X_0 \mp \frac{\sqrt{3}}{2}X_0'\right)^2\right] \end{aligned}\right\}.$$

(C) $$\left\{\begin{aligned} X_3 &= -\frac{1}{2}\left(-\frac{1}{2}X_0 \pm \frac{\sqrt{3}}{2}X'\right) \mp \frac{\sqrt{3}}{2}\left[\mp\frac{\sqrt{3}}{2}X_0 - \frac{1}{2}X_0' + S\left(-\frac{1}{2}X_0 \pm \frac{\sqrt{3}}{2}X_0'\right)^2\right] \\ X_3' &= \pm\frac{\sqrt{3}}{2}\left(-\frac{1}{2}X_0 \pm \frac{\sqrt{3}}{2}X'\right) - \frac{1}{2}\left[\mp\frac{\sqrt{3}}{2}X_0 - \frac{1}{2}X_0' + S\left(-\frac{1}{2}X_0 \pm \frac{\sqrt{3}}{2}X_0'\right)^2\right] \end{aligned}\right\}.$$

After the addition of the three steps (A), (B), and (C) shown above, only first-order correction terms in ε are retained to give:

$$\Delta X_3 = \varepsilon X_0' \pm \frac{\sqrt{3}}{2}S\left(-\frac{1}{2}X_0 \mp \frac{\sqrt{3}}{2}X_0'\right)^2 \mp \frac{\sqrt{3}}{2}S\left(-\frac{1}{2}X_0 \pm \frac{\sqrt{3}}{2}X_0'\right)^2$$

$$\Delta X_3' = -\varepsilon X_0 + SX_0^2 - \frac{1}{2}S\left(-\frac{1}{2}X_0 \mp \frac{\sqrt{3}}{2}X_0'\right)^2 - \frac{1}{2}S\left(-\frac{1}{2}X_0 \pm \frac{\sqrt{3}}{2}X_0'\right)^2.$$

The cancellation of signs shows that there is no fundamental difference between the 1/3rd and 2/3rd resonances.

$$\Delta X_3 = \varepsilon X_0' + \frac{\sqrt{3}}{2} S\left(-\frac{1}{2}X_0 - \frac{\sqrt{3}}{2}X_0'\right)^2 - \frac{\sqrt{3}}{2} S\left(-\frac{1}{2}X_0 + \frac{\sqrt{3}}{2}X_0'\right)^2$$

$$\Delta X_3' = -\varepsilon X_0 + S X_0^2 - \frac{1}{2} S\left(-\frac{1}{2}X_0 - \frac{\sqrt{3}}{2}X_0'\right)^2 + \frac{1}{2} S\left(-\frac{1}{2}X_0 + \frac{\sqrt{3}}{2}X_0'\right)^2 .$$

Finally, the expressions for the changes in position and divergence of the ion, known as the 'spiral step' and 'spiral kick', over three revolutions, are obtained as:

Spiral step:
$$\Delta X_3 = \varepsilon X_0' + \frac{3}{2} S X_0 X_0' . \tag{19}$$

Spiral kick:
$$\Delta X_3' = -\varepsilon X_0 + \frac{3}{4} S\left(X_0^{\,2} - X_0'^{\,2}\right) \tag{20}$$

8 Kobayashi Hamiltonian

The time needed for three revolutions in the machine is short compared to the spill time and can be safely used as the basic unit of time. The elementary changes occurring in this unit of time are the smallest that need to be resolved to understand the physics of the extraction. Thus the subscripts in Eqs. (19) and (20) are no longer needed and the kicks can be treated as a continuous function that is derived from a Hamiltonian $\boldsymbol{H}$, such that

$$\Delta X_3 \Rightarrow \left(\frac{\Delta X}{\Delta t}\right)_{\Delta t = 1\,(3\ \text{turn})} = \frac{\partial \boldsymbol{H}}{\partial X'} = \varepsilon X' + \frac{3}{2} S X X' \ , \tag{21}$$

$$\Delta X'_3 \Rightarrow \left(\frac{\Delta X'}{\Delta t}\right)_{\Delta t = 1\,(3\ \text{turn})} = -\frac{\partial \boldsymbol{H}}{\partial X} = -\varepsilon X + \frac{3}{4} S\left(X^2 - X'^2\right) . \tag{22}$$

The Kobayashi Hamiltonian [3,4] is found by integrating the above partial differentials Eqs. (21) and (22). It should be noted that in this formulation time is dimensionless.

Kobayashi Hamiltonian:
$$\boldsymbol{H} = \frac{\varepsilon}{2}\left(X^2 + X'^2\right) + \frac{S}{4}\left(3XX'^2 - X^3\right) . \tag{23}$$

The Hamiltonian is time independent and a constant of the motion. Contours of constant $\boldsymbol{H}$ show the particle trajectories in normalized phase space at the sextupole. This presentation of the motion is known as a 'phase-space map'. The first term in Eq. (23) describes the unperturbed particle motion in the linear, machine (i.e., $S = 0$). These trajectories are circles of radius $\sqrt{(2\boldsymbol{H}/\varepsilon)}$ in normalized phase space. The second term contains the perturbation that distorts the circular phase-space trajectories into a triangular form as illustrated in Figs. 2 and 6. At a certain level of excitation, the triangle 'breaks' into open phase-space trajectories forming the three separatrices. A change in sign of either the modified tune distance ε or the normalized sextupole strength S is equivalent to a rotation of the phase-space trajectories by 180º.

All the properties of the system can be derived from the Hamiltonian. In particular, when $\boldsymbol{H}$ has the value $[(2\varepsilon/3)^3/S^2]$, it factorizes into three straight lines

$$\left(\frac{S}{4}X + \frac{\varepsilon}{6}\right)\left(\sqrt{3}X' + X - \frac{4\varepsilon}{3S}\right)\left(\sqrt{3}X' - X + \frac{4\varepsilon}{3S}\right) = 0 \ , \tag{24}$$

called the 'separatrices' that define the boundaries between the 'last stable triangle' and unstable regions in phase space. The phase-space area of the stable region is the 'acceptance' of the system at the given momentum. It could also be regarded as the 'dynamic aperture'. The size of the stable region is determined by the ratio $|\varepsilon/S|$. For an ion that has exactly the resonant tune the stable region shrinks to zero.

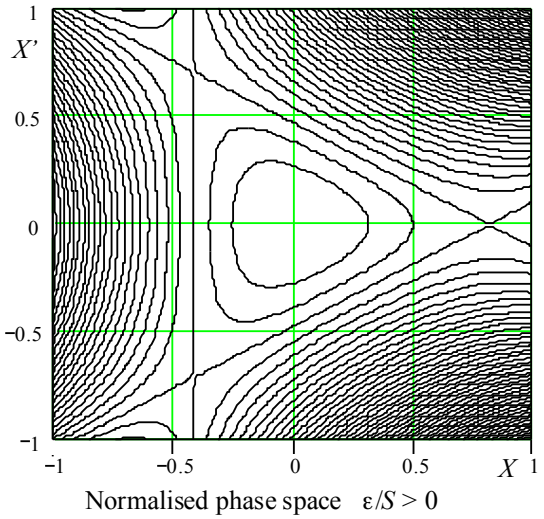
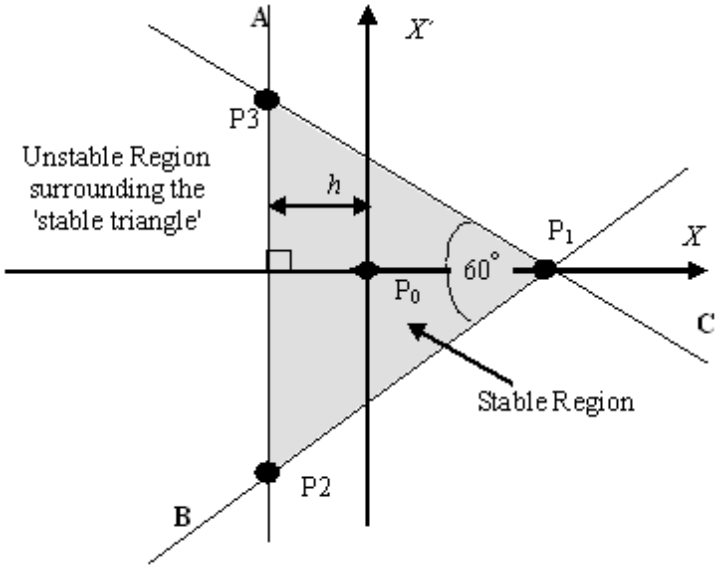


Fig. 6: Phase-space map at the resonance sextupole (plotted from the Kobayashi Hamiltonian)

Figure 7 defines the geometry of the separatrices at the sextupole and the four stable fixed points, P_0 to P_3.



$$P_1 = \left(\frac{4}{3}\frac{\varepsilon}{S}, 0\right)$$

$$P_2 = \left(-\frac{2}{3}\frac{\varepsilon}{S}, -\frac{2\varepsilon}{\sqrt{3}S}\right)$$

$$P_3 = \left(-\frac{2}{3}\frac{\varepsilon}{S}, \frac{2\varepsilon}{\sqrt{3}S}\right)$$

$$h = \frac{2}{3}\frac{\varepsilon}{S} = \frac{4\pi}{S}\delta Q$$

Fig. 7: Geometry of the phase-space map at the resonance-excitation sextupole

The geometry of the last stable triangle is conveniently described by introducing the distance h between the upright separatrix A and the X'-axis. A change in the sign of h is equivalent to a 180° rotation of the stable triangle around the origin.

$$h = \frac{2}{3}\frac{\varepsilon}{S} = \frac{4\pi}{S}\delta Q \,. \tag{25}$$

The area (i.e., the acceptance) of the last stable triangle can also be expressed in terms of h as,

Acceptance of last stable triangle: $\text{Area} = 3\sqrt{3}h^2 = \dfrac{48\sqrt{3}\pi}{S^2}(\delta Q)^2\pi$. (26)

Providing the sextupole is powered in an adiabatic way, the emittance will be conserved and Eq. (26) gives the fraction of the original beam emittance that remains stable.

Since Figs. 6 and 7 are in normalized phase space, the unperturbed trajectories are circles and X is equivalent to X'. Hence, we can rotate the diagram to any position around the machine using the betatron phase to determine the position without altering its shape.

9 Comment on resonances

If this were a lecture on resonances, a more advanced Hamiltonian would be used in order to describe the general case of mixed horizontal and vertical resonances driven by normal and skew fields over much larger apertures. Some examples showing the variety of patterns generated by the relatively simple half-integer resonances are illustrated in Fig. 8.

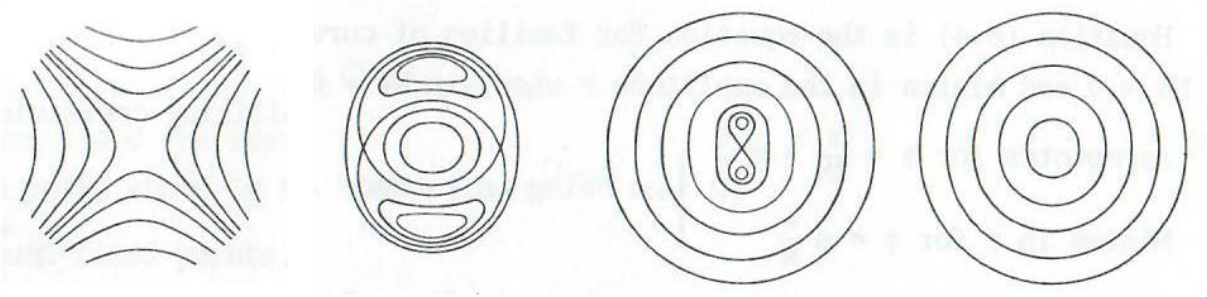


Fig. 8: Examples of the various phase-space views excited by the half-integer resonance

In this lecture, a simplified Hamiltonian has been used. This is justified since only a purely radial 1/3rd integer resonance is being considered and, furthermore, the trajectories are being intercepted within the physical aperture of the machine before more exotic features such as distant stable islands can be formed. It is stressed, that this is a simplified situation that is only used to obtain basic parameters such as the spiral step and to provide some insight into how the slow extraction works. In practice, it is essential to back up these estimations with numerical calculations using a lattice program.

10 The virtual sextupole

In most rings, sextupoles are mounted in dispersion regions for controlling the chromaticity and they are positioned so as to cause as little resonance excitation as possible, but their mutual compensation is never perfect. For slow extraction, a dedicated sextupole for exciting the resonance will be needed and this sextupole is best positioned in a zero-dispersion region, so as not to disturb the chromaticity correction. The dedicated resonance sextupole, which will be dominant, the residual excitation from the chromaticity sextupoles and any sextupole errors from the machine (mainly from the dipoles) can be combined numerically into a single virtual sextupole. This virtual sextupole will behave according to the theory presented so far.

For the 1/3rd integer resonance $3Q_x = p$, the driving term is found by integrating the normal sextupole components around the whole ring according to [5],

$$\kappa(s_0) = \frac{1}{24\sqrt{\pi C}} \int_{s_0}^{s_0+C} \beta_x^{3/2} \left[\frac{-1}{|B_0\rho|} \left(\frac{\mathrm{d}^2 B_z}{\mathrm{d}x^2} \right)_0 \right] \exp(3\mathrm{i}\mu_x)\mathrm{d}s \ , \quad (27)$$

using the earlier definition for S in Eq. (12) and assuming short elements gives,

$$\kappa(s_0) = \frac{-1}{12\sqrt{\pi C}} \sum_{\mathrm{n}} S_{\mathrm{n}} \exp\left(3\mathrm{i}\mu_{\mathrm{x,n}}\right). \tag{28}$$

The virtual sextupole can be found by equating its driving term to the sum of all the driving terms around the ring as given by Eq. (28) and separating the real and imaginary parts to give,

$$S_{\mathrm{virt}} \cos\left(3\mu_{\mathrm{x,virt}}\right) = \sum_{\mathrm{n}} S_{\mathrm{n}} \cos\left(3\mu_{\mathrm{x,n}}\right), \tag{29}$$

$$S_{\mathrm{virt}} \sin\left(3\mu_{\mathrm{x,virt}}\right) = \sum_{\mathrm{n}} S_{\mathrm{n}} \sin\left(3\mu_{\mathrm{x,n}}\right). \tag{30}$$

Using Eqs. (29) and (30), the betatron phase (and hence the position) and the strength of the virtual sextupole can be found as

Virtual sextupole phase:

$$\tan\left(3\mu_{\mathrm{x,virt}}\right) = \frac{\sum_{\mathrm{n}} S_{\mathrm{n}} \sin\left(3\mu_{\mathrm{x,n}}\right)}{\sum_{n} S_{\mathrm{n}} \cos\left(3\mu_{\mathrm{x,n}}\right)}. \tag{31}$$

Virtual sextupole strength:

$$S_{\mathrm{virt}}^2 = \left[\sum_{\mathrm{n}} S_{\mathrm{n}} \cos\left(3\mu_{\mathrm{x,n}}\right)\right]^2 + \left[\sum_{\mathrm{n}} S_{\mathrm{n}} \sin\left(3\mu_{\mathrm{x,n}}\right)\right]^2. \tag{32}$$

11 Configuring the extraction

The basic theory in the earlier sections can be exploited in a number of practical configurations. These are listed briefly below with some comments on their use:

11.1 Moving the beam

This method has the advantage of leaving the transverse optical parameters of the machine constant, and hence also those of the resonance. There are two possibilities:

11.1.1 Momentum-driven

In this scenario the beam is accelerated towards the stationary resonance by a betatron core, or by stochastic noise, or possibly by phase displacement acceleration or a RF micro-bucket acceleration system. The principal requirement is to accelerate the beam smoothly and in a controlled way over say one to two seconds. The betatron core having a large inductance smooths out ripple and the steps made by the digital to analogue converter (DAC), but for the same underlying reason it is not ultra-fast to respond. Stochastic noise can be particularly slow acting and this method has been used for ultra-slow extractions lasting 30 minutes or more.

11.1.2 Amplitude-driven

In this case, the beam is excited by transverse stochastic noise or RF excitation at the revolution frequency, so that its betatron amplitudes grow. The chromaticity is set to near zero, so that the resonance line acts as a threshold in amplitude above which the ions become unstable. RF excitation of the

amplitude works well in the range of one to a few seconds and it is quick to respond when switched off [6].

11.2 Moving the resonance

This method is not recommended as it alters the transverse optics of the machine during the extraction. Because the optics is changing, on-line compensation systems are needed to keep the extracted beam at a constant position, angle, momentum spread, and intensity.

11.2.1 Quadrupole-driven

This is probably the first method ever used. The tune of the machine is changed by a quadrupole so as to move the resonance across the beam. In addition to the objections in section 11.2, this method is particularly sensitive to ripple in the power converter driving the extraction quadrupole.

11.2.2 Sextupole-driven

The resonance excitation is changed by increasing the strength of the resonance sextupole. This method is included only for academic completeness.

11.3 Equipment configuration

The momentum-driven and the amplitude-driven are the two most important configurations, but for brevity, this paper will focus on the momentum-driven extraction with all the extraction equipment in the positive, outer half of the aperture. There are of course other possibilities, but in general the principles remain unchanged. Placing the ES in the 1st quadrant requires the virtual sextupole to be positioned so that the chosen extraction separatrix advances around the machine to the ES where it has the angle of 45º±15º with respect to the *x*-axis. At this point, the ES intercepts the beam segment corresponding to the last three turns. This segment will be typically 10 mm in radial length. The ES wall, which is made of thin wires usually of 0.1 mm diameter will intercept an equivalent amount of beam (~1%). The beam segment inside the ES is given a radial kick typically of a few milliradians. Once this is done, the beams continue to the magnetic septum (MS). So as to maximize the gap between the end of the extraction separatrix and the segment of extracted beam, there is ideally a horizontal betatron phase advance of about 90º between the ES and MS to convert the kick to a physical gap at the MS of typically 20 mm. Figure 9 summarizes these points.

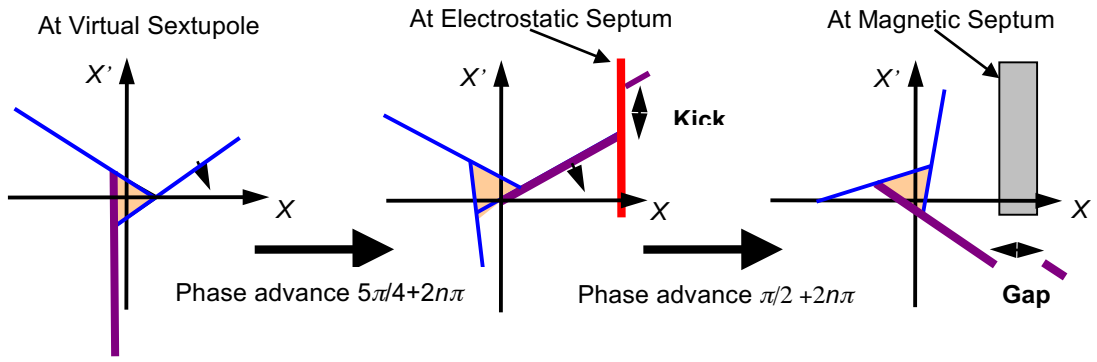


Fig. 9: Configuring the ES and MS for a slow extraction scheme

Of course, the ES and MS must also coincide with available drift spaces in the lattice and have suitable Twiss parameters. There are also further conditions to be considered such as the Hardt condition (see sections 12−14). The final configuration will inevitably be a series of compromises.

12 A more generalized Hamiltonian

The previous section referred to the Hardt condition, which is an important feature in slow extraction schemes. To derive this condition, it is necessary to extend the simple Kobayashi Hamiltonian given earlier to include the dispersion function [7]. To first order in momentum, the equilibrium orbit in normalized coordinates is given by:

$$X_{\mathrm{EQ.O}} = D_n \frac{\delta p}{p}, \quad X'_{\mathrm{EQ.O}} = D'_n \frac{\delta p}{p} , \tag{33}$$

where $(D_\mathrm{n}, D'_\mathrm{n})$ is the normalized dispersion vector and $\delta p/p$ is the fractional momentum deviation. A second coordinate system (X_β, X'_β) is introduced with its origin on the off-momentum equilibrium orbit. The particle coordinates are then split into a constant term, given by the dispersion function and a betatron term as shown in Fig. 10.

$$X = X_\beta + D_\mathrm{n} \frac{\delta p}{p} \quad \Leftrightarrow \quad X_\beta = X - D_\mathrm{n} \frac{\delta p}{p} \tag{34}$$

$$X' = X'_\beta + D'_\mathrm{n} \frac{\delta p}{p} \quad \Leftrightarrow \quad X'_\beta = X' - D'_\mathrm{n} \frac{\delta p}{p} . \tag{35}$$

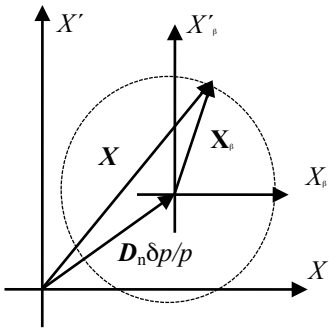


Fig. 10: Coordinates of the betatron motion

The general Hamiltonian is derived following the same steps as for the Kobayashi Hamiltonian except that the betatron motion and the total motion now have to be separated. The analysis starts with the betatron motion Eqs. (34) and (35) for n turns in the unperturbed machine, as before,

$$\begin{pmatrix} X_\beta \\ X'_\beta \end{pmatrix}_\mathrm{k} = \boldsymbol{M}_\mathrm{n} \begin{pmatrix} X_\beta \\ X'_\beta \end{pmatrix}_0 . \tag{36}$$

The main difference appears when calculating the kick of the sextupole that now depends on the total motion of the ion,

Dispersion region: $$\Delta X_\beta = 0, \quad \Delta X'_\beta = SX^2 = S\left(X_\beta + D_n \frac{\delta p}{p} \right)^2 . \tag{37}$$

This yields a more general Hamiltonian in the betatron coordinates (X_β, X'_β):

$$\boldsymbol{H} = \frac{\varepsilon}{2}\left(X_\beta^2 + X_\beta'^2\right) + \frac{S}{4}\left(3X_\beta X_\beta'^2 - X_\beta^3\right) - \frac{3}{2}S\left(D_\mathrm{n} \frac{\delta p}{p} \right)\left(X_\beta^2 + X_\beta'^2\right) . \tag{38}$$

Re-ordering the terms in Eq. (38) shows that the dispersion-dependent term affects only the 'circular' trajectories and this can be considered as a change in the tune distance, ε.

General form: $$\boldsymbol{H} = \frac{1}{2}\left[\varepsilon - 3S\left(D_{\mathrm{n}}\frac{\delta p}{p}\right)\right]\left(X_\beta^2 + X_\beta'^2\right) + \frac{S}{4}\left(3X_\beta X_\beta'^2 - X_\beta^3\right). \tag{39}$$

Let $\delta\widetilde{Q}$ be the change in tune implied by the dispersion-dependent term in Eq. (39), so that

. $$\delta\widetilde{Q}_x = -\frac{1}{6\pi}3S\left(D_{\mathrm{n}}\frac{\delta p}{p}\right) = -\frac{1}{4\pi}\beta_{\mathrm{x}}\ell_{\mathrm{S}}k'D_{\mathrm{x}}\frac{\delta p}{p}\,. \tag{40}$$

By introducing the chromaticity $Q'_{\mathrm{x}} = \left(\frac{dQ_{\mathrm{x}}}{dp/p}\right)$ as the linear change of the betatron tune with momentum, the chromatic effect of the sextupole can be expressed as

$$Q'_{\mathrm{x}} = \left(\frac{d\widetilde{Q}_{\mathrm{x}}}{dp/p}\right) = -\frac{1}{4\pi}\beta\ell_{\mathrm{S}}k'D_{\mathrm{x}}\,, \tag{41}$$

which corresponds exactly to the classic linear form for the tune shift introduced by a sextupole.

The general Hamiltonian in Eq. (39) correctly describes the physics of a third-integer resonance to first order in a perfect lattice without any further restrictions on sextupole locations or particle momenta. The phase space at the sextupole has the same qualitative shape as the earlier simple theory, but it is scaled by the tune shift introduced by the sextupole. In readiness for the next topic, the geometry and general equations of the separatrices are given in Fig. 11. From these results, it is clear that any change in the sextupole strength will also change the tune distance of the particles and therefore the scale of the extraction phase space. This a second reason for putting the resonance sextupole in a dispersion-free region (the first reason was not to affect the chromaticity).

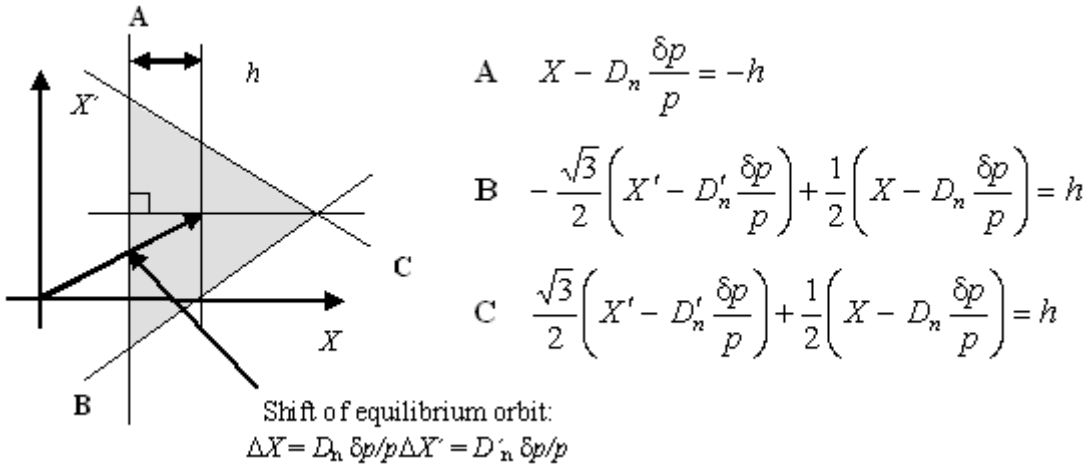


Fig. 11: Geometry of the separatrices at the sextupole for an off-momentum beam

The above can be further extended in an exactly analogous way for the effect of closed-orbit errors. This can be used to put limits on the closed-orbit variation that can be tolerated at the resonant sextupole.

13 Equations for separatrices

A general equation for a separatrix in the phase-space map can be constructed from the standard form for a straight line that uses the perpendicular distance (in this case *h*) from the origin Eq. (42) and Fig. 12.

Standard form: . (42)

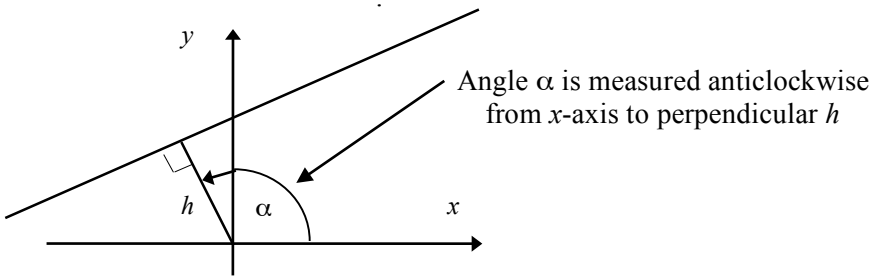


Fig. 12: Perpendicular form for a straight line

Combining Eq. (42) with the shift of origin corresponding to the dispersion vector yields the general equation for a separatrix.

$$\left(X - D_n(s)\frac{\delta p}{p}\right)\cos\alpha + \left(X' - D_n'(s)\frac{\delta p}{p}\right)\sin\alpha = h \ . \tag{43}$$

The separatrices rotate clockwise with the betatron phase advance $\Delta\mu$ measured from the sextupole. At any given position $\alpha = (\alpha_0 - \Delta\mu)$ and by inspection of Figs. 11 and 12 the values for α_0 at the sextupole are:

(A) $\alpha_0 = 180°$ (anticlockwise);

(B) $\alpha_0 = 300°$ (anticlockwise);

(C) $\alpha_0 = 420°$ (anticlockwise);

so that the three separatrices can be drawn at any position using,

$$\left(X - D_n(s)\frac{\delta p}{p}\right)\cos(\alpha_0 - \Delta\mu) + \left(X' - D_n'(s)\frac{\delta p}{p}\right)\sin(\alpha_0 - \Delta\mu) = h \ . \tag{44}$$

14 Hardt condition

Referring back to Fig. 3, it can be seen that in a momentum-driven extraction all ions on the line AB become unstable at the same time. Each point between A and B has its own momentum, its own stability threshold, its own transit time in the resonance and its own last stable triangle. This spread in parameters can be useful for smoothing ripple, but unless the extraction separatrices are all aligned there will be significant particle losses on the ES. Figure 13 illustrates the general case in real space. The last stable triangles are spread out and when the extraction separatrices reach the ES they all have different angles. Since the ES can only be aligned for one angle, losses will occur.

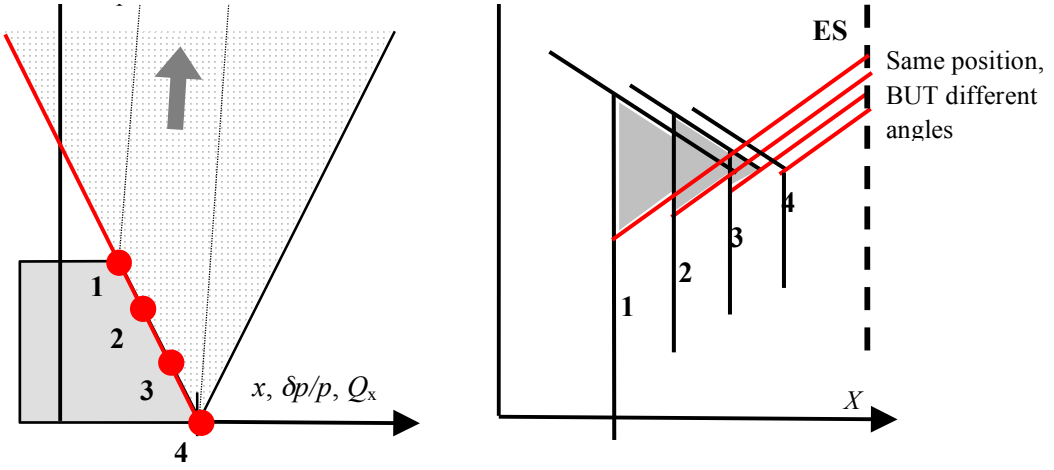


Fig. 13: A momentum-driven extraction without alignment of the separatrices

Minimum losses would be obtained if all the separatrices could be super-positioned. This is achieved by the Hardt condition [8]. Since the stable triangles have different momenta they are affected by the dispersion at the ES and by a suitable choice of the dispersion vector all the separatrices of the different momenta can be made to enter the ES with the same angle. The restrictions on the lattice functions to satisfy the Hardt condition can be derived with a purely mathematical approach from the general expression for the separatrix already given in Eq. (44). First Eq. (25) is rewritten using the chromaticity as,

$$h = \frac{2}{3}\frac{\varepsilon}{S} = \frac{4\pi}{S}\delta Q = \frac{4\pi}{S}Q'\frac{\delta p}{p} \,. \tag{45}$$

The substitution of Eq. (45) into Eq. (44) gives the general expression for a separatrix as a function of particle momentum and the chromaticity of the machine,

$$\left(X - D_n(s)\frac{\delta p}{p}\right)\cos(\alpha - \Delta\mu) + \left(X' - D_n'(s)\frac{\delta p}{p}\right)\sin(\alpha - \Delta\mu) = \frac{4\pi}{S}Q'\frac{\delta p}{p} \,. \tag{46}$$

To superimpose the extraction separatrices, the momentum dependence has to be removed from Eq. (46). This is true if,

Hardt condition:
$$D_n\cos(\alpha - \Delta\mu) + D_n'\sin(\alpha - \Delta\mu) = -\frac{4\pi}{S}Q' \,. \tag{47}$$

In practice, there is usually enough flexibility in D_n, D_n', and Q' to satisfy Eq. (47) with non-zero values. This corresponds to the Hardt condition being applied to the case of the momentum-driven extraction.

There is a second possibility of making $D_n = D_n' = Q' = 0$. Clearly zero chromaticity does not work for the momentum-driven extraction, but the amplitude-driven extraction can work with zero, or rather near-zero, chromaticity. The amplitude-driven extraction also uses a narrower beam in momentum, which makes the problem of aligning the separatrices less critical.

15 Numerical design

Although the above theory explains how slow extraction works and indicates the main configurations for the beams and equipment, this is not enough for designing a practical scheme. Most lattice programs will quickly reveal that separatrices are curved and the various parameters are not predicted with sufficient accuracy. It is therefore necessary to resort to a lattice program and to perform some trial and error calculations to set the spiral step precisely, to fine tune the Hardt condition and so on. Figure 14

shows the Hardt condition calculated for a momentum-driven extraction. Only the separatrices for the largest and the zero-amplitude ions are shown. It can be seen that these separatrices are curved and are not superimposed over their full length, but they have been made to cross at the ES using the Hardt condition theory and some trial-and-error refinements.

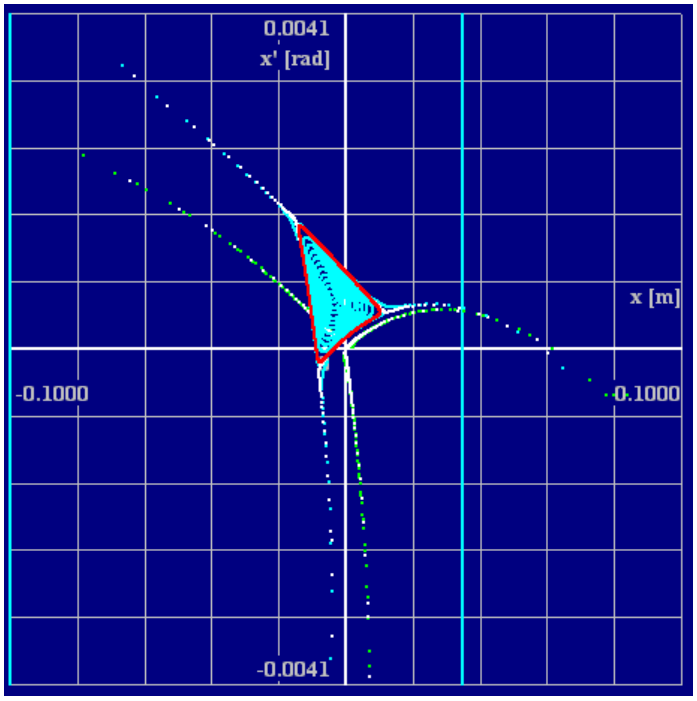


Fig. 14: The Hardt condition for a momentum-driven extraction (screen shot, WinAgile)

16 Getting to the MS

Another practical issue that depends on numerical calculation is the passage of the extracted beam segment between the ES and the MS. Depending on the lattice design, the extracted segment will stray beyond the edge of the machine aperture. The aperture between the ES and the MS is checked visually using a graphical method developed by C. Steinbach (CERN), see Fig. 15.

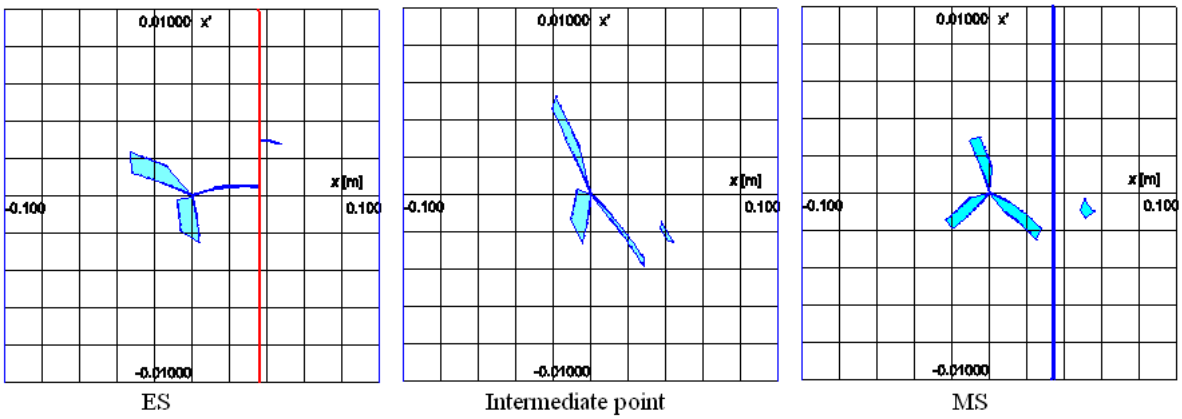


Fig. 15: Visual check of horizontal aperture from ES to MS (screen shots from WinAgile)

In Fig. 15, the phase-space areas between the maximum and the zero-amplitude separatrices are marked in light blue as well as the extracted beam segment. The efficiency of the method relies on the computer program being given all the aperture data for all the lattice positions between the ES and MS.

17 A 'strip' spill

It is now time to look more closely at how the ions cross the resonance and how this affects the spill. For mathematical convenience, the origin of the Hamiltonian is shifted to the upper fixed point and the extraction is considered via this fixed point, see Eq. (48) and Fig. 16.

$$\boldsymbol{H}_{\text{Shifted}} = \frac{S}{4}\left(4h^3 + 6hX^2 + 6\sqrt{3}hX\,X' + 3X\,X'^2 - X^3\right). \tag{48}$$

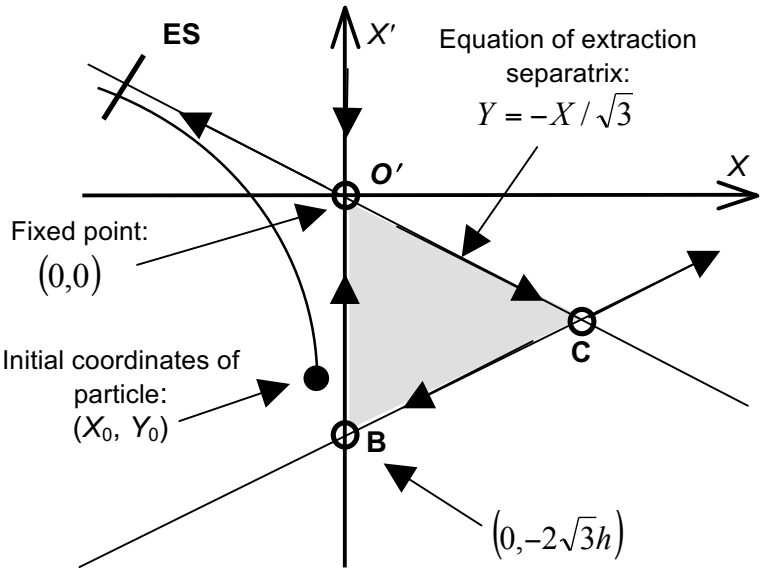


Fig. 16: Kobayashi Hamiltonian with shifted origin

The extraction process can be studied by considering what happens when the last stable triangle shrinks by a small step and a narrow strip of ions (coloured red) that were previously stable find themselves outside, in the unstable region, see Fig. 17. Only one of the three sides of the triangle is considered here. The black arrows show the general direction of movement of the ions. The grey arrows show the movement of the separatrices.

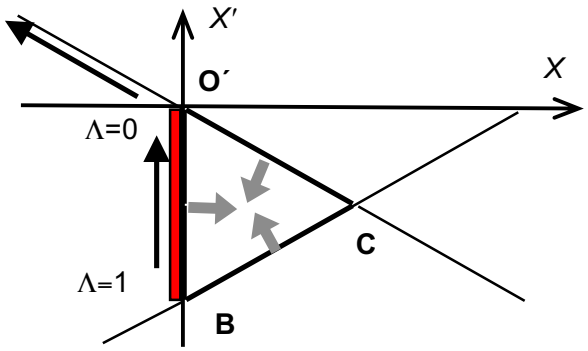


Fig. 17: Effect of shrinking the last stable triangle

In the translated frame of reference, the equations of motion become:

$$\frac{dX}{dt} = \frac{\partial \boldsymbol{H}}{\partial X'} = \frac{S}{4}\left(6\sqrt{3}hX + 6XX'\right). \tag{49}$$

$$\frac{dX'}{dt} = -\frac{\partial \boldsymbol{H}}{\partial X} = -\frac{S}{4}\left(12hX + 6\sqrt{3}hX' + 3X'^2 - 3X^2\right). \tag{50}$$

Extracted ions will follow paths of constant $\boldsymbol{H}$ close to the separatrices. The trajectories come directly from the initial conditions (X_0, X'_0) substituted into the Hamiltonian Eq. (48).

$$\frac{S}{4}(4h^3 + 6hX_0^2 + 6\sqrt{3}hX_0\,X_0' + 3X_0\,X_0'^2 - X_0^3) = \frac{S}{4}(4h^3 + 6hX^2 + 6\sqrt{3}hX\,X' + 3X\,X'^2 - X^3).$$

Close to the fixed point O′, ($|X|$, $|X'| << h$) and the third-order terms in X and X' can be neglected, so that

$$X' = \frac{X_0^2 + \sqrt{3}X_0X_0' - X^2}{\sqrt{3}X}. \tag{51}$$

Substituting Eq. (51) into the earlier expression for dX/dt in Eq. (49) gives the time profile. Note that time is dimensionless and is measured in steps of 3 turns.

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \frac{S}{4}\left(6\sqrt{3}hX + \frac{6}{\sqrt{3}}X_0^2 + 6X_0X_0' - \frac{6}{\sqrt{3}}X^2\right). \tag{52}$$

Fortunately, Eq. (52) is a standard form that can be integrated.

$$\int \frac{\mathrm{d}x}{ax^2 + bx + c} = \frac{1}{\sqrt{b^2 - 4ac}} \ln\left|\frac{2ax + b - \sqrt{b^2 - 4ac}}{2ax + b + \sqrt{b^2 - 4ac}}\right|. \tag{53}$$

Within the strict assumptions made earlier ($|X|$, $|X'| << h$) , the result is only valid close to O′, but, since the particle approaches the separatrix asymptotically the third-order terms in the Hamiltonian cancel out, so they can also be neglected far from O′ along the outgoing separatrix and the integration can be extended right up to the ES.

The above analysis is used to find how long it takes for a particle to leave the machine. This is called the resonance ‘transit time’. The transit time is split into two parts:

- the time to travel from O′ to the ES; and
- the time to move along the side of the stable triangle from an arbitrary point towards O′.

Firstly, this is done under static conditions (while the stable triangle remains constant in size and position). In a second step these expressions must be modified to take into account the dynamic conditions of a shrinking stable triangle. The static result is shown in Eq. (54) and further results can be found in Refs. [9,10].

$$T_{\text{static}} \approx \frac{1}{\varepsilon\sqrt{3}} \ln\left| \underbrace{\frac{1}{(1-\Lambda_0)^2}}_{\substack{\text{Position on side of} \\ \text{triangle}}} \quad \underbrace{\left(\frac{n}{n+3}\right)}_{\substack{\text{Position of} \\ \text{electrostatic septum}}} \quad \underbrace{\frac{3}{\lambda_0}}_{\substack{\text{Distance} \\ \text{to triangle}}} \right|. \tag{54}$$

Firstly, Eq. (54) gives an estimate of the time needed for the first ions to arrive at the ES from the corner of the stable triangle.

Delay for first ions to arrive:	$t_0 \approx \dfrac{1}{\varepsilon\sqrt{3}}$.	(55)

Secondly, it yields the same estimate for the length of time needed for all the ions along one side of the triangle to reach the corner. When these results are combined with estimates of the particle density along the sides of the triangle (the ions tend to concentrate at the corners and move speedily along the sides), the time profile of a strip spill is obtained, see Fig. 18. This profile shows a narrow spike of current after the delay given by Eq. (55) followed by a low flat spill of the same length as the initial delay. Typically, the initial delay is a few hundred time units (1 time unit being 3 turns).

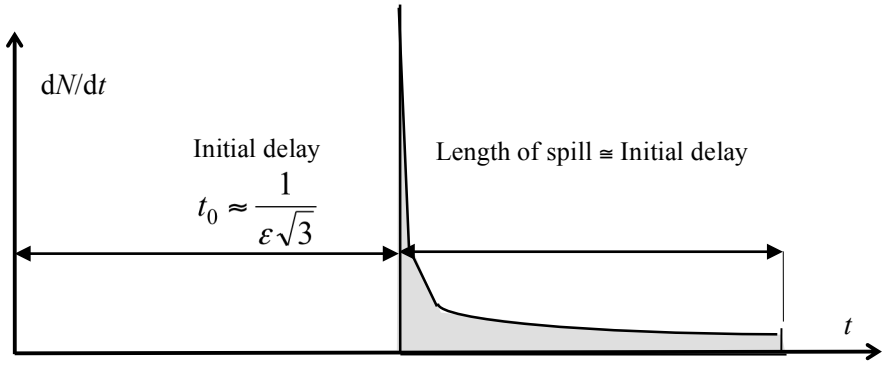


Fig. 18: Time profile of a strip spill

In the momentum-driven extraction (see Fig. 3), there is a range of momenta with stable triangles extending from zero to a maximum value (between points A and B). The smaller the area of the stable triangle, the smaller ε and the longer the transit time becomes. Combining all the strip spills leads to a so-called 'band' spill. Figure 19 shows an example.

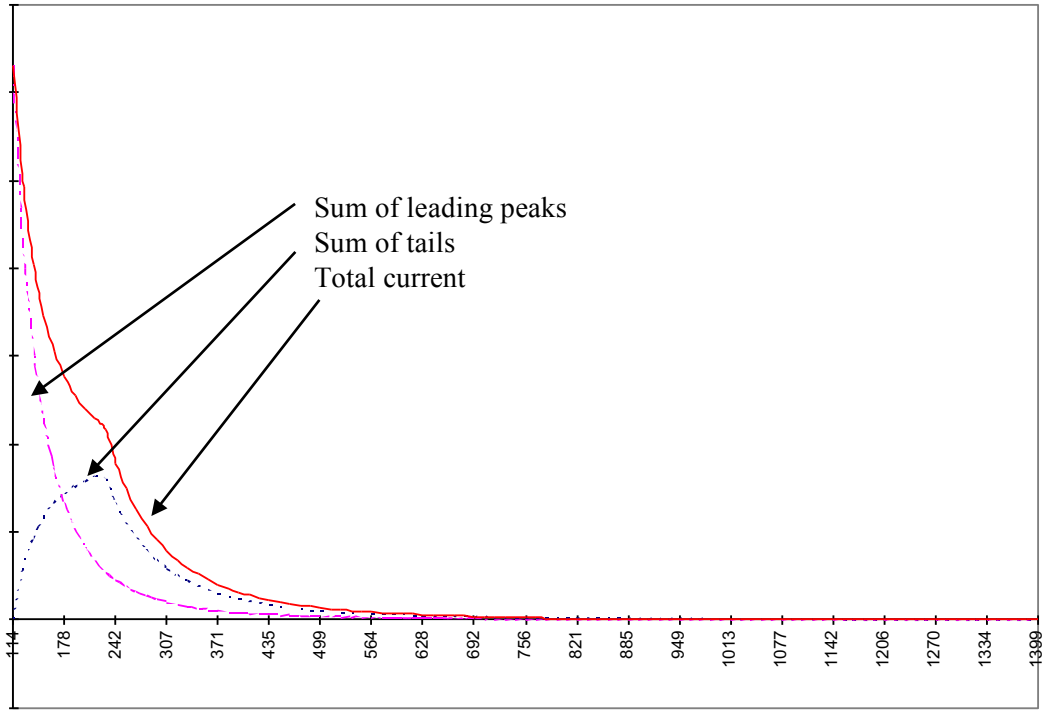


Fig. 19: Example of the time profile of a 'band' spill

18 Emittance of the extracted beam

The conservation of phase space (Liouville's Theorem) affords a simple and quick method for estimating and understanding the emittance of the extracted beam. Consider a coasting beam with a relatively large momentum spread. Assume that resonance is configured to give a significantly narrower momentum spread in the extracted beam. During the extraction, the separatrix acts like a knife shaving off the beam and its phase space. The transverse phase space and the longitudinal phase space ($\Delta p/p$, t) are jointly conserved as a phase-space volume. In the transverse plane, one can neglect the coupling to the vertical plane (in line with earlier approximations), so that the 'waiting' beam and the extracted beam have the same vertical emittance. The effect of the extraction can be seen by equating the phase-space volumes before and after the extraction and then looking at how the horizontal emittance in particular was affected. This is expressed quantitatively in Eq. (56) and illustrated schematically in Fig. 20.

$$E_{\mathrm{x,stack}}\left(\frac{\Delta p}{p}\right)_{\mathrm{stack}} T_{\mathrm{rev}} = E_{\mathrm{x,spill}}\left(\frac{\Delta p}{p}\right)_{\mathrm{spill}} T_{\mathrm{spill}} \ . \tag{56}$$

To give a quantitative impression, some orders of magnitude would be useful. Let the 'waiting' beam's horizontal emittance be $E_{\mathrm{x,beam}} = 10\pi \times 10^{-6}$ [m rad], $(\Delta p/p)_{\mathrm{beam}} = 0.005$, $T_{\mathrm{rev}} = 0.5 \times 10^{-6}$ [s], $(\delta p/p)_{\mathrm{spill}} = 0.001$, $T_{\mathrm{spill}} = 500 \times 10^{-6}$ [s], so that $E_{\mathrm{x,spill}} = 50\pi \times 10^{-12}$ [m rad]. Thus, the long spill time compared to the short revolution time is balanced by the small transverse emittance of the spill compared to the much larger emittance of the 'waiting' beam. Under perfect conditions, this emittance would be seen as zero. However, the small horizontal emittance is often blown-up by ripple, noise, and imperfections in the optics. How big the horizontal emittance ends up is basically a question of power converter quality, optical precision, and various effects such as hysteresis and magnet time constants.

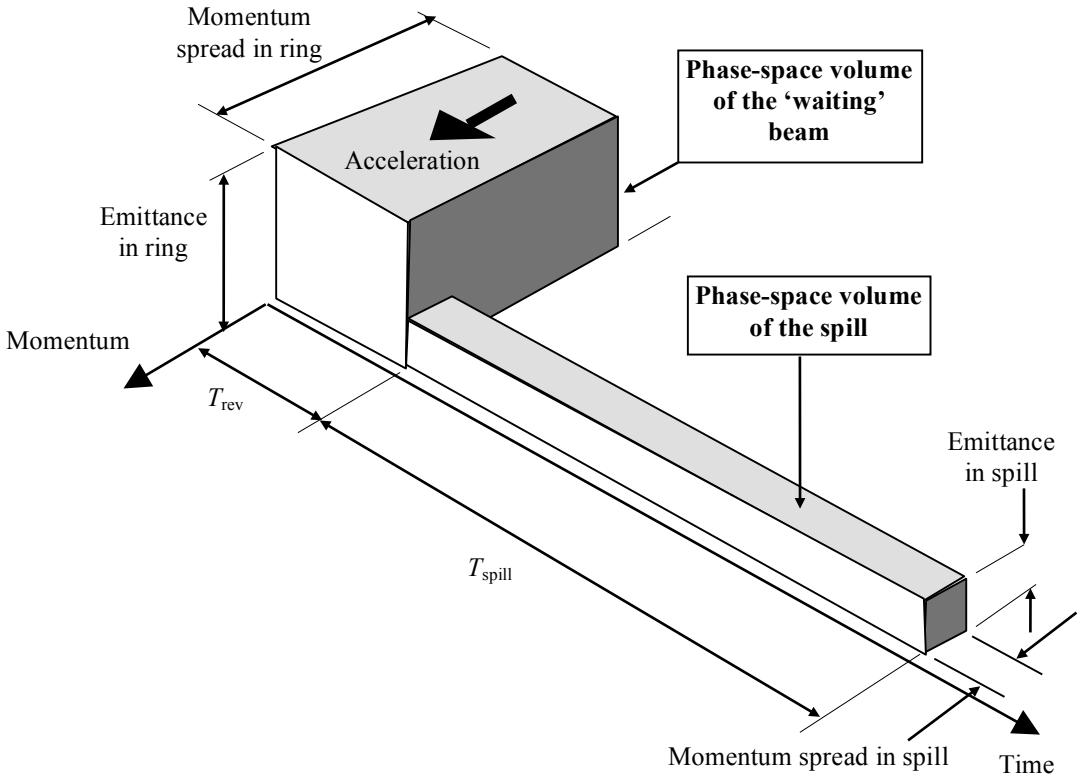


Fig. 20: Schematic view of the phase-space volume during extraction

19 A descriptive view of spill quality

19.1 Momentum-driven extraction

Momentum-driven, slow extraction can be likened to the shaving action on a wood-turning lathe, see Fig. 21. The projected width of the shaving is equivalent to the extracted $\delta p/p$, which is smaller than the width of the wood (i.e., the original beam $\Delta p/p$). The thickness of the shaving is related to the horizontal emittance of the extracted beam and the X-section of the wood is the original beam emittance. A typical extraction over one second in a machine with a revolution time of about one micro-second would require of the order of one million turns. This means that the shaving is extremely thin and the slightest vibration can cause the chisel to jump and chop the shaving, which gives rise to the legendary reputation of slow extractions for their sensitivity to ripple.

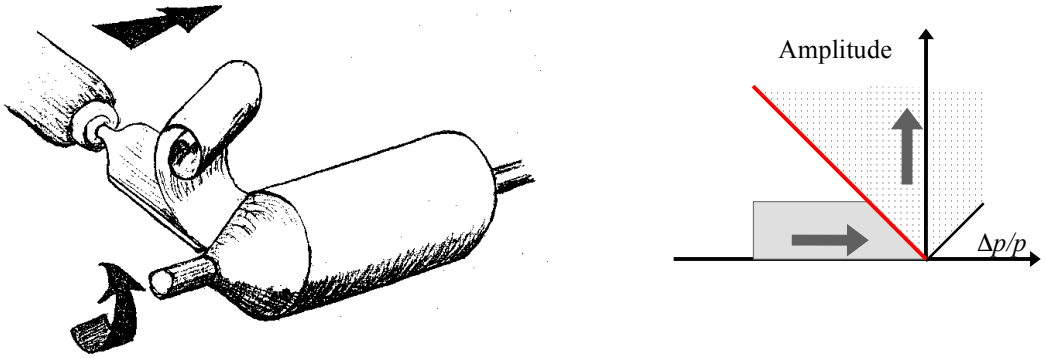


Fig. 21: Model of momentum-driven slow extraction

19.2 Amplitude-driven extraction

Amplitude-driven, slow extraction can also be likened to the shaving action on a wood-turning lathe, see Fig. 22. The width of the shaving is equal to the width of the wood and is equivalent to the $\Delta p/p$. The thickness of the shaving is the horizontal emittance of the extracted beam and the X-section of the wood is the original beam emittance. For ~1,000,000 turns, the shaving is extremely thin. As before, a typical extraction over one second in a machine with a revolution time of about one micro-second would require of the order of one million turns. This means that the shaving is extremely thin and the slightest vibration can cause the chisel to jump and chop the shaving.

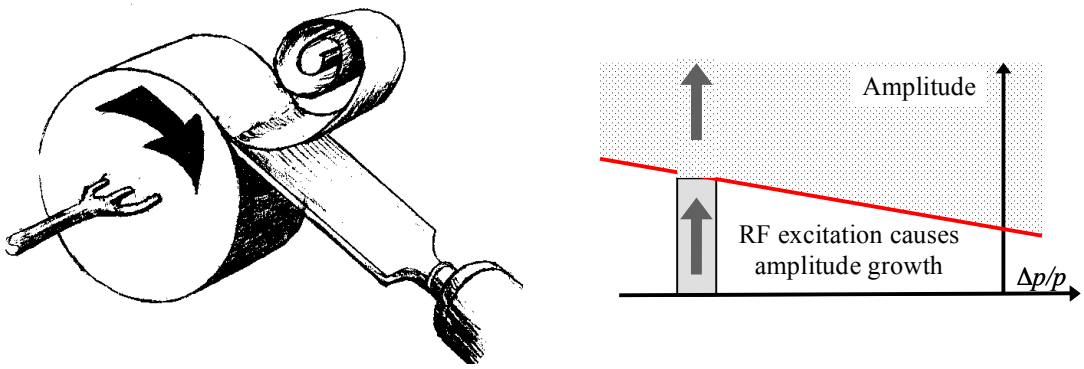


Fig. 22: Model of amplitude-driven slow extraction

20 Sensitivity to ripple

A simple analysis of the effects of ripple at low frequencies can be made by assuming that the particle flux that enters the resonance from the 'waiting' beam appears instantaneously in the observed spill. This is a fair approximation when the transit times in the resonance are small compared to the ripple frequency. Transit times depend upon the circumference of the machine, the strength of the resonance, the emittance, and so on. For a small medical synchrotron, the transit times typically vary from 100 turns up to 4,000 turns. There will always be a few particles that are quicker and few that are slower. For ease of computation, assume that most particles leave the machine within 2,000 turns. For a revolution time of 0.5 μs, this represents a generous delay of 1 ms. Thus, for 50 Hz, 100 Hz, and 300 Hz ripples (common in power converters), it is reasonable to say that extraction is instantaneous. With this simplified picture in mind, consider the schematic model shown below in Fig. 23. Whether the relative motion between the beam and the resonance is due to one partner, or both, is unimportant for the moment.

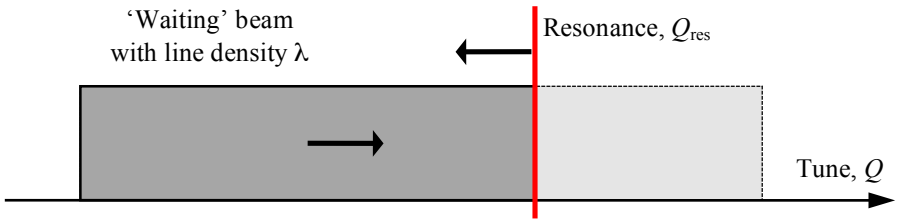


Fig. 23: Simple model for ripple

The relative motion between the beam and resonance will comprise a constant velocity, $(\mathrm{d}Q/\mathrm{d}t) = \dot{Q}_0$ plus a ripple term that may come from either the position of the beam via the main dipole field ripple or from the position of the resonance via the quadrupole ripple. Let the ripple have a sinusoidal form,

$$Q_r = \delta Q \sin(\omega t) \quad \text{so that} \quad \dot{Q}_r = \omega \delta Q \cos(\omega t) \,. \tag{57}$$

The flux of ions entering the resonance will be,

$$\phi = \frac{\mathrm{d}N}{\mathrm{d}t} = \lambda(\dot{Q}_0 + \dot{Q}_r) = \lambda(\dot{Q}_0 + \omega \delta Q \cos(\omega t)) \,, \tag{58}$$

where λ is the line density in the 'waiting' beam. By inspection, it is clear that while $\dot{Q}_0 > \omega\delta Q$ the spill will be continuous but modulated, whereas once $\dot{Q}_0 < \omega \delta Q$ the spill will be periodically 'chopped'. Thus, the critical ripple for a 'chopped' spill is given by

$$(\omega \delta Q)_{\text{critical}} = \dot{Q}_0 \,. \tag{59}$$

To give a quantitative impression, consider a typical small machine with T_{spill} = 1 s, Q_{res} = 1.666 and a tune spread necessary to consume the beam of ΔQ = 0.01. The average tune speed would be $\dot{Q}_0 = 0.01 s^{-1}$. The relationship of the tune ripple to the current ripple would be close to $\Delta Q / Q \approx \Delta I / I$. Some results are listed in Table 1. It is immediately apparent that just to stop the spill from being chopped requires exceeding tight specifications for the power supplies. Table 1 implies that extremely strict specifications for power supply ripple, transients and DAC steps are needed to prevent the chopping of the spill. For a project, it would be necessary to consider many more aspects before writing the power converter specifications. Some of these aspects will be considered in the next section.

Table 1: Conditions for 100% modulation of the spill at low frequencies ($dQ_0/dt = 0.01$, $Q_{res} = 1.666$)

Frequency [Hz]	$(\delta Q/Q_{res})_{critical} = \Delta I/I$
50	$2 \cdot 10^{-5}$
100	10^{-5}
300	$3.2 \cdot 10^{-6}$
1000	$9.6 \cdot 10^{-7}$

At higher frequencies, the spikes on the band spill (see Fig. 19) still appear as individual spikes, but the long tails overlap creating a uniform base spill containing about half of the ions topped by the narrow spikes. At much higher frequencies, the spikes start to overlap.

21 Handling ripple

21.1 Power converters

Switch-mode power converters have an advantage over the more conventional units, especially if switch-mode frequency can be placed above 20 kHz. Care must also be taken to ensure there is always a sufficient number of DAC steps in the current range to be ramped.

21.2 Eddy current smoothing

Normally, eddying currents are avoided in accelerators; to reduce field distortion and resistive losses during ramping. However, during a slow extraction, very stable conditions in the 1,000 Hz region are required for the main lattice elements and for the betatron core. Adjusting the lamination thickness of the main lattice magnets can be of help. This is of course a compromise—some resistive losses and distortion during ramping for some field smoothing on the extraction flat top.

21.3 Intrinsic smoothing

Referring back to the ‘strip’ spill in Fig. 18 and the ‘band’ spill in Fig. 19, the former is a close approximation of what happens in a RF, amplitude-driven extraction while the latter corresponds to what occurs in the momentum-driven extraction with a betatron core. Clearly, there is some advantage in the momentum-driven extraction with the band spill, because the narrow peaks are spread out. This is known as ‘intrinsic smoothing’.

21.4 Front-end acceleration

Figure 24 gives a pictorial view of a momentum-driven extraction. In the analogy, the ‘beam ions’ are moved slowly into the resonance by a plunger that pushes the ions across a vibrating platform. In the case shown, the vibrations dominate the slow acceleration and they wobble on the edge of the platform shaken by the vibrations (ripple) until they fall off the platform (and into the resonance) in an uncontrolled way.

The spill can be desensitized to the ripple by increasing the velocity of the ‘beam ions’ as they reach the edge of the platform, so that their velocity is greater than the perturbation caused by the ripple. Figure 25, shows the same analogy as Fig. 24, except that the ‘beam ions’ now sprint forward just before the edge of the platform, so that the vibrations becomes less influential. This is called ‘front-end’ acceleration and can be done by ‘empty-bucket’ channelling (see next section) or by RF noise.

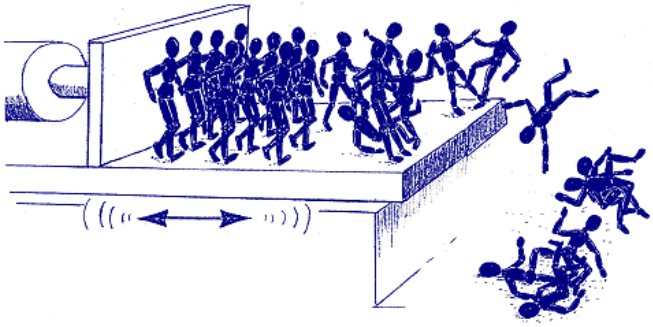


Fig. 24: No addition 'front-end' acceleration

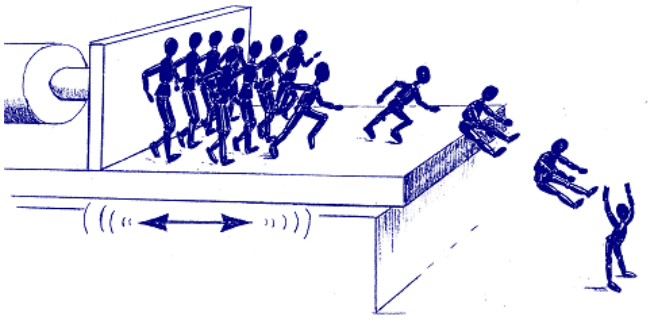


Fig. 25: 'Front-end' acceleration stabilizes ions as they pass into resonance

21.5 Empty-bucket channelling

The 'front-end' acceleration mentioned in the previous section can be achieved by 'empty-bucket' channelling [11, 12]. The betatron core slowly accelerates the beam into the resonance (upwards in Fig. 26). The empty RF buckets are nearly stationary. If they were to be filled with beam, they would accelerate their beam in the opposite sense to the betatron core at the same speed. When the slowly moving beam meets the empty buckets, it has to stream round the buckets through a narrow phase-space gap at a much higher speed to maintain the flow. This is similar to the rapid movement of water between the piers of a bridge.

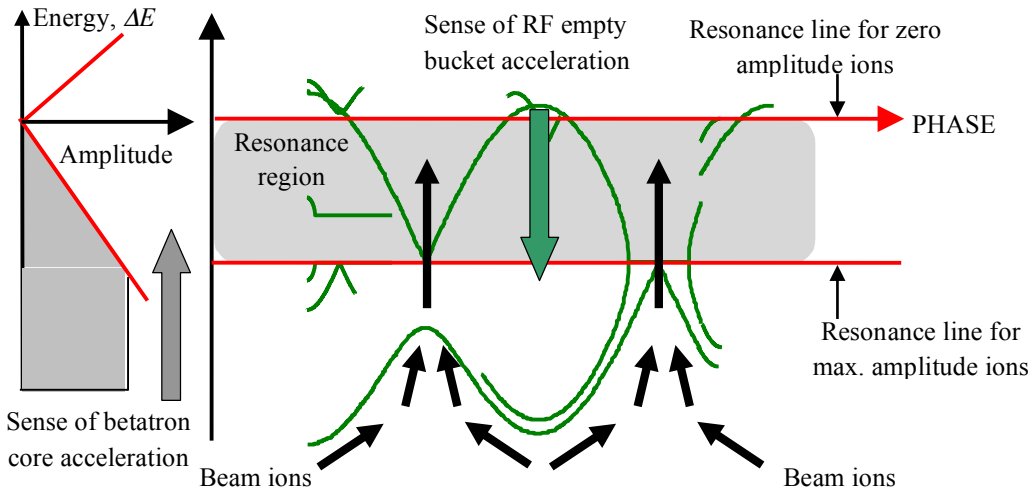


Fig. 26: Empty-bucket channelling

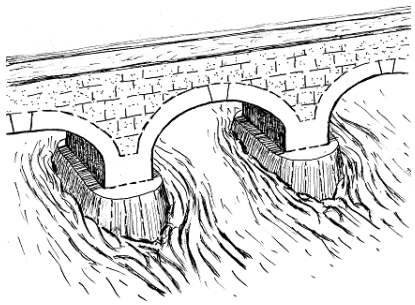


Fig. 27: The bridge analogy to empty-bucket channelling

22 Small horizontal beam size

The mono-energetic extracted beam segment (see Figs. 5 and 9) is typically 10 mm long and 0.1 mrad wide and is known as the 'bar of charge'. These dimensions are very approximate and depend on the power converter ripple and the optics errors as mentioned in section 19. For the momentum-driven extraction there is a continuum of segments of different momenta. If the Hardt condition is applied these segments are all superimposed and the 'bar of charge' is again 10 mm by 0.1 mm.

Those designing an extraction should note that the detailed optics at the entrance to the ES are rather complicated. At this position, there is one set of Twiss parameters that belongs to the ring. In the vertical plane these parameters are handed on to the extraction line, BUT NOT in the horizontal plane. In the horizontal plane, careful tracking will reveal that the segments of different momenta have different lengths and have their low momentum sides aligned. First one should choose a central segment to define the momentum and position of the central orbit for the extracted beam. Next the relative positions of the segments with respect to this central orbit define the dispersion vector for the entry to the extraction line. Finally, the rectangular shape of the segments is not convenient for fitting to a betatron ellipse. If one tries to fit inscribed or circumscribed ellipses to all the segments a rather complicated situation is created with potentially exotic optics. One solution is to forget ellipses and to treat the 'bar of charge' as the diameter of an unfilled ellipse, see Fig. 28.

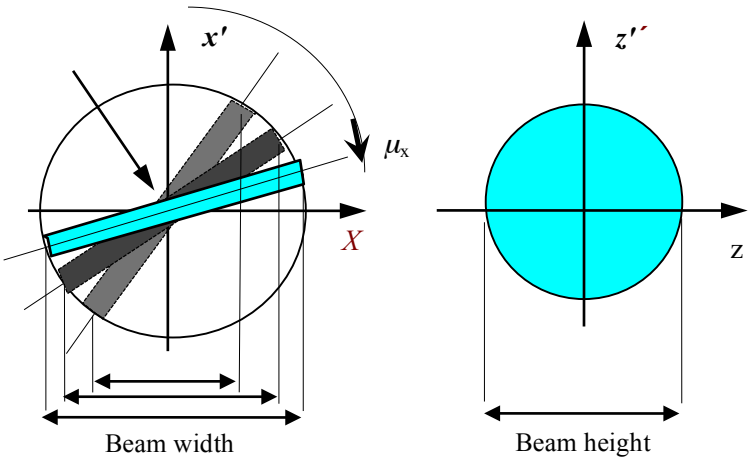


Fig. 28: Controlling beam width and height

Using the concept of an unfilled ellipse creates well-behaved optics that are relatively insensitive to changes in the width of the 'bar of charge' that will vary for example because ripple is more serious at low extraction energies than at top energy. For the unfilled ellipse, the width of the beam is defined by the projection of the 'bar of charge' on the *x*-axis. If a phase shifter is introduced in the line, then the

'bar' can be rotated. This rotation changes the beam width at the patient from a minimum value set by the width of the 'bar' up to a maximum value set by the length of the segment (see Fig. 28). The beam height is the projection of the Gaussian beam distribution in the vertical plane on the z-axis and can be changed by a 'betatron stepper'.

23 Scanning

As explained above, the extracted beam spot distributions are near-uniform over a narrow rectangle in the horizontal phase space and near-Gaussian in the vertical phase space. For voxel scanning, overlapping Gaussian spots give a smooth deposition of beam, whereas overlapping rectangles give 'hot/cold' ridges. The procedure is to scan with the Gaussian distribution aligned to the 'slow' scanning direction (from row to row), since errors are more likely to build up over time between rows and then to scan in the fast direction (along a row) taking mini-steps (say 1/5th of the voxel width). Since the beam spot is not perfectly rectangular, but has some rounding at the edges, this scheme works well.

24 Rotational matching to a gantry

In the extraction line of a medical facility, a rather special situation is met in which a fixed transfer line must be matched to a section of line called the gantry that has to be able to rotate through a full $0 \rightarrow 2\pi$ without affecting the beam spot at the patient, see Fig. 29. In cyclotron-based facilities, the problem is solved by making the beam distribution rotationally symmetric at the interface between the fixed and rotating lines. A rotationally symmetric beam requires equal emittances, equal lattice parameters, equal particle distributions and zero dispersion. Unfortunately, the slow-extracted beam from a synchrotron, as described in this lecture, is far from fulfilling these requirements (rectangular/Gaussian). A lot can be written on this topic about various restricted solutions and similar problems, but the only truly rotational solution that also includes the dispersion vector is the 'rotator' method attributed to Lee Teng [13]. This solution is extremely elegant and can be approached from a purely mathematical viewpoint.

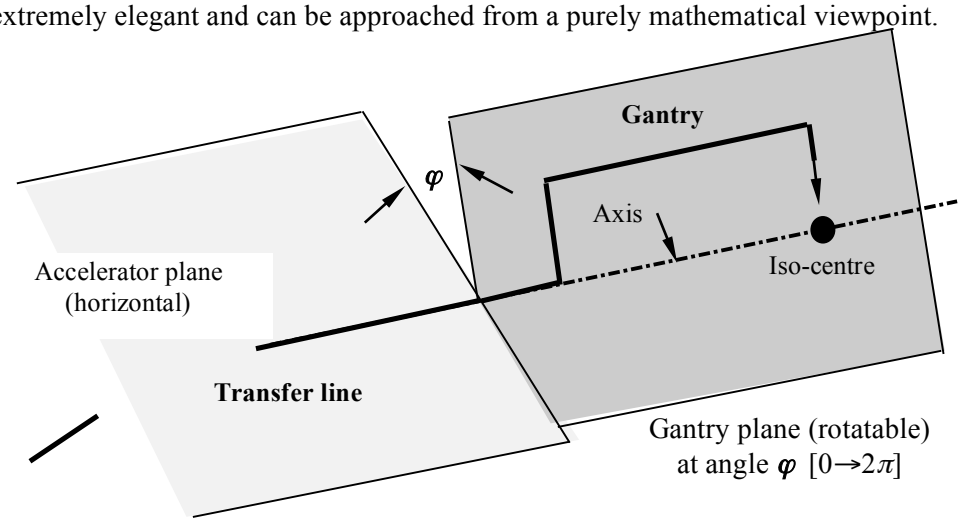


Fig. 29: Schematic view of an iso-centric gantry and transfer line

Consider a section of bending-free transfer line with a betatron phase advance of 2π in the x-plane and π in the z-plane. This section of line will be called a 'rotator'. According to the phase advances, the transfer matrix for the rotator will be;

$$M_{\text{rotator}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{60}$$

Let this line be rotated by an angle ν with respect to the fixed transfer line and let ν be just half of the rotation angle of the gantry, 2ν. A schematic view is shown in Fig. 30.

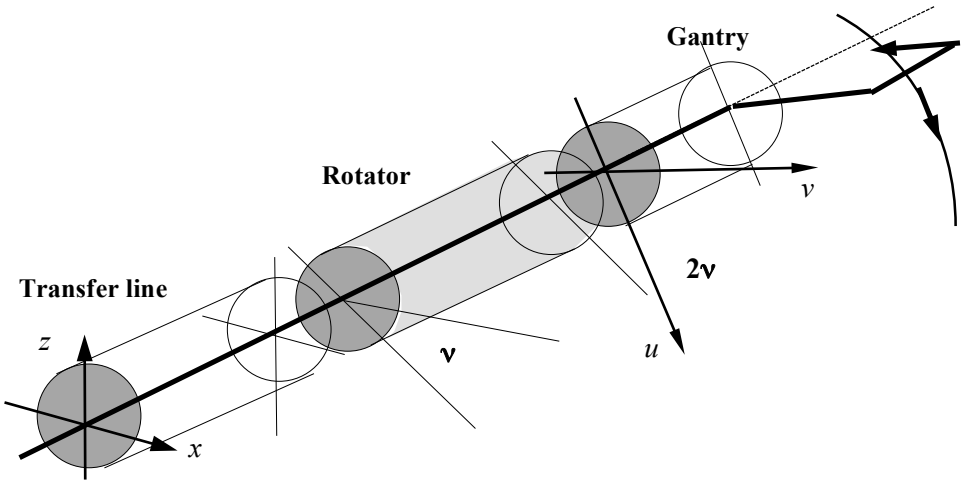


Fig. 30: Equipment layout for the rotator method

The transfer matrix from the exit of the fixed beam line to the interface point at the entry to the gantry is found by multiplying the rotation matrix for the angle ν, the rotator and a second rotation matrix also for the angle ν.

$$\begin{pmatrix} \cos\nu & 0 & \sin\nu & 0 \\ 0 & \cos\nu & 0 & \sin\nu \\ -\sin\nu & 0 & \cos\nu & 0 \\ 0 & -\sin\nu & 0 & \cos\nu \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \cos\nu & 0 & \sin\nu & 0 \\ 0 & \cos\nu & 0 & \sin\nu \\ -\sin\nu & 0 & \cos\nu & 0 \\ 0 & -\sin\nu & 0 & \cos\nu \end{pmatrix}$$

$$= \begin{pmatrix} \cos^2\nu + \sin^2\nu & 0 & \sin\nu\cos\nu - \cos\nu\sin\nu & 0 \\ 0 & \cos^2\nu + \sin^2\nu & 0 & \sin\nu\cos\nu - \cos\nu\sin\nu \\ -\sin\nu\cos\nu + \cos\nu\sin\nu & 0 & -\cos^2\nu - \sin^2\nu & 0 \\ 0 & -\sin\nu\cos\nu + \cos\nu\sin\nu & 0 & -\cos^2\nu - \sin^2\nu \end{pmatrix}.$$

After a simple reduction of the terms, the final overall transfer matrix maps the incoming normal modes directly to those of the gantry without any cross-coupling and independently of the rotation angle.

Rotator:

$$\begin{pmatrix} u \\ u' \\ v \\ v' \end{pmatrix}_{\text{Gantry}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ x' \\ z \\ z' \end{pmatrix}_{\text{Input}}. \tag{61}$$

Since a dispersion vector in a bending-free region behaves as a betatron oscillation to first order, the rotator will automatically match the dispersion vector (*D, D'*) into the gantry at the same time as it matches the normal modes. There are no fundamental limitations on the beam distributions or any of the Twiss or dispersion functions. The matching is entirely independent of the gantry rotation angle. This is important for the slow extraction from a synchrotron because the beam shape (rectangular/Gaussian) will always have the same orientation in the gantry nozzle whatever the gantry angle. Note, however, that the beam is coupled inside the rotator.

The rotator as presented is a purely mathematical solution. For example, four FODO cells with $\pi/2$ phase advance horizontally and $\pi/4$ phase advance vertically would satisfy the conditions, but would not be a practical solution because the beam sizes inside the rotator would be impractically large. To achieve well-behaved beam sizes within the rotator the trick is to make the Twiss alpha functions zero in both planes at the entry and exit. Figure 31 shows a well-behaved triple-triplet rotator design.

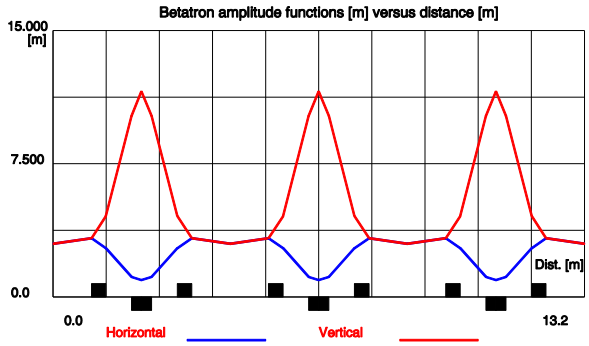


Fig. 31: Triple-triplet rotator ($\beta_x = \beta_z = 3$ m, $\alpha_x = \alpha_z = 0$, $L = 13.2$ m, $k_F = -k_D$)

25 Beam cut-off

Cutting the beam off is a security problem in medical machines. Amplitude-driven, slow extraction has the advantage that the RF excitation driving the amplitude growth can be switched off quickly and since the extraction works only with large amplitude ions the transit time in the resonance is short making the overall switching time a few tens of microseconds, whereas, momentum-driven, slow extraction is a little slower to respond. Not only is the betatron core slower than RF, there are a few small amplitude ions trapped in the resonance that can take a millisecond or two to transit. In both cases, the switch-off can be made to satisfy treatment safety, but to satisfy the wider needs of security there has to be an independent cut-off mechanism that is less reliant on software links.

The usual solution is to place four identical ferrite dipoles powered in series in a single straight section in the transfer line to excite a closed-orbit bump that holds the beam clear of a dump block, see Fig. 32. By switching off the dipoles, or in the event of a power failure, the beam drops onto the dump. The switching time can be less than 200 microseconds.

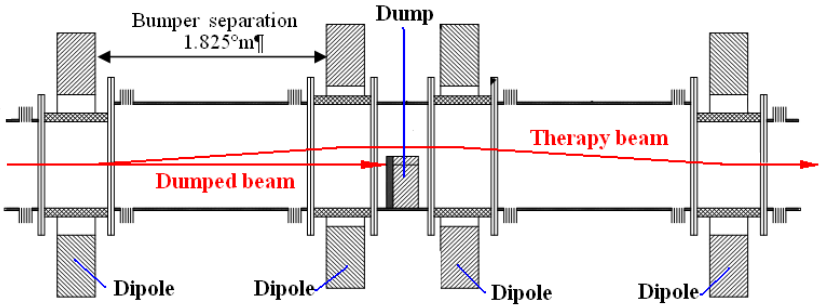


Fig. 32: Beam chopper

References

[1] The possibility of resonant extraction from the CPS, CERN-AR-Int-GS-61-5.

[2] C.L. Hammer and L.J. Laslett, *Rev. Sci. Instr.* **32** (1962) 144−149, https://doi.org/10.1063/1.1717299.

[3] Y. Kobayashi and H. Takahashi, Improvement of the emittance in the resonant ejection, Proc. VIth Int. Conf. on High Energy Accelerators, Cambridge, Massachusetts, 1967, pp. 347−51.

[4] M.Q. Barton, Beam extraction from synchrotrons, Proc. VIIIth Int. Conf. on High Energy Accelerators, CERN, Geneva, 1971, pp. 85−8.

[5] G. Guignard, A general treatment of resonances in accelerators, CERN 78-11 (CERN, Geneva, 1978), https://doi.org/10.5170/CERN-1978-011.

[6] K. Hiramoto and M. Nishi, *Nucl. Instr. Meth. in Phys. Res.* **A3320** (1992), p. 154, https://doi.org/10.1016/0168-9002(92)90023-W.

[7] PIMMS, Proton-ion medical machine study, Study Leader and Editor P.J. Bryant, CERN 2000-006, (CERN, Geneva, 2000), https://doi.org/10.5170/CERN-2000-006.

[8] Ultraslow extraction out of LEAR (transverse aspects*),* CERN Internal Note PS/DL/LEAR Note 81-6, (1981).

[9] M. Pullia, Transit time for third order resonance extraction, CERN/PS 96-36 (DI).

[10] M. Pullia, Time profile of the slowly extracted beam, CERN/PS 97-50 (DI).

[11] R. Cappi and C. Steinbach, Low Frequency Duty Factor Improvement for the CERN PS Slow Extraction Using RF Phase Displacement Techniques, 1981 Particle Accelerator Conference, Washington, 11−13 March 1981.

[12] RF Empty Bucket Channelling Combined With a Betatron Core to Improve Slow Extraction in Medical Synchrotrons, CERN/PS 97-68 (DI).

[13] L.C. Teng, Private communication, Laboratory notebook (Jan. 1970) and Int. Rep. LL-134 (Oct. 1986).

Proceedings of the CAS–CERN Accelerator School: Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018)

Emittance Preservation

V. Kain
CERN, Geneva, Switzerland

Abstract
The transverse emittance is a critical parameter for the performance of synchrotron light sources, linear and circular colliders, and many other applications. The best performance is often achieved with the smallest emittances. Many effects lead however to emittance growth. In this lecture emittance growth from mismatch at injection into circular machines will be discussed. The concept of emittance and how to calculate it will be introduced and the formulae for emittance blow-up for the most common injection errors will be derived.

Keywords
Emittance; brightness; accelerator performance; emittance blow-up.

1 Introduction

One of the key figures of merit in circular and linear colliders is the collider *luminosity*. It is defined as

$$\mathcal{L} = \frac{N_+ N_- f}{2\pi \Sigma_x \Sigma_y}\,, \tag{1}$$

where $\Sigma_{x,y} = \sqrt{\sigma^{*2}_{x,y+} + \sigma^{*2}_{x,y-}}$ describes the convolution of the colliding particle distributions with horizontal and vertical beam sizes $\sigma^*_{x,y}$ at the interaction point, N_+, N_- the number of particles in the two beams, and f the revolution frequency. In the case of round Gaussian beams and the same distribution for both beams, the formula for luminosity simplifies to

$$\mathcal{L} = \frac{N_+ N_- f}{4\pi \beta^* \varepsilon}\,, \tag{2}$$

where β^* is the β function at the collision point and ε is the emittance of the beam. The luminosity thus depends directly on the emittance in the vertical and the horizontal plane. The smaller the emittance, the higher the luminosity and the higher is the collision rate at the experiment.

As long as the dynamics of the particle beams is governed by conservative forces, the area occupied by the beam in phase space remains constant. We will see later that the area in phase space that is occupied by the beam is related to the emittance of the beam. Many effects exist nevertheless that alter the distribution in phase space. Synchrotron radiation, for example, will lead to a decrease of the emittance. Various beam cooling mechanisms, such as stochastic and electron cooling, are also available in specific circumstances to decrease the beam emittance in a controlled manner. A big class of effects, however, leads to an uncontrolled increase of the emittance. Intra-beam scattering, power supply noise, crossing resonances, beam instabilities, alignment errors, and dispersion for e^- linacs are examples of these issues; another category of emittance growth is due to errors at injection into synchrotrons or linacs, where we typically speak of *mismatch* of various parameters. In this lecture we will discuss how mismatch at injection can introduce emittance blow-up. We will start with a general definition and discussion of emittance and how to calculate it for a given particle distribution. This will then be input to calculate emittances for altered particle distributions due to mismatch. The effects will be explained mostly referring to the horizontal plane. The same considerations can be applied to the vertical plane. The material used for the preparation of these lectures is mainly from Refs. [1–3].

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-005.345

2 Emittance

For many calculations and applications, the canonical conjugate variable pair called *action-angle variables* (J, ϕ) prove to be a practical concept. The relation between Cartesian (x and x') and action-angle variables in the horizontal plane is given in Fig. 1. It shows the well-known elliptic particle trajectory in horizontal phase space. The area A of the phase-space ellipse is proportional to action J_x via $A = 2\pi J_x$.

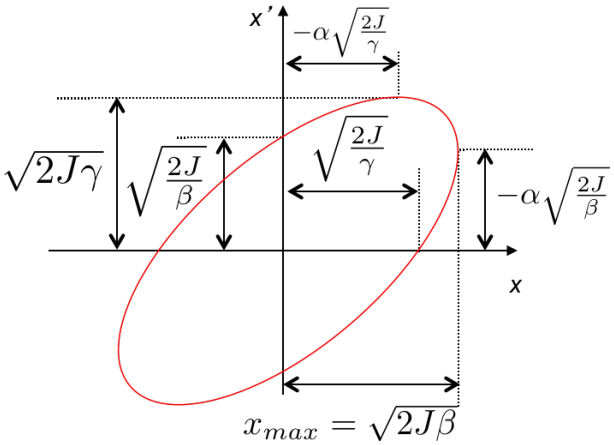


Fig. 1: Phase-space in the horizontal plane. The area of the ellipse defines the action J_x, with $A = 2\pi J_x$

The transformation between action-angle variables and Cartesian variables is summarised in Eqs. (3) and (4). The action J_x is a measure of the particle amplitude of motion, see Eq. (4).

$$\begin{aligned} 2J_x &= \gamma_x x^2 + 2\alpha_x x' x + \beta_x x'^2 \\ \tan\phi_x &= -\beta_x x \frac{x}{x'} - \alpha_x \end{aligned} \tag{3}$$

$$\begin{aligned} x &= \sqrt{2\beta_x J_x}\cos\phi_x \\ x' &= -\sqrt{\tfrac{2J_x}{\beta_x}}(\sin\phi_x + \alpha_x \cos\phi_x)\,. \end{aligned} \tag{4}$$

The advantage of action-angle variables becomes apparent when observing them under symplectic transformations as for example under transfer matrices. Symplectic operations M are defined as

$$M^T \cdot S \cdot M = S \tag{5}$$

where S is defined as

$$S = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 0 \end{pmatrix}. \tag{6}$$

Symplectic operations are phase-space area-preserving, and thus action-preserving, operations. The concept of action-angle variables is also useful for our discussion as it allows to define the emittance in very simple terms. The emittance is defined as the mean action of all particles in the beam:

$$\varepsilon_x = \langle J_x \rangle\,. \tag{7}$$

It is related to the width of the transverse distribution and can hence also be calculated from the second-order moments of the distribution in phase space with the σ-matrix

$$\sigma = \begin{pmatrix} \langle x^2 \rangle & \langle x x' \rangle \\ \langle x x' \rangle & \langle x'^2 \rangle \end{pmatrix} \tag{8}$$

and the emittance as

$$\varepsilon_x = \sqrt{|\sigma|} = \sqrt{\langle x^2\rangle\langle x'^2\rangle - \langle xx'\rangle^2}\,. \tag{9}$$

We refer to this definition as rms-emittance. For phase-space distributions that are centred around 0 in x and x', both definitions, (7) and (9), are equivalent.

3 Mismatch at injection

Injection into a circular accelerator is a delicate process. Imprecise delivery of the beam for instance or mismatch in the optics parameters lead to emittance blow-up and possibly to losses. In the following, the methodology to calculate the emittance blow-up due to injection errors will be introduced for the most important and most typical imperfections. The treatment can then be applied to other problems. Most illustrations and calculations make use of *normalised coordinates*, another practical concept. In normalised phase space the elliptical trajectories from real phase space become circles, see Fig. 2. The transformation from real phase-space coordinates x, x' to the normalised coordinates $\overline{x}, \overline{x}'$ is given in Eq. (10)

$$\begin{pmatrix}\overline{x}\\ \overline{x}'\end{pmatrix} = N\cdot\begin{pmatrix}x\\ x'\end{pmatrix} = \sqrt{\frac{1}{\beta_x}}\cdot\begin{pmatrix}1 & 0\\ \alpha_x & \beta_x\end{pmatrix}\cdot\begin{pmatrix}x\\ x'\end{pmatrix}\,. \tag{10}$$

Whereas the equation of the ellipse for real phase space is,

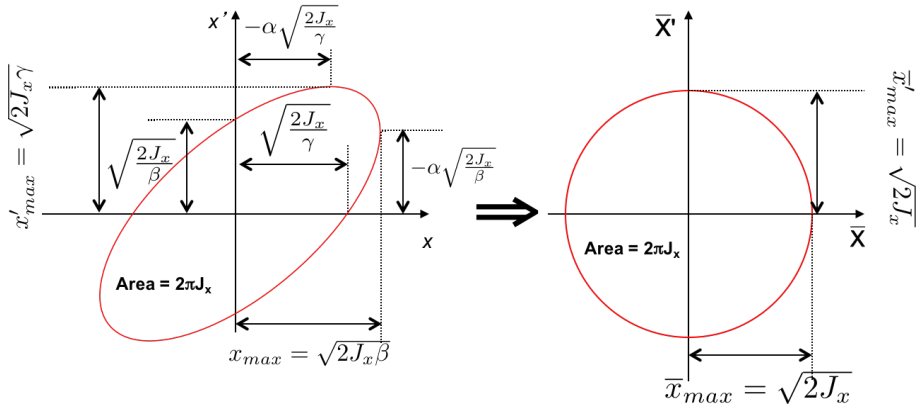


Fig. 2: In normalised phase space the trajectories are circles. The area stays the same and is linked to the action of the particle.

$$2J_x = \gamma_x x^2 + 2\alpha_x p_x + \beta_x p_x^2\,, \tag{11}$$

it becomes the equation of a circle in normalised phase space

$$2J_x = \overline{x}^2 + \overline{x}'^2\,. \tag{12}$$

The area enclosed by the trajectory in real as well as normalised phase space is $A = 2\pi J_x$.

3.1 Steering (dipole) error

If the beam is not injected on the closed orbit as indicated in Fig. 3, i.e., $x \neq x_{CO}$ and/or $x' \neq x'_{CO}$, the injected beam will perform so-called injection oscillations around the closed orbit for several turns until the beam has filamented and the emittance has effectively grown. In phase space at the location of the injection point, a steering error will lead to a particle distribution that is off-centred with respect to the matched distribution as indicated in Fig. 4. In Fig. 5 snapshots of the evolution of the mismatched

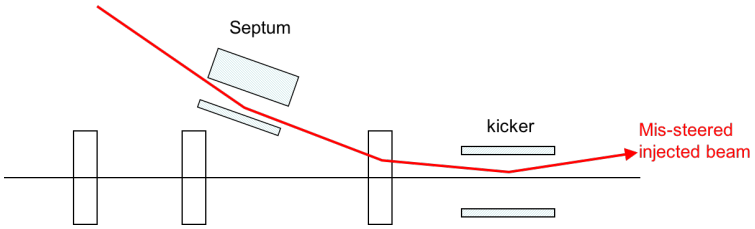


Fig. 3: The injected beam arrives with an offset in position and angle at the injection point. This will result in injection oscillations.

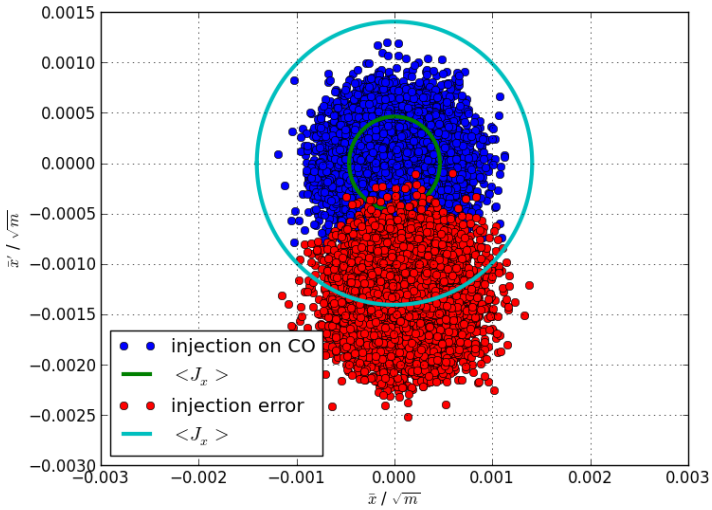


Fig. 4: The blue particle distribution is beam injected without steering error, whereas the red one arrives with an injection angle error. The mean action for both distributions is also indicated.

distribution (red distribution in figure) turn after turn are shown. In this example the particles experience only linear fields and the centroid of the red particle distribution will move on a circular trajectory. The centroid of the distribution is performing *injection oscillations*, as it can be measured turn-by-turn by a single beam position monitor, Fig. 6, while the width of the distribution remains constant, Fig. 7. Figure 8 finally shows the evolution of the emittance turn-by-turn. Injection is supposed to happen at turn 20 in that figure. Both definitions (7) and (9) are applied to the red particle distribution and shown in Fig. 8. Whereas the rms-emittance corresponds to what we see in Fig. 7, the width of the distribution remains constant, the usefulness of the emittance via mean action does not seem apparent. This will be different in the presence of non-linear fields in the circular accelerator, where the particles' tune will depend on their amplitude.

In case of non-linear fields, the area occupied by the particle distribution in phase space will change, Fig. 9. *Filamentation* will take place until the distribution will cover again all phases, see Fig. 10. The projection into the horizontal plane after many turns, as it can be measured by a profile monitor, is shown in Fig. 11. It is no longer Gaussian. Smaller injection errors than in the example just lead to the creation of non-Gaussian tails, a phenomenon which is frequently seen in day-to-day accelerator operation. At the same time as the filamentation occurs, the oscillation amplitude of the centroid will reduce. In fact the oscillation amplitude will decay exponentially following $A(t) = A_0 \cdot e^{-\frac{t}{\tau}}$, where τ is the filamentation time. The *damped* motion of the centroid under the influence of non-linear fields

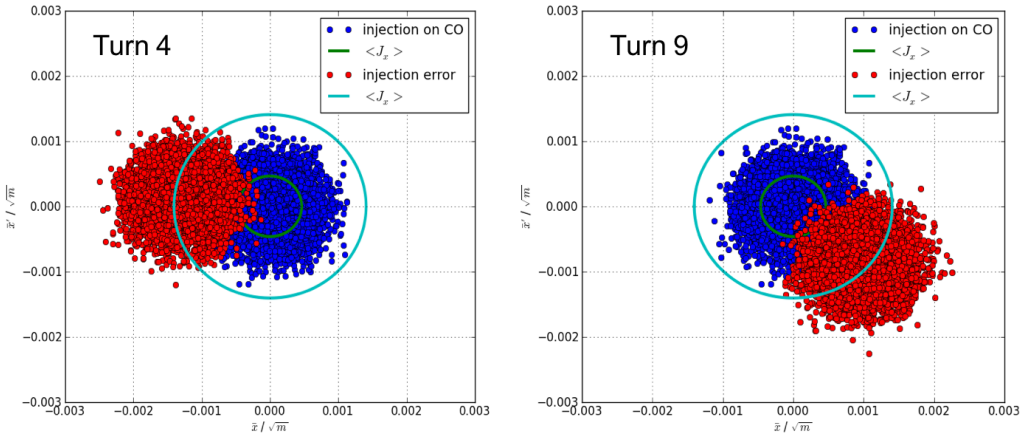


Fig. 5: The centroid of the particle distribution will move on a circular trajectory in normalised phase space turn after turn.

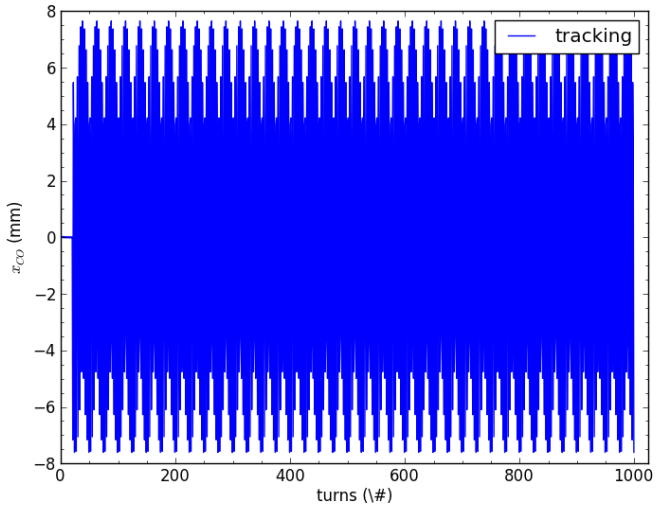


Fig. 6: Turn-by-turn beam position measurements with a beam position monitor reveal *injection oscillations* in case of injection with steering error.

is shown in Fig.12. As before, we can plot the evolution of the emittance using the mean action of the particle distribution and the rms-emittance as defined in (9), see Fig. 13. The important conclusion from the short phenomenological introduction of the effect on the emittance growth from an injection steering error is that the emittance from mean action corresponds to the rms-emittance after full filamentation. Or as said earlier the emittance definition from mean action is equivalent to the rms-emittance in case of a particle distribution where mean angle and position are zero.

This input will now be used to derive the formula to calculate the emittance blow-up for a given injection steering error. A collection of particles is injected with an error in position and angle as indicated in Fig. 14. For an injection error $\Delta a[\sigma]$ in units of $\sigma = \sqrt{\beta\varepsilon}$, the misinjected beam is offset in normalised phase space by $L = \Delta a\sqrt{\varepsilon}$. The particle coordinates of the misinjected beam with respect

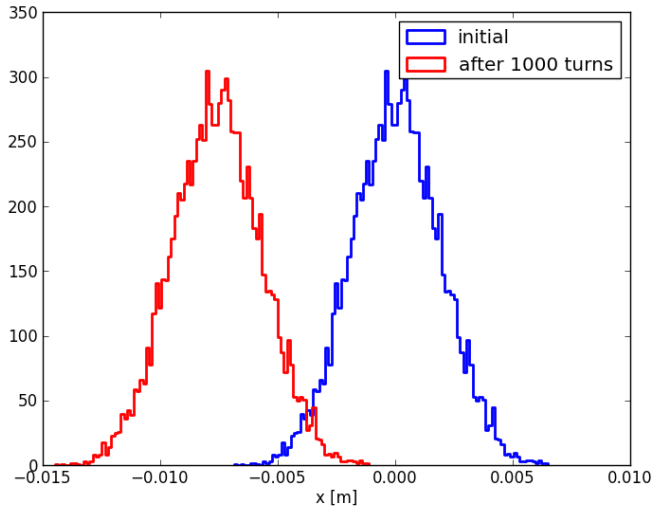


Fig. 7: Horizontal distributions for different turns as could be measured by a turn-by-turn profile monitor. The width of the distribution stays the same, but the mean is changing form turn to turn.

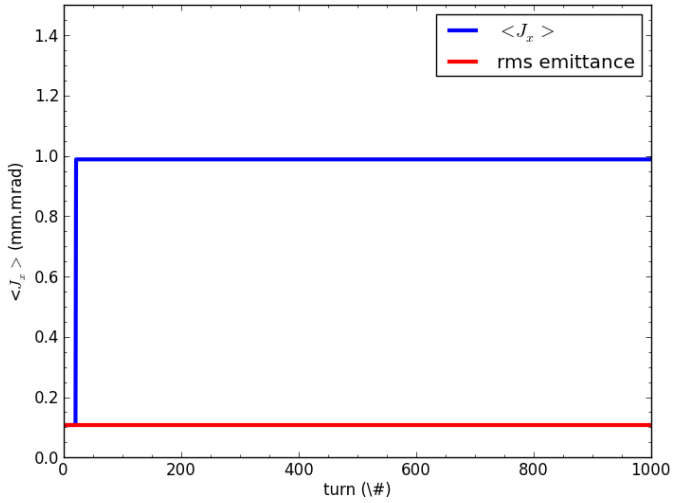


Fig. 8: Evolution of mean action and rms emittance as a function of turn. The injection occurs at the time turn = 20.

to the matched distribution are

$$\begin{aligned} \overline{x}_{new} &= \overline{x}_0 + L\cos\theta \\ \overline{x}'_{new} &= \overline{x}'_0 + L\sin\theta \end{aligned} \tag{13}$$

and the action for one particular particle becomes

$$\begin{aligned} 2J_{new} &= \overline{x}^2_{new} + \overline{x}'^2_{new} = (\overline{x}_0 + L\cos\theta)^2 + (\overline{x}'_0 + L\sin\theta)^2 \\ &= \overline{x}^2_0 + \overline{x}'^2_0 + 2L(\overline{x}_0\cos\theta + \overline{x}'_0\sin\theta) + L^2\,. \end{aligned} \tag{14}$$

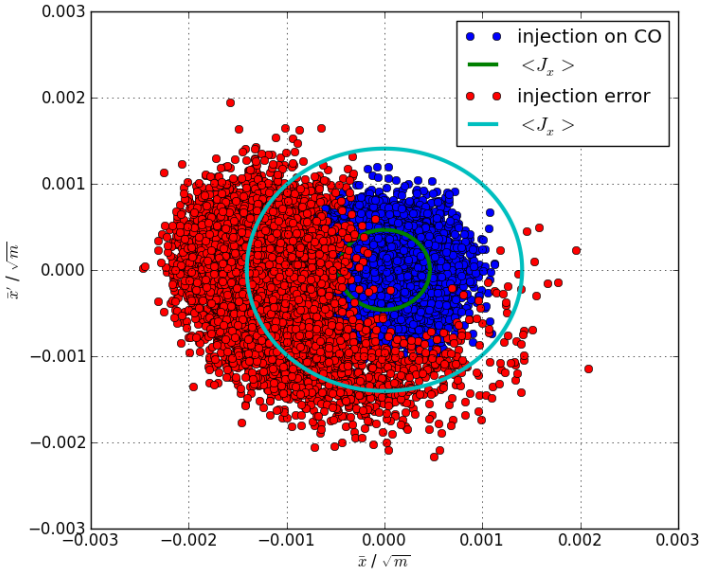


Fig. 9: Particle distribution with injection error in red after 100 turns in the presence of non-linear fields in the circular accelerator. The distribution is filamenting.

For the emittance we need the mean action of all particles:

$$\begin{aligned} 2\langle J_{new}\rangle &= \langle \overline{x}_0^2\rangle + \langle \overline{x}_0'^2\rangle + \langle 2L(\overline{x}_0\cos\theta + \overline{x}_0'\sin\theta)\rangle + L^2 \\ &= 2\varepsilon_0 + 2L(\langle \overline{x}_0\cos\theta\rangle + \langle \overline{x}_0'\sin\theta\rangle) + L^2 \\ &= 2\varepsilon_0 + L^2 \end{aligned} \tag{15}$$

using $\langle\cos\theta\rangle = 0$ and $\langle\sin\theta\rangle = 0$. And the emittance increase is hence

$$\begin{aligned} \varepsilon_{new} &= \langle J_{new}\rangle = \varepsilon_0 + L^2/2 \\ &= \varepsilon_0(1 + \Delta a^2/2)\,. \end{aligned} \tag{16}$$

Equation (16) can also be written for real phase-space coordinates, where Δx and $\Delta x'$ are the errors at injection with respect to the horizontal closed orbit at injection:

$$\frac{\varepsilon}{\varepsilon_0} = 1 + \frac{1}{2}\frac{\Delta x^2 + (\beta\Delta x' + \alpha\Delta x)^2}{\beta\varepsilon_0}\,. \tag{17}$$

As we have seen now, injection oscillations are detrimental for emittance preservation and therefore need to be corrected. Frequently the input for the correction algorithm is not the oscillation measured at a single pick up over many turns, but rather the difference of the first turn trajectory from the injection closed orbit measured with many pick ups. In Fig. 15 an example of the injection oscillation display in the LHC control room is shown. It contains the trajectory reading in the transfer line between the LHC pre-injector and the LHC as well as the injection oscillations in the first 3 km long part of the LHC. The usual trajectory correction algorithms are used to find the best dipole correctors in the transfer line to minimise the injection oscillation amplitudes in the LHC. The horizontal injection oscillation amplitude in the example is 2 mm, corresponding roughly to 2 σ. Let's apply Eq. (16) to calculate the emittance blow-up for $\Delta a = 2\sigma$. According to Eq. (16) the emittance blow-up will be $\varepsilon_{new} = \varepsilon_0(1 + \Delta a^2/2) = 3\varepsilon_0$. For comparison, the allowed emittance growth budget in the LHC is only 10%. Clearly, these injection oscillations would need to be corrected.

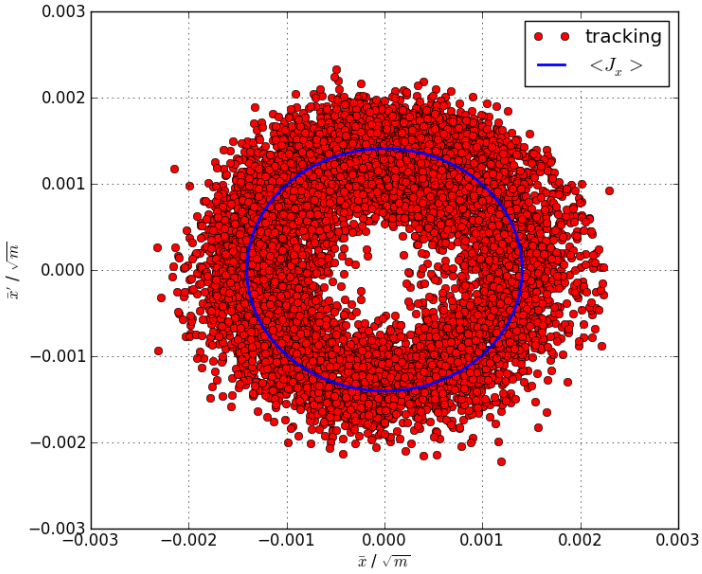


Fig. 10: After filamentation the beam covers all phases in phase space and the area is larger than initially

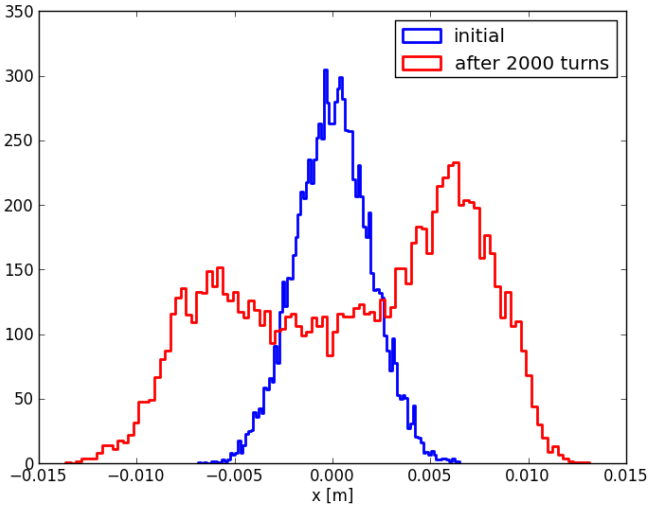


Fig. 11: Filamentation leads to a larger beam size measured at a profile monitor. The distribution has also become non-Gaussian. The mean of the distribution remains constant after filamentation.

The kind of injection oscillation correction discussed so far will however only remove the static contributions to the steering errors. Small variations of the injected trajectory might also occur on a shot-by-shot basis or if the injected beam consists of several bunches, the equipment involved in the transfer might introduce slightly different trajectories and hence different injection oscillations for the different bunches. These injection errors can only be corrected by the transverse feedback system in the receiving circular machine. The feedback damping time needs to be faster than the filamentation time to avoid emittance blow-up.

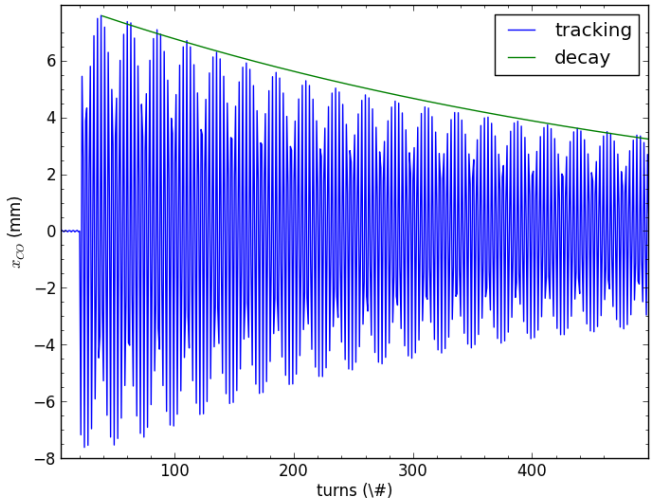


Fig. 12: In the presence of non-linear fields, damping of the centroid motion occurs after an injection error. The decay of the oscillation amplitudes can be fitted with $A(t) = A_0 \cdot e^{-\frac{t}{\tau}}$, where τ is the filamentation time.

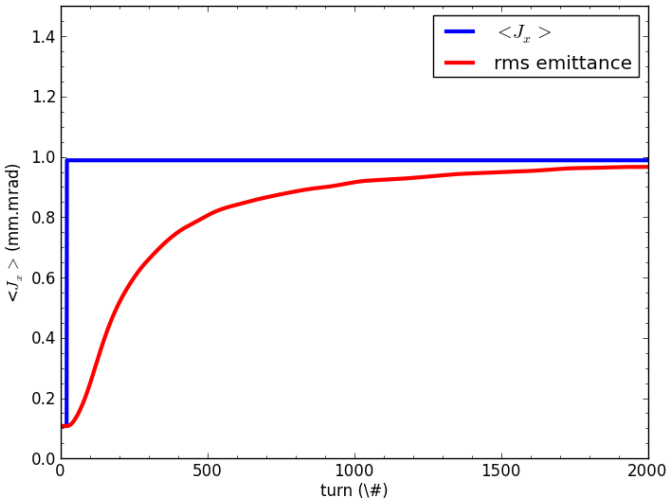


Fig. 13: The evolution of the emittance versus turn using the definition of mean action as well as rms-emittance for an injection steering error in the presence of non-linear fields where the beam distribution is filamenting.

The principal behind a transverse feedback is illustrated in Fig. 16. The trajectory excursion is measured with one or more beam position monitor(s) at a certain position along the ring and the required correction is calculated in a processing unit and sent to a fast transverse kicker that will correct the trajectory error with respect to the orbit. Depending on the bandwidth of the system even bunch-by-bunch up to intra-bunch oscillations can be damped in this manner. The result of the action of a transverse feedback on the injection error and the resulting injection oscillations is shown in Fig. 17 and the effect on the emittance in Fig. 18. The emittance will nevertheless grow in the presence of a transverse damper as it can only correct the oscillations with limited gain. The reduction in growth with respect to the situation without damper depends on the ratio of filamentation time τ_{DC} and damping time

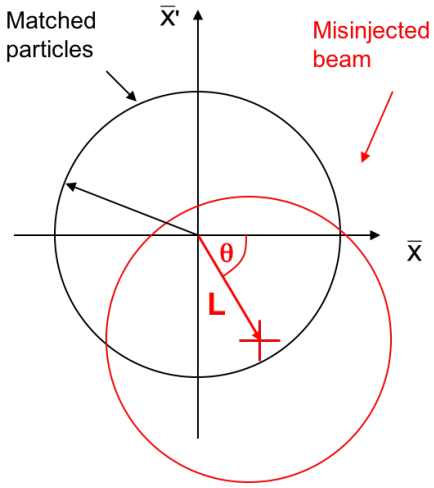


Fig. 14: The misinjected beam is offset in normalised phase space by $L = \Delta a\sqrt{\varepsilon}$ with respect to the matched distribution.

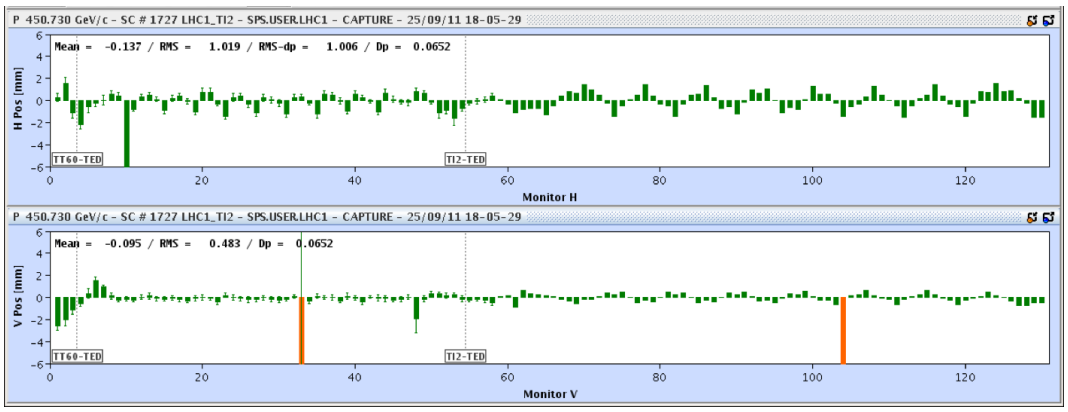


Fig. 15: A screen shot of the injection oscillation display for LHC beam 1. It shows the transfer line trajectory in the horizontal and vertical plane for the transfer between the LHC injector and the LHC itself, as well as the first turn at the pick ups of the first 3 km in the LHC minus the closed orbit at injection energy. The plot spans 6 km.

τ_d of the feedback system. The emittance blow-up for injection error Δx and $\Delta x'$ in the presence of a transverse feedback is [4]

$$\frac{\varepsilon}{\varepsilon_0} = 1 + \frac{1}{2}\frac{\Delta x^2 + (\beta\Delta x' + \alpha\Delta x)^2}{\beta\varepsilon_0}\left(\frac{1}{1+\tau_{DC}/\tau_d}\right)^2 . \tag{18}$$

3.2 Emittance blow-up from betatron mismatch

Optical errors can occur in the transfer line and the ring, such that the beam can be injected with a mismatch. The shape of the injected beam then corresponds to different twiss parameters α and β than the closed solution of the ring at the injection point. At the moment of injection the area of the particle distribution in phase space might be the same as for the desired matched distribution, see Fig. 19, only the orientation is different. Filamentation will introduce an emittance increase. The coordinates of the

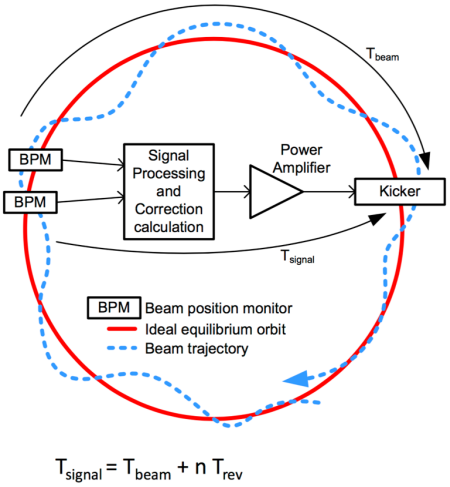


Fig. 16: The principle of the transverse feedback system: the trajectory excursion is measured with one or more beam position monitor(s) at a certain position along the ring and the required correction is calculated in a processing unit and sent to a fast transverse kicker that will correct the trajectory error with respect to the orbit.

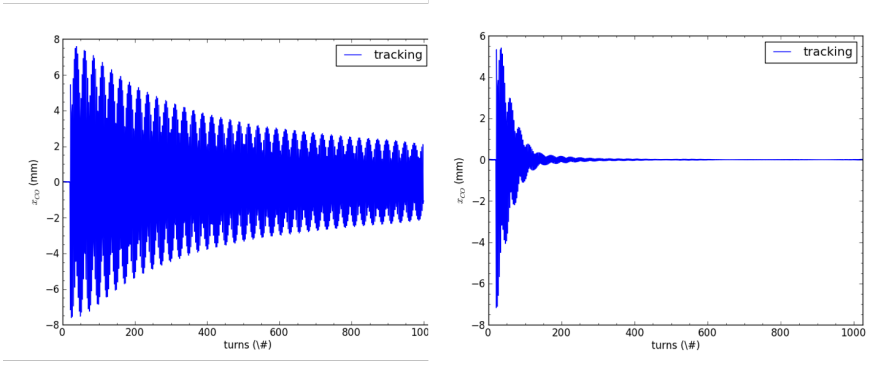


Fig. 17: The result of the action of a transverse feedback on the injection error and the resulting injection oscillations: the left plot shows the injection oscillations without transverse feedback and the left plot with. The damping time is faster than the filamentation time.

particles on the injected ellipse are

$$\begin{aligned} x_2 &= \sqrt{2\beta_2 J_x}\cos\phi \\ x_2' &= -\sqrt{\tfrac{2J_x}{\beta_2}}(\sin\phi + \alpha_2\cos\phi)\,. \end{aligned} \tag{19}$$

Transforming the coordinates in Eq. (19) to the normalised phase space defined by the matched twiss parameters α_1, β_1 results in an ellipse in normalised phase space and not a circle.

$$\begin{pmatrix} \overline{x_2} \\ \overline{x_2}' \end{pmatrix} = \sqrt{\frac{1}{\beta_1}} \cdot \begin{pmatrix} 1 & 0 \\ \alpha_1 & \beta_1 \end{pmatrix} \cdot \begin{pmatrix} x_2 \\ x_2' \end{pmatrix}\,. \tag{20}$$

The ellipse is defined by

$$2J_x = \overline{x}_2^2[\frac{\beta_1}{\beta_2} + \frac{\beta_2}{\beta_1}(\alpha_1 - \alpha_2\frac{\beta_1}{\beta_2})^2] + \overline{x'}_2^2\frac{\beta_2}{\beta_1} - 2\overline{x}_2\overline{x}_2'[\frac{\beta_2}{\beta_1}(\alpha_1 - \alpha_2\frac{\beta_1}{\beta_2})]\,, \tag{21}$$

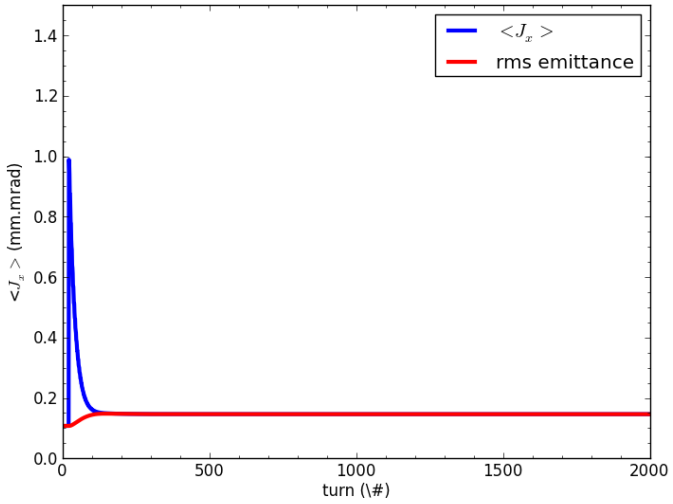


Fig. 18: The evolution of the emittance versus turn for an injection steering error using the definition of mean action as well as rms-emittance in the presence of non-linear fields and transverse feedback system. The final emittance blow-up depends on the ratio of filamentation time versus damping time τ_{DC}/τ_d.

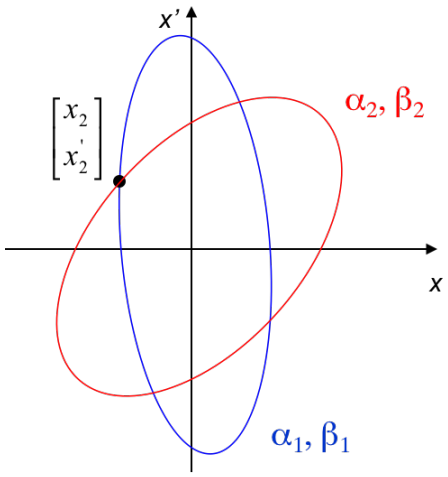


Fig. 19: Betatron mismatch illustrated in real phase space. The blue ellipse represents the matched twiss parameters (α_1, β_1), the injected beam (red curve) arrives however on an ellipses describing different twiss parameters (α_2, β_2).

which can be characterised by γ_{new}, β_{new} and α_{new} following Eq. (11):

$$\begin{aligned} \alpha_{new} &= -\frac{\beta_2}{\beta_1}(\alpha_1 - \alpha_2\frac{\beta_1}{\beta_2}) \\ \beta_{new} &= \frac{\beta_2}{\beta_1} \\ \gamma_{new} &= \frac{\beta_1}{\beta_2} + \frac{\beta_2}{\beta_1}(\alpha_1 - \alpha_2\frac{\beta_1}{\beta_2})^2 \,. \end{aligned} \tag{22}$$

From the general ellipse properties [5], the parameters a, b and λ can be introduced (Fig. 20), where a and b are the half axes of the ellipse which are related to the radius of the matched circle $A = \sqrt{2J}$ via

$$\begin{aligned} a &= A/\lambda \\ b &= A \cdot \lambda \end{aligned} \tag{23}$$

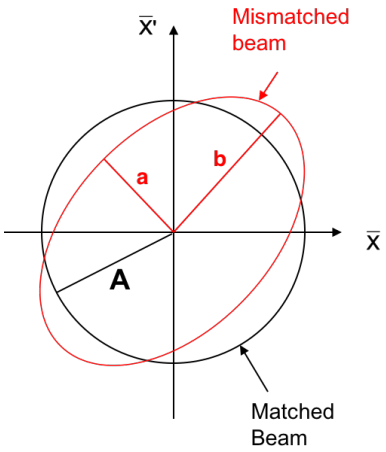


Fig. 20: Betatron mismatch in normalised phase space: The mismatched beam is enclosed by an ellipse. The half axes of the ellipse are related to the matched circle of same area with $a = A/\lambda$ and $b = A \cdot \lambda$.

and

$$\begin{aligned} a &= \frac{A}{\sqrt{2}}(\sqrt{H+1}+\sqrt{H-1}) \\ b &= \frac{A}{\sqrt{2}}(\sqrt{H+1}-\sqrt{H-1}) \end{aligned} \tag{24}$$

where

$$H = \frac{1}{2}(\gamma_{new}+\beta_{new}) = \frac{1}{2}(\frac{\beta_1}{\beta_2}+\frac{\beta_2}{\beta_1}(\alpha_1-\alpha_2\frac{\beta_1}{\beta_2})^2+\frac{\beta_2}{\beta_1}) \tag{25}$$

such that

$$\begin{aligned} \lambda &= \frac{1}{\sqrt{2}}(\sqrt{H+1}+\sqrt{H-1}) \\ \frac{1}{\lambda} &= \frac{1}{\sqrt{2}}(\sqrt{H+1}-\sqrt{H-1})\,. \end{aligned} \tag{26}$$

The coordinates of the mismatched distribution in normalised phase space at the injection point can then be written simply as

$$\begin{aligned} \overline{x}_{new} &= \lambda \cdot A\sin(\phi+\phi_1) \\ \overline{x}'_{new} &= \frac{1}{\lambda} \cdot A\cos(\phi+\phi_1)\,. \end{aligned} \tag{27}$$

After filamentation the particles' trajectories will be circles concentric to the circle of the matched distribution. The equation of the circle for one particular particle with the new coordinates is

$$2J_{new} = \overline{x}^2_{new} + \overline{x}'^2_{new} = \lambda^2 \cdot 2J_0\sin^2(\phi+\phi_1) + \frac{1}{\lambda^2}2J_0\cos^2(\phi+\phi_1)\,. \tag{28}$$

The emittance after filamentation is again the average over all phases, where we use that $\langle\sin^2(\phi+\phi_1)\rangle = \langle\cos^2(\phi+\phi_1)\rangle = 1/2$

$$\begin{aligned} \varepsilon_{new} = \langle J_{new}\rangle &= \tfrac{1}{2}(\lambda^2\langle 2J_0\sin^2(\phi+\phi_1)\rangle + \tfrac{1}{\lambda^2}\langle 2J_0\cos^2(\phi+\phi_1)\rangle) \\ &= \langle J_0\rangle(\lambda^2\langle\sin^2(\phi+\phi_1)\rangle + \tfrac{1}{\lambda^2}\langle\cos^2(\phi+\phi_1)\rangle) \\ &= \tfrac{1}{2}\varepsilon_0(\lambda^2+\tfrac{1}{\lambda^2})\,. \end{aligned} \tag{29}$$

For the final result λ can be substituted with the twiss parameters of the matched and the mismatched distribution:

$$\varepsilon_{new} = \frac{1}{2}\varepsilon_0(\lambda^2+\frac{1}{\lambda^2}) = H\cdot\varepsilon_0 = \frac{1}{2}\varepsilon_0(\frac{\beta_1}{\beta_2}+\frac{\beta_2}{\beta_1}(\alpha_1-\alpha_2\frac{\beta_1}{\beta_2})^2+\frac{\beta_2}{\beta_1}) \tag{30}$$

where subscript 1 refers to the matched ellipse and 2 to the mismatched one.

An example of a turn-by-turn profile measurement with betatron mismatch is shown in Fig. 21. Before the beam has fully filamented, the beam size measured by the instrument is oscillating indicating mismatch.

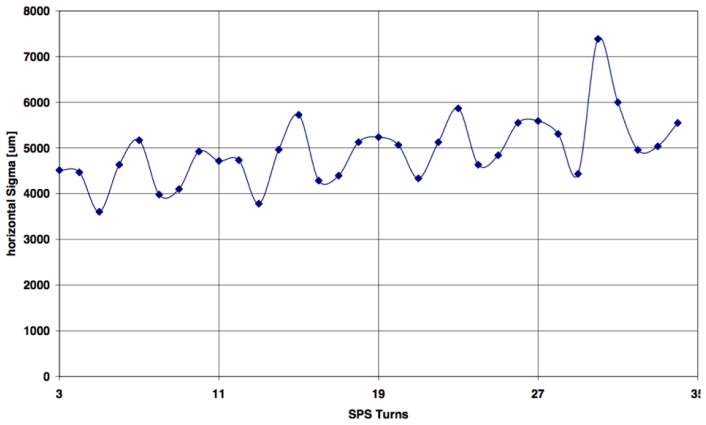


Fig. 21: Measurement of beam size versus turn after injection in the horizontal plane measured with a turn-by-turn profile monitor. The oscillation of the beam size indicates mismatch. The slope of the curve is due to emittance growth from the scattering in the foil [6].

3.3 Emittance blow-up from thin scatterer

It is sometimes necessary to have thin scatterers in the beam as thin beam screens of Al_2O_3 or Ti used to generate beam profile measurements, metal windows used to separate vacuum of transfer lines from vacuum in circular machines or foils to strip electrons to change charge state. The emittance of the beam that passes through will increase due to multiple Coulomb scattering. The rms angle increase from

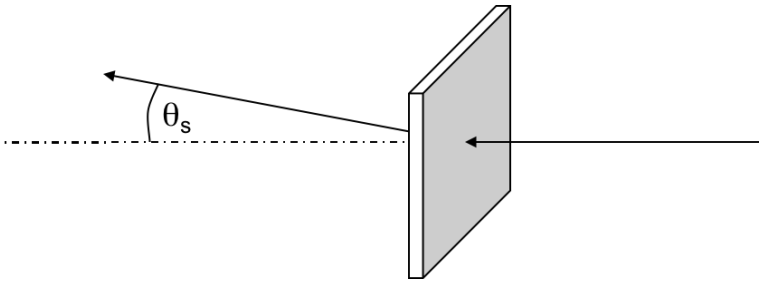


Fig. 22: The emittance of the beam that passes through a thin foil will increase due to multiple Coulomb scattering

multiple Coulomb scattering is

$$\sqrt{\langle\Theta_s^2\rangle}[mrad] = \frac{14.1}{\beta_c p[MeV/c]} Z_{inc}\sqrt{\frac{L}{L_{rad}}}(1+0.11\cdot log_{10}\frac{L}{L_{rad}})\,, \tag{31}$$

where $\beta_c = v/c$, p is the momentum, Z_{inc} is the particle charge, L the target length, and L_{rad} the radiation length of the target. Each particle gets a random change in angle Θ_s, but there is no effect on the positions at the scatterer, Fig. 23. So, the coordinates in normalised phase space at the scatterer location will transform to

$$\begin{aligned} \overline{x}_{new} &= \overline{x}_0 \\ \overline{x}'_{new} &= \overline{x}'_0 + \sqrt{\beta}\Theta_s\,. \end{aligned} \tag{32}$$

And after filamentation the particles will have an emittance $\varepsilon = \langle J_{new}\rangle$ as illustrated in Fig. 24. The action for one of the scattered particles will be

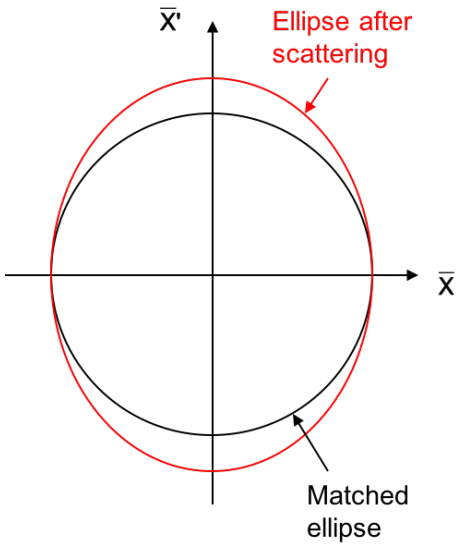


Fig. 23: Each particle gets a random angle change Θ_s, but there is no effect on the positions at the scatterer

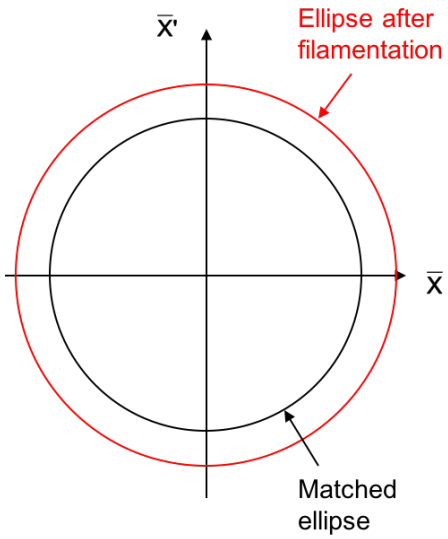


Fig. 24: Filamentation further increases the area in phase space the beam is occupying, after the increase in the angular distribution at the scatterer.

$$\begin{aligned} 2J_{new} &= \overline{x}_{new}^2 + \overline{x}_{new}'^2 \\ &= \overline{x}_0^2 + (\overline{x}_0' + \sqrt{\beta}\Theta_s)^2 \\ &= \overline{x}_0^2 + \overline{x}_0'^2 + 2\sqrt{\beta}(\overline{x}_0'\Theta_s) + \beta\Theta_s^2 \end{aligned} \tag{33}$$

and the mean action therefore, where we use that the initial angle and scattering angle are uncorrelated and the average angle $\langle \overline{x}_0' \rangle = 0$,

$$\begin{aligned} 2\langle J_{new} \rangle &= \langle \overline{x}_0^2 \rangle + \langle \overline{x}_0'^2 \rangle + 2\sqrt{\beta}\langle \overline{x}_0'\Theta_s \rangle + \beta\langle \Theta_s^2 \rangle \\ &= 2\varepsilon_0 + 2\sqrt{\beta}\langle \overline{x}_0' \rangle \langle \Theta_s \rangle + \beta\langle \Theta_s^2 \rangle \\ &= 2\varepsilon_0 + \beta\langle \Theta_s^2 \rangle \,. \end{aligned} \tag{34}$$

The new emittance after filamentation and after having traversed a thin scatterer will be

$$\varepsilon_{new} = \varepsilon_0 + \frac{\beta}{2}\langle\Theta_s^2\rangle\,. \tag{35}$$

In order to reduce the blow-up from stripping foils and other thin scatterers, they have to be placed at locations with small β-function. A small β-function means a large spread in angles, so the additional scattering angle has a small effect on the distribution. An example for the optics design of the insertion in a transfer line with a stripper foil for LHC ion transfer between the PS and SPS accelerator is shown in Fig. 25. For LHC heavy ions, Pb^{54+} is stripped to Pb^{82+} at 4.25 GeV/u using a 0.8 mm thick Al

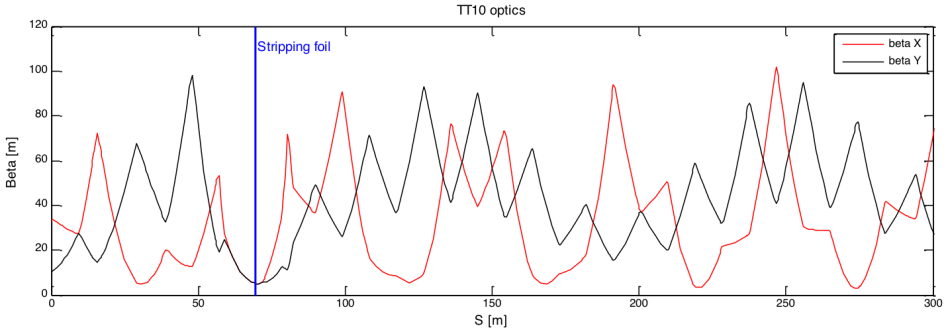


Fig. 25: This figure shows the optics in the transfer line between the accelerators PS and SPS at CERN for the LHC heavy ions. The stripper foil is located at a location with $\beta_{x,y} \approx 5$ m.

foil, in the PS to SPS line. The emittance blow-up is minimised with a low-β insertion ($\beta_{x,y} \approx 5$ m) in the transfer line. Using the formulae from above, the expected emittance blow-up is nevertheless 8%.

3.4 Emittance blow-up from other effects

The techniques that were introduced in the previous sections can be applied to various other effects and errors at injection to calculate the emittance blow-up. In the following results of some of these derivations are summarised.

3.4.1 Dispersion mismatch

In case of a dispersion mismatch by ΔD and $\Delta D'$ and a momentum spread of $\frac{\delta p}{p}$ of the beam, the emittance blow-up will be

$$\frac{\varepsilon}{\varepsilon_0} = 1 + \frac{1}{2}\frac{\Delta D^2 + (\beta\Delta D' + \alpha\Delta D)^2}{\beta\varepsilon_0}(\frac{\delta p}{p})^2\,. \tag{36}$$

3.4.2 Energy mismatch

With dispersion at the injection point and a momentum offset of the beam of $\Delta p/p$, the emittance blow-up will amount to

$$\frac{\varepsilon}{\varepsilon_0} = 1 + \frac{1}{2}\frac{D^2}{\beta\varepsilon_0}(\frac{\Delta p}{p})^2\,. \tag{37}$$

3.4.3 Geometrical mismatch

In case there is a tilt angle Θ between the beam reference systems of the injected beam and the ring at the injection point, cross-plane coupling will occur and through filamentation emittance blow-up. The

following formula gives the result in the horizontal plane [7, 8]

$$\varepsilon_x = \varepsilon_{0x}\cos^2\Theta + \frac{1}{2}\varepsilon_{0y}\sin^2\Theta \cdot \left(\frac{\beta_y}{\beta_x}(1+\alpha_x^2) + \frac{\beta_x}{\beta_y}(1+\alpha_y^2) - 2\alpha_x\alpha_y\right) . \tag{38}$$

3.5 Summary

In this lecture the generic method to calculate emittance growth from mismatch at injection was introduced. The method was applied for the most prominent injection errors and important formulae were derived. At the end of the lecture a collection of formulae for various other typical injection errors was given.

References

[1] A. Wolski, *Beam Dynamics in High Energy Particle Accelerators*, (Imperial College Press, London, 2014), https://doi.org/10.1142/p899.

[2] B. Goddard, Transfer Lines, CAS, 2004.

[3] P. Bryant, in Proceedings of the CAS-CERN Accelerator School: General Accelerator Physics, v.2, Gif-sur-Yvette, France, 3-14 September 1984, edited by P.J. Bryant and S. Turner, CERN-85-19-V-2 (CERN, Geneva, 1985), pp. 358-376, https://doi.org/10.5170/CERN-1985-019-V-2.

[4] L. Voss, Transverse emittance blow-up from double errors in proton machines, CERN, 1998.

[5] C. Bovet et al., A selection of formulae and data useful for the design of A.G. synchrotrons, 1970, CERN.

[6] G. Arduini et al., Mismatch Measurement and Correction Tools for the PS-SPS Transfer of the 26 GeV/c LHC Beam, 1999, CERN.

[7] B. Goddard et al., Expected emittance growth and beam tail repopulation from errors at injection into the LHC, IPAC proceedings, 2005.

[8] K. Fuchsberger et al., Coupling at injection from tilt mismatch between the LHC and its transfer lines, 2009, CERN.

Proceedings of the CAS–CERN Accelerator School: Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018)

Septa

M. Paraliev
Paul Scherrer Institute, Villigen PSI, Switzerland

Abstract
In the Septa I and Septa II lectures we introduce the basic concept of particle accelerator septa devices—key elements of beam injection and extraction systems. The basics of charged particle beam electrostatic and magnetic deflection are presented and a comparison between the two schemes is given. Different septa types and their specifics are described and most are illustrated with real septa designs from accelerator laboratories around the world. Design examples are used to emphasize the important septa parameters and their limitations. Finally, the presented equations for electrostatic and magnetic deflection are mathematically derived.

Keywords
Electrostatic septum; septum magnet; injection; extraction.

1 Introduction

For efficient beam switching, injection, or extraction a special type of beam deflecting device is needed. Such devices are characterized by two distinctive deflecting regions in order to switch, merge, or separate charged particles beams. Typically one region provides (ideally) zero deflection and the other—a significant constant deflection. However, there are designs where both regions provide deflection but in opposite directions. The main objective of the device is to ensure an abrupt deflecting field change between the two regions, with a separating barrier (septum) that is as thin as possible and a minimum leakage field.

The article gives an overview of the conventional septa deflecting schemes and designs. It uses example hardware implementations to discuss some specific design features of the septa.

The essential mathematical expressions that describe electrostatic and magnetic deflections are presented. All quantities are in SI units unless other is stated.

2 Septum

The term *septum* (plural *septa*) means ‘a partition’, ‘a wall’, ‘a barrier’ that separates two cavities or two chambers. It is used in different science and technology fields such as biology, mechanics, particle physics, etc. The word has a Latin origin and it is a derivate of the word *saepio* (*sēpiō*) meaning ‘to surround’, ‘to enclose’, ‘to fence in’ something. In particle accelerators, a septum divides two distinctive field regions in order to selectively deflect charged particle beams to one or other side of it. Often the device that embodies the septum is called a ‘septum’ as well (e.g., electrostatic septum, septum magnet, etc.). Septa, often accompanied by kicker devices, are an essential part of charged particle beam injection and extraction systems.

2.1 Basic concept and terminology

A magnetic septum has a lot in common with dipole (bending) magnets but it is more complex due to the necessity of having an abrupt field change. Figure 1 shows schematically the general layout and basic parameters of a septum. The *bending angle* θ (usually in radians or milliradians) is one of the most

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-005.363

important parameters of a septum. It expresses how much the trajectory of the deflected charged particle beam (hereafter referred to by only '*beam*') diverges from the straight trajectory. The bending angle depends on particle type (rest mass and charge), particle momentum (kinetic energy), deflecting field (type, strength, and direction), and effective length of the septum. *Bending radius R* (usually in metres) is another way of expressing the deflection of the *beam*. *Sagitta, s* (usually in metres or millimetres), gives the largest distance between the arch of the *beam* trajectory and its chord. Together with the *deflecting gap width, w* (usually in metres or millimetres), it is an important parameter for designing a septum large enough to accommodate fully the bent beam, especially if the deflecting gap does not follow the *beam* trajectory. *Septum thickness, t* (usually in millimetres), describes the distance between the aperture walls of the straight and the deflected *beam*. Septum thickness is one of the important parameters of the device and it is often a delicate engineering compromise to meet the controversial mechanical, thermal, electrical, and leakage field requirements of the septum. A thin septum is always an advantage, whether it is for slow (scraping) extraction or for fast (e.g., kicker based) extraction. In the first case, a thin septum reduces *beam* losses and radiation loading. In the second, it reduces the required spatial separation of the two *beams*, relaxing the requirements of the kicker. *Septum length, l* (usually in metres), gives the distance the *beam* travels under the deflecting field. In practice, the length of the field region is different from the mechanical length of the device due to the mechanical construction and fringe fields at the septum entrance and exit. An integral equivalent value, called the *effective length*, is used to describe the proper *beam*−field interaction distance.

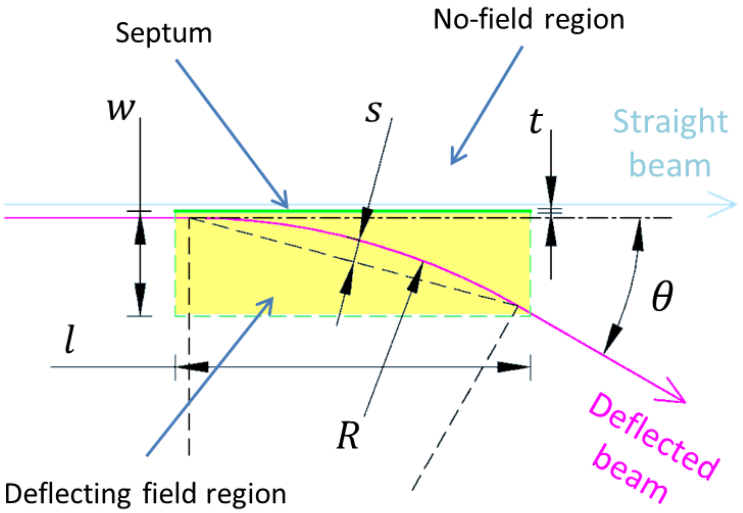


Fig. 1: Schematic representation of a septum

2.2 Key objectives

The key objectives to a septum may be summarized as follows.

(i) *Good field region* – The bending field region should provide an adequate deflecting field (type, strength, and direction) in order to achieve the required *beam* deflection. The field should be constant (homogeneous) not to affect the *beam* optics (although there are some exceptions when the septum has an optics functionality too).

(ii) *Field free region* – In order not to disturb the trajectory of the 'straight' beam, the 'field free region' (or 'zero field region') should have very low residual deflecting field. This is of key importance for circulating *beam* accelerators, where the unwanted trajectory disturbance can accumulate (depending on the machine tune) each time the *beam* passes through the septum.

(iii) *Accommodate the beam* – The aperture of the device should be large enough to accommodate the straight and the bent *beams* taking in account *beam* trajectory and *beam* size (depending on *beam* emittance and the particular beta function at the location of the septum)

(iv) *Beam impedance* – Irregularities in the *beam* chamber introduce variation of the electrical impedance seen by the *beam*. This can create unwanted electric resonances that are excited by the passing *beam*. Especially in circular accelerators these resonances can lead to *beam* instabilities.

(v) *Vacuum* – Vacuum quality is an important parameter for the particle accelerators. When designing septa, care should be taken to avoid materials that are not compatible with a vacuum (emitting or having a large capacity to absorb gases, sublimating etc.) If a high vacuum is required, the design should be bake-out compatible and it should provide enough vacuum 'conductance' to the corresponding vacuum pumps. If vacuum conductance cannot be further improved, a number of special measures can be taken (such as non-evaporable getter (NEG) coating etc.) to reduce locally the background gas pressure.

(vi) *Positioning and mechanical stability* – The correct position of the septum with respect to the *beam* is crucial for its proper and efficient operation. High resolution, remotely controlled mechanical actuators are often used to position septa. Proper mechanical support should be provided to bear the electromagnetic forces. Delicate septa (thin foils or wires) should be attached using special tensioners to cope with the thermal deformation. In pulsed devices, mechanical absorbers may be necessary to limit mechanical shockwave propagation.

(vii) *Synchrotron radiation* – If the charged particle *beam* is deflected it emits electromagnetic waves (synchrotron radiation). This process might affect the properties of the *beam* or it could interact with the vacuum chamber or other devices. Since synchrotron radiation power depends strongly on the deflecting angle, usually it does not have a significant effect in small deflection angle septa.

(viii) *Radiation effects* – Septa are often the most irradiated components of a particle accelerator. In some cases, the excessive radiation can cause material degradation especially for some electrical insulators. Sensitive electronic components could be affected as well if placed in proximity to the septa. Proper shielding should be used to ensure personnel radiation safety outside of the accelerator tunnel during operation. Material activation is another problem that should be taken in account. After being in operation, the septa could be activated and could remain radioactive for extensive periods of time after they are out of service. Remaining radioactive emission could make septa very difficult to service and handle due to the radiation hazards.

(ix) *Thermal management* – Proper thermal management of the septa is crucial for their reliable operation. Excessive heat load could be generated by *beam* irradiation and/or by flowing electrical currents (in electromagnetic septa). Proper cooling should be provided to keep the septa temperature within safe limits.

(x) *Machine and personal protection* – Often septa must meet very demanding requirements and their operating point could be very close to the material damage threshold. In these cases, a proper machine protection system should be provided to prevent damage in case of system failure (fault in power supply, cooling system etc.) or operator error (*beam* miss-steering, incorrect operation conditions etc.). Septa themselves could be part of the machine or personnel protection system, for example disabling the injection in a circular accelerator. The risks associated with high energy densities and radiation effects in septa and the associated equipment should be carefully evaluated and proper measures should be taken to protect the personnel.

(xi) *Reliability/Serviceability/Reparability* – The complexity and the demanding requirements that septa should meet make them prone to failures. Constructors should carefully asses the reliability of the device as well as the procedures for maintenance and repair. Proper safety measures should

be taken to handle and recycle activated components. Depending on the estimated mean time between failure, expected device lifetime, facility importance, and projected facility life span, the adequate number of reserve devices can be foreseen. Resources should be planned for the necessary long-term maintenance.

(xii) *Building costs versus operation costs* – As in any engineering process, care should be taken to optimize construction costs against running costs. Often assessing the running costs is not a straightforward task. Besides the obvious costs for electricity, there are 'hidden' costs such as: for running and maintaining the supporting systems (cooling, controls etc.), consumables (e.g., in cases of high radiation environment, insulating oil, and cables), maintaining qualified supporting personnel, and so on. Proper assessment of the running costs may influence strongly the device design.

(This list is not exhaustive.)

3 Types of septa

Typically, charged particle beams are deflected using electric and/or magnetic fields. However, there are other, more 'exotic' methods to deflect *beams* as well (such as using bent crystals, see Refs. [1−2] etc.) but these are beyond the scope of this overview.

The force $\vec{F}$ exerted on a point charge by electromagnetic fields is given by Eq. (1), which is named after the prominent Dutch physicist Hendrik Lorentz. Some historians suggest that the formula derivation has an even earlier origin and attribute it to Oliver Heaviside or James Maxwell.

$$\vec{F} = q\vec{E} + q\vec{v}\times\vec{B}\,, \tag{1}$$

where q is the electric charge of the particle, $\vec{E}$ is the electric field, $\vec{v}$ is the instantaneous velocity of the particle, and $\vec{B}$ is the magnetic flux density. The first term on the right-hand side of Eq. (1) accounts for the electric field interaction. The exerted electric force depends only on the strength of the electric field and particle charge. The second term accounts for the magnetic interaction, and the force depends on the magnetic field strength as well as on the particle's charge and instantaneous velocity

In accelerators, charged particles often travel with relativistic speeds and the relativistic effects have to be taken in account when dealing with their dynamics.

3.1 Electrostatic deflection

The first option for deflecting *beams* is to use an electric field (electrostatic deflection, also called electric deflection). The force excreted on the particle $\vec{F}_E$ is given by Eq. (2) and it is in the direction shown in Fig. 2.

$$\vec{F}_E = q\vec{E}\,, \tag{2}$$

where q is the electric charge of the particle and $\vec{E}$ is the electric field. The deflecting force is collinear with the electric field. Positive charges are deflected in the direction of the electric field lines, negative charges are deflected in the opposite direction.

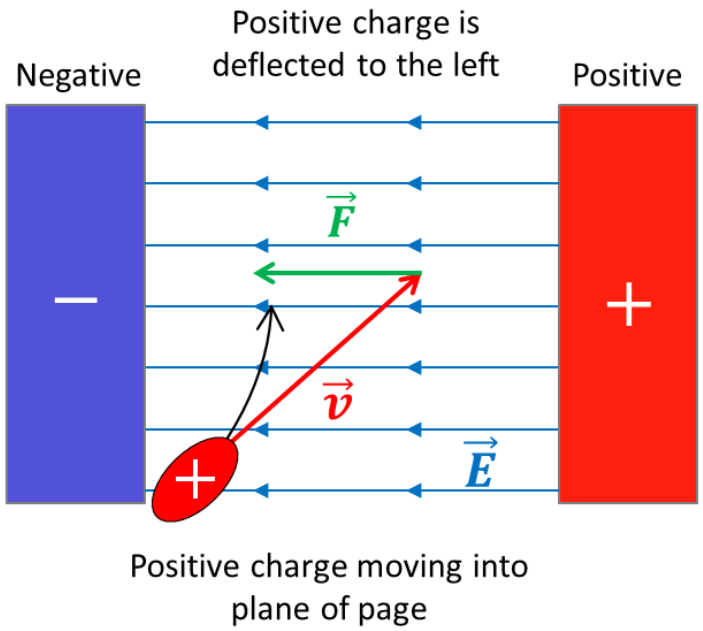


Fig. 2: Electric deflection

Conventions:

- Force on a ***positive*** point charge.
- Electric field lines go from the ***positive*** electrode to the ***negative*** one.
- Opposite electric charges attract each other and ***like*** electric charges repel.

Equation (3) expresses the bending angle θ_E of a particle with elementary charge, where E is the electric field, l_{eff} is the effective length of the device, β is a dimensionless relativistic coefficient that gives the fraction of the speed of light at which the particles travel, p is the *beam* momentum in gigaelectronvolts over the speed of light (GeV/c), U is the deflecting voltage and d is the distance between the deflecting electrodes, see Refs. [3,4].

$$\theta_E \approx \frac{E \cdot l_{\text{eff}}}{10^9 \cdot \beta \cdot p} = \frac{U \cdot l_{\text{eff}}}{10^9 \cdot \beta \cdot p \cdot d} \,. \tag{3}$$

Equation (3) uses *the small angle approximation* ($\tan(\theta) \approx \theta$) which is valid for angles up to ~0.17 rad (~10°) with an error below 1%. Attention should be payed to use of the correct measurement units (especially for *beam* momentum).

Note:

– Momentum units conversion $p[\frac{\text{kg.m}}{\text{s}}] = \frac{q_e}{c} p[\frac{\text{eV}}{\text{c}}]$.

Derivation of the electric deflection equation is included in Appendix A.

3.2 Magnetic deflection

The second option for deflecting *beams* is to use a magnetic field (magnetic deflection). The force exerted on the particle $\vec{F}_M$ is given by Eq. (4) and it is in the direction shown in Fig. 3.

$$\vec{F}_M = q\vec{v}\times\vec{B}\,, \tag{4}$$

where q is the electric charge of the particle, $\vec{v}$ is the instantaneous velocity of the particle, and $\vec{B}$ is the magnetic flux density. The deflecting force is a cross product of $\vec{v}$ and $\vec{B}$ and it is perpendicular to the propagation direction and magnetic field. Positive charges are deflected in the direction defined by the *right-hand rule*; negative charges are deflected in the opposite direction.

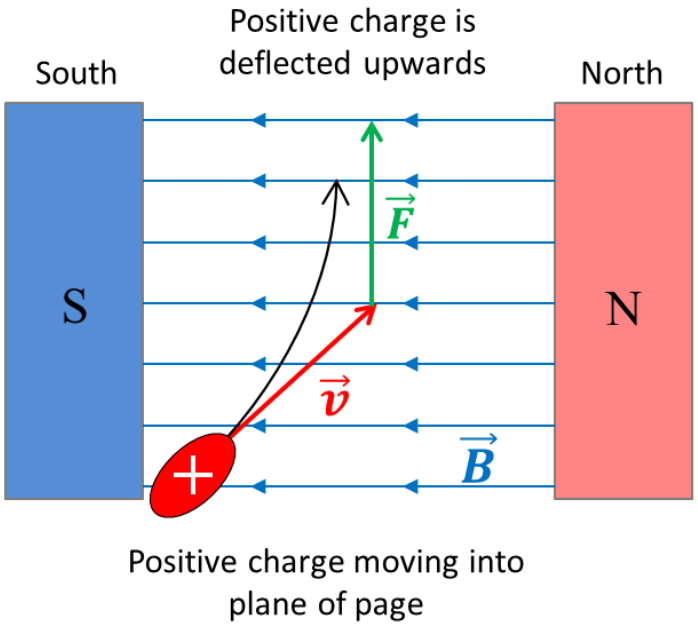


Fig. 3: Magnetic deflection

Conventions:

- Force on a ***positive*** point charge.
- Magnetic ***field*** lines go from the ***North pole*** to the ***South pole*** of the magnet (in the space outside the magnet).
- ***Right-hand rule***: If the thumb points in the direction of motion and the index finger points in the direction of the magnetic field, the magnetic force goes in the direction of the middle finger.

Equation (5) expresses the bending angle θ_M of a particle with elementary charge, where B is the magnetic flux density, l_{eff} is the effective length of the device, p is the *beam* momentum in gigaelectronvolts over the speed of light (GeV/c), n is the number of turns of the magnet winding, I is the deflecting current, and d is the distance between the magnetic poles, see Refs. [3−6].

$$\theta_M \approx \frac{0.3\cdot B\cdot l_{\text{eff}}}{p} \approx \frac{3.76\cdot n\cdot I\cdot l_{\text{eff}}}{10^7\cdot p\cdot d}\,. \tag{5}$$

Equation (5) uses *the small angle approximation* ($\tan(\theta) \approx \theta$) which is valid for angles up to ~0.17 rad (~10°) with an error below 1%. Attention should be payed to use of the correct measurement units (especially for *beam* momentum).

Note:

- Magnetic flux density approximation $B \approx \frac{\mu_0 nI}{d}$.
- Momentum unit conversion $p\left[\frac{\text{kg.m}}{\text{s}}\right] = \frac{q_\text{e}}{\text{c}} p\left[\frac{\text{eV}}{\text{c}}\right]$.

Derivation of the magnetic deflection equation is included in Appendix B.

3.3 Electrostatic versus magnetic deflection

The duality *of electromagnetism* suggests that the two deflection methods should be somehow equivalent. Let's compare the deflection capability of an electric and magnetic field with the same energetic characteristics. Equations (6) and (7) give the energy density, W_E and W_M, of an electric and magnetic field in free space, see Ref. [7]:

$$W_E = \frac{\varepsilon_0 E^2}{2}, \tag{6}$$

$$W_M = \frac{B^2}{2\mu_0}, \tag{7}$$

where ε_0 is the permittivity of free space, E is the electric field, B is the magnetic flux density, and μ_0 is the permeability of free space. The relation between an electric and magnetic field with the same energy density could be found by equating Eqs. (6) and (7):

$$E = \frac{1}{\sqrt{\mu_0 \varepsilon_0}} B \quad \text{or} \quad E = cB, \tag{8}$$

where c is the speed of light in free space. Using the scalar form of Eqs. (2) and (4), we can see that electric deflecting force F_E and magnetic force F_M for fields with the same energy density depends only on the charged particle instantaneous velocity v and it is *the same for relativistic particles ($v = c$).*

$$F_E = qE = qcB, \tag{9}$$

$$F_M = qvB. \tag{10}$$

Note:

- Speed of light in free space is equal to 299 792 458 m/s *exactly*!
- On the other hand, μ_0 and ε_0 are *irrational numbers*.

Since most particle accelerators deal with relativistic *beams* and since, practically, producing a magnetic field is *much easier* than producing an electric one with the same energy density, the majority of accelerator' deflecting devices are magnetic. Equation (11) expresses the equivalence of electric and magnetic field deflection, where the electric field E is in gigavolts per metre (GV/m), B is the magnetic flux density, and β is the dimensionless relativistic coefficient that gives the fraction of the speed of light at which the particles travel.

$$\frac{3.3}{\beta} E = B. \tag{11}$$

Table 1 compares the deflecting capability of an electric and magnetic field for electrons and protons with different momenta. An electric field of 10 MV/m is taken as a comparison benchmark since it is widely accepted as the practical maximum electric field limit.

Table 1: Electric and magnetic deflection for different particle momenta

β	γ	$p_{\text{electrons}}$ (MeV/c)	p_{protons} (GeV/c)	Electric field (MV/m)	Equivalent magnetic field (T)	
0.001	1.000	0.0005	0.0009	10.00	33.356	Non-relativistic
0.01	1.000	0.0051	0.0094	10.00	3.336	
0.1	1.005	0.0514	0.0944	10.00	0.334	
0.3	1.048	0.1607	0.2955	10.00	0.111	
0.5	1.155	0.2950	0.5425	10.00	0.067	
0.9	2.294	1.0552	1.9401	10.00	0.037	Relativistic
0.99	7.089	3.5864	6.5944	10.00	0.034	
0.999	22.366	11.4185	20.9955	10.00	0.033	
0.9999	70.712	36.1328	66.4386	10.00	0.033	
0.99999	223.607	114.2698	210.1114	10.00	0.033	
0.999999	707.107	361.3552	664.4349	10.00	0.033	

Electric field deflection has an advantage for low β *beams* (low energy, heavy particles) as well as in some special cases when there are geometrical advantages. Table 2 summarizes the basic pros and cons of the two septa types.

Table 2: Advantages and disadvantages of electrostatic and magnetic septa

Electrostatic septum	Magnetic septum
+ Nearly perfect 'zero-field' region	+ Strong deflection
+ Thin septum	+ More effective for relativistic beams
+ Low mass density (low beam interaction)	+ In-vacuum and in-air design is possible
+ Better for non-relativistic beams	– Thick septum
– Difficult to have high fields	– Field leakage
– Less effective for relativistic beams	– Non-uniform field region
– High voltages handling	– Interaction with other magnets
– Strictly in-vacuum design	– High currents handling

4 Electrostatic septum

As illustrated schematically in Fig. 4, an electric field is established between a high voltage (HV) electrode and a septum foil. The injected or extracted *beam* passes through the electric field region and is deflected. Using electrostatic screening (*Faraday cage effect*) a zero-field region is created for the (circulating) *beam* that goes straight. Due to the practically perfect electric field screening from even very thin conductor foils, the straight *beam* passes undisturbed. Due to the high electric fields electrostatic septa designs are exclusively 'in-vacuum'.

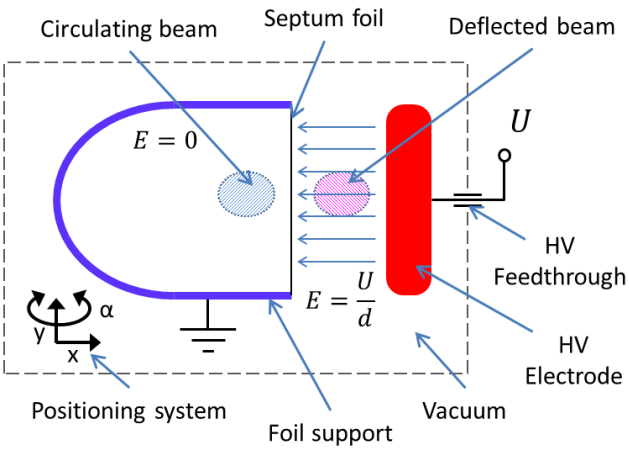


Fig. 4: Schematic view of electrostatic septum

A thin foil or wire array is used to minimize the interaction with the *beam* (reduces *beam* losses and radiation levels). To align the septum with the *beam*, often the septum and its support are attached to a remotely controlled, precision positioning system. Since the zero-field region is a relatively confined volume, some holes away from the HV electrode are necessary to ensure good vacuum conductivity and to maintain a low background gas pressure.

4.1 Foil septum

Foil septa benefit from the very thin septum conductor (the foil) used to separate the field and zero-field region. Due to its small cross-section the foil (foil thickness down to 100 μm) interacts less with the *beam* and reduces losses and radiation. Figure 5 shows an example of a foil septum, 'Septum 23', used in the proton synchrotron ring at CERN for resonant slow extraction (see Refs. [8−9]). Table 3 summarizes its technical parameters.

Table 3: Technical specification of 'Septum 23'

Parameter	Value
Electrode length, mm	778
Gap width, mm	17
Beam momentum, GeV/c	24
Deflection angle, mrad	0.28
Septum thickness, μm	100
Vacuum pressure, mbar	10^{-9}
Voltage, kV	260
Electric field, MV/m	up to 15
Septum foil material	Molybdenum
Electrode material	Anodized Peraluman 300
In situ bake-able	Yes

'Septum 23' (Fig. 5) uses polished 100 μm molybdenum foil (as septum) and an anodized aluminium alloy HV electrode that enables an electric field up to 15 MV/m after conditioning. Polished stainless steel deflectors confine the electric field and help to extend the 'good field' region to 40 mm. They also protect HV active parts from titanium deposition (from sublimation vacuum pumps). The septum is equipped with an in-vacuum infrared lamp in order to allow in situ bake-out. The foil is

attached using tensioners that reduce the mechanical stress in the foil due to different thermal loadings in different operating modes (bake-out or *beam* thermal loading), see Ref. [8]. A series of holes in the foil support provide good vacuum conductivity between the 'zero-field' region and the rest of the chamber. The septum and the HV electrode are equipped with a precise positioning system (translation resolution 100 μm) in order to align well with the *beam*. The septum can also be moved away from the *beam* in case it is not used to minimize *beam* losses. Due to radiation degradation effects, the insulating oil in the HV feedthroughs is regularly changed (every 3 months) and the close HV cables are replaced annually, see Ref. [9]. The maintenance of the septum is difficult because it is radioactively 'activated'.

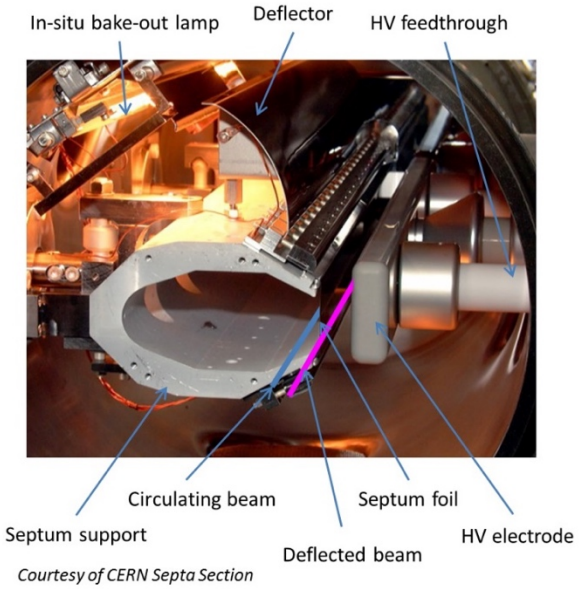


Fig. 5: 'Septum 23' proton synchrotron ring at CERN

4.2 Foil septum with inverted design

Another variation of the foil septum has a so-called 'inverted' design. In this case, the field is confined inside the foil holder, as shown in Fig. 6. This provides better control over the electric field distribution and enables one-sided mechanical support, leaving space for additional in-vacuum components in the 'zero-field' region, but makes the construction of the HV feedthrough more complex. A typical example (Fig. 6) is the injection electrostatic septum (ES) designed for the National Centre of Oncological Hadrontherapy (CNAO) in Italy, see Refs. [10,11]. Table 4 summarizes its technical data.

Table 4: Technical specification of injection SE for CNAO

Parameter	**Value**
Electrode length, mm	800
Gap width, mm	25
Beam momentum, MeV/*c*	20
Deflection angle, mrad	60
Septum thickness, μm	100
Vacuum pressure, mbar	10^{-9}
Voltage, kV	69
Electric field, MV/m	2.8
Septum foil material	Molybdenum
Electrode material	Stainless steel

This particular design comprises two straight sections to follow the trajectory of the deflected *beam*. The whole septum support is attached to a positioning system with ±5 mm range.

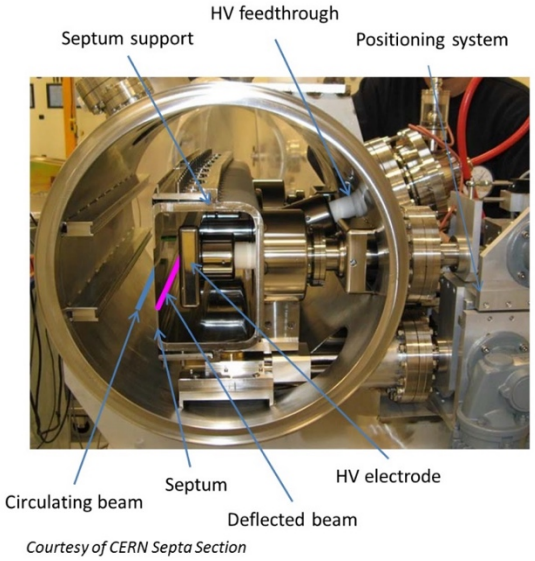


Fig. 6: Foil septum with 'inverted' design, CNAO injection ES

4.3 Diagonal foil septum

A diagonal foil electrostatic septum (ER.SEH10, Ref. [12]) is used in the low energy ion ring (LEIR) at CERN. It allows for a combined longitudinal and transverse injection scheme. The foil septum is tilted by 30° to provide the necessary diagonal electric field. The foil and the HV electrode are mechanically polished to improve their electrical breakdown performance. The septum−HV electrode system is conditioned to a higher than nominal voltage in order to ensure reliable operation. A carbon resistor is used to decouple the HV electrode from the supply cable to limit the discharge energy in case of breakdown between the HV electrode and septum.

Table 5: Specification of injection electrostatic septum ER.SEH10 for LEIR

Parameter	Value
Electrode length, mm	720
Gap width, mm	40
Beam momentum, MeV/nucleon	4.2
Deflection angle, mrad	28.9
Septum thickness, μm	100
Vacuum pressure, mbar	10^{-12}
Voltage, kV	51
Electric field, MV/m	1.12
Septum foil material	Molybdenum
Electrode material	Titanium

Special mechanical assembly is used to keep the septum foil in tension and to compensate for the different thermal expansions of the foil and the stainless steel support. Even after a 300°C bake-out, the

alignment of the foil remains within 100 µm. Both septum and cathode have independent alignment systems. Figure 7 shows the design of the septum.

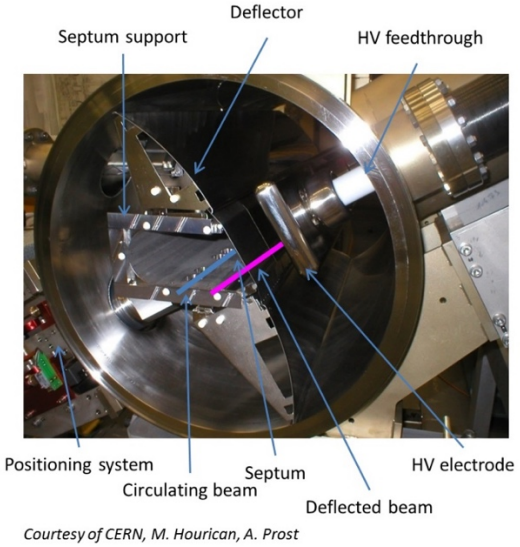


Fig. 7: Design of LEIR injection electrostatic septum ER.SEH10

4.4 Wire septum

In wire septa, the foil is replaced by an array of thin metallic wires. There are several advantages to using a wire array as a septum: Due to the gap between the wires the effective density of the septum is reduced. This leads to less interaction with the *beam* and lower *beam* losses and radiation. Often the wires have individual tensioners that in case of a failure take the broken wire away from the *beam*, keeping the septum operational. Wire septa allow better vacuum conductance for the zero-field region. There are some drawbacks associated with the wire septa as well. The wire array is not a continuous field barrier and allows some field penetration through the septum. This depends on the wire diameter d and wire spacing D. Let's define a wire density coefficient k as the ratio between the wire diameter d and the distance between the wires D. Figure 8 summarizes the results of 3D electrostatic simulation of the leakage field at a distance of 2× D for different values of the wire density coefficient k.

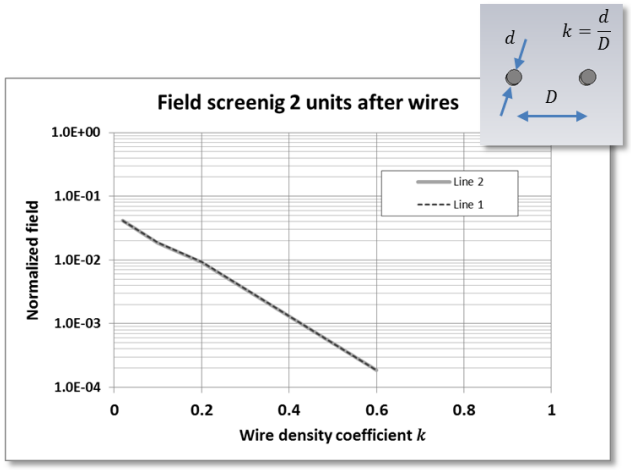


Fig. 8: Field penetration through wire septum at distance 2× D after the septum as a function of the wire density coefficient, *k*.

Even not very dense wire arrays ($k = 0.1$) provide ~53 times field reduction at a distance of $2\times D$. Even though this field leakage is small it could move some ionized background gas molecules from the circulating *beam* (zero-field) area to the field gap. They will then be accelerated by the main septum field and they will be smashed into the HV electrode, increasing the chance of electrical breakdown. Special additional electrodes (ion traps) can be used to counter the leakage field and to clear the ions from the zero-filed area of the septum.

A typical example of a wire septum is the ZS electrostatic septum in the super proton synchrotron (SPS) at CERN, allowing for both fast and slow extraction (see Ref. [13]). To obtain the required deflection, 5 identical wire septa are used. Table 6 gives the technical parameters of each of these septa and Fig. 9 shows their construction.
ZS septa are prone to electrical breakdown due to background gas ionization and they are fitted with upper and lower ion trap electrodes. A spark protection system (Ref. [14]) deactivates the devices if breakdowns are too frequently detected. The septa are equipped with a remote controlled mover system in order to be aligned with the *beam*. If not used, they can be completely retracted. Due to the different possible modes of operation, measures have to be taken to protect the ZS wire septum from *beam* induced overheating, caused by machine failure or operator error (Ref. [15]).

Table 6: Specification of the individual septa unit part of ZS electrostatic SPS extraction septum

Parameter	**Value**
Electrode length, mm	2997
Wire material	W74Re26
Wire diameter, μm	50 to 100
Wire spacing, mm	1.5
Gap width, mm	20
Beam momentum, GeV/*c*	450
Deflection angle, μrad	~70
Vacuum pressure, mbar	10^{-12}
Voltage, kV	220
Ion traps voltage, kV	3 to 6.5
Electric field, MV/m	up to 11
Electrode material	Anodized aluminium

There are even more challenging proposed electrostatic septum designs using extremely thin, low atomic number material wire arrays, curved under the force of an electric field, such as the mini-wire-septum proposed by H. Schönauer that can provide an additional collimation function (see Refs. [16,17].

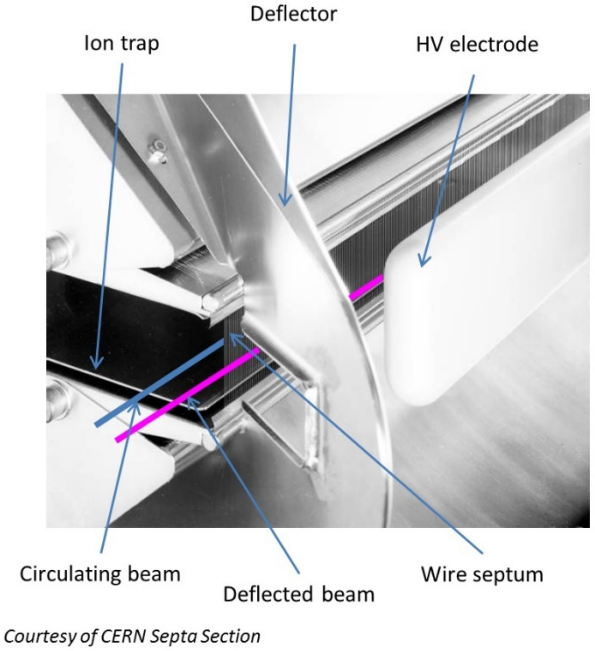


Fig. 9: Wire septum ZS, SPS, CERN

4.5 Practical considerations

(i) *Very thin septa* – Small septa thicknesses (or diameter) require a stress relief mechanism (tensioners) to cope with the deformation due to thermal stress.

(ii) *Low atomic number material* – Low atomic number septum materials are welcomed because they interact less with the *beam* and consequently generate less radiation.

(iii) *High electric field* – Special surface treatment of the septum and HV electrode followed by conditioning should be considered to operate reliably at high electric field without breakdown.

(iv) *Good field region* – Homogeneity of the field in the 'good field' region could be improved using a suitable gap geometry and deflectors.

(v) *Field leakage compensation* – Wire septa may need leakage field compensation.

(vi) *Beam impedance* – Measures should be taken to keep the *beam* impedance changes small, especially in circular accelerators, to avoid *beam* instabilities.

(vii) *HV handling* – Mechanical components that serve as electrical isolators should have large enough creepage distances to avoid surface discharges. Triple points (metal−vacuum−isolator) should be carefully designed to cope with the enhanced electric field in these regions.

(viii) *Insulators degradation* – Materials should be carefully chosen to withstand radiation. Organizational measures should be taken to regularly replace the vulnerable materials (oil, cables).

(ix) *Alignment* – A proper positioning system should ensure good septum alignment.

(x) *Good vacuum* – The design should meet the particular vacuum requirements, including bake-out capabilities.

(xi) *Background gas ionization* – Measures should be taken to minimize the amount of ionized background gas molecules (good vacuum, ion traps) that could initiate a breakdown.

(xii) *Machine protection* – Proper machine protection should be used to protect delicate septa.

(xiii) *Activation* – Maintenance should be done according to the radiation safety regulations.

(This list is not exhaustive.)

5 Magnetic septum

In a typical magnetic septum, illustrated schematically in Fig. 10, the deflected *beam* goes through a homogeneous magnetic field that is established between two magnetic poles. The straight (circulating) *beam* passes next to main magnetic circuit 'seeing' as little as possible the magnetic field. Often magnetic screening techniques are used to shield the straight *beam* further.

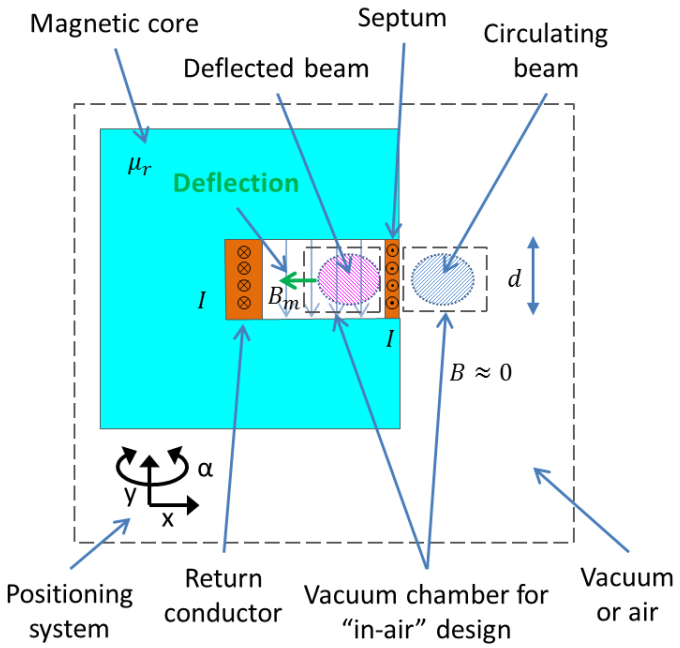


Fig. 10: Schematic view of magnetic septum

Magnetic septa are classified according to the magnetic field variation in time, as shown in Fig. 11. Essentially, they can have an 'in-vacuum' or 'in-air' design.

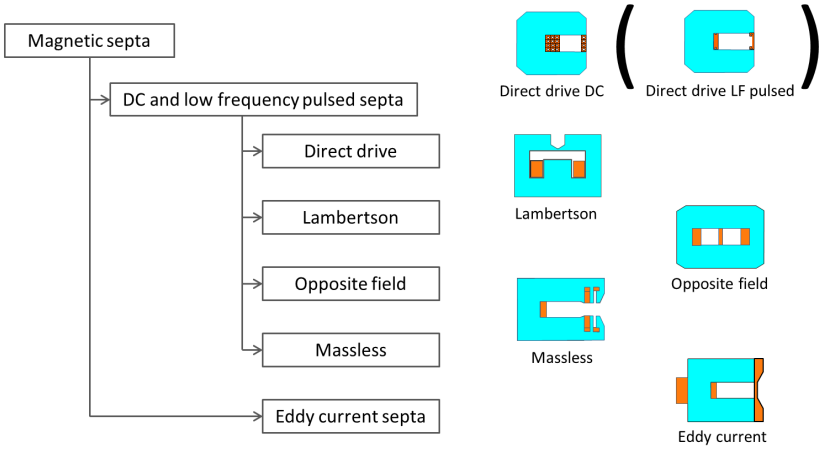


Fig. 11: Magnetic septa classification and schematic symbols

5.1 Direct drive DC septum

The direct drive DC septum (or 'C' type active septum magnet) is a type of window frame magnet with one of the legs removed. The septum is used as one of the magnet's conductors and the return conductor is inside the magnetic core, see Ref. [18]. Field quality in the field region is good due to the magnet geometry. The septum is relatively thick in order to handle high currents. In circular machines the DC operation can push the acceptable leakage field limits further down because the circulating *beam* will be disturbed continuously each time it passes the septum. A static magnetic simulation example (Fig. 12) is used to illustrate the field distribution along line A.

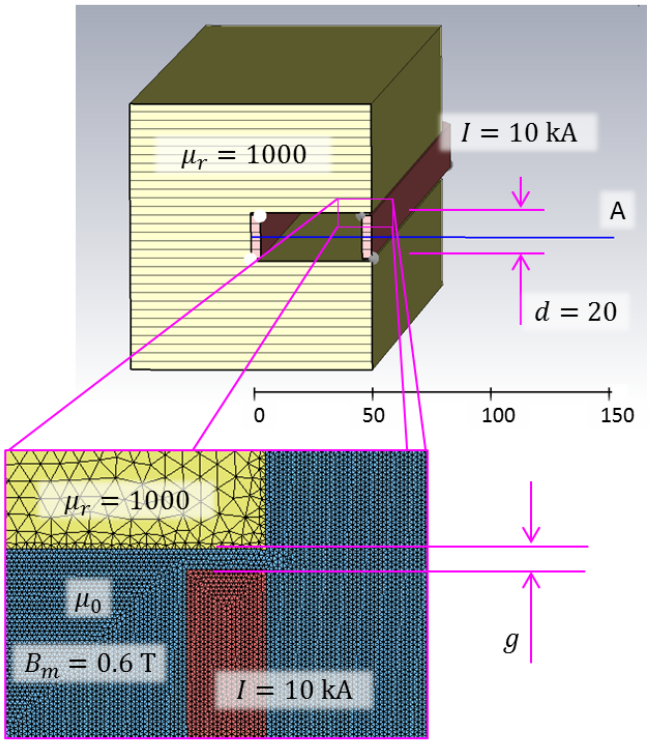


Fig. 12: Simplified static model of the magnet and evaluation of the leakage field as a function of the gap between the septum conductor and the magnetic core.

We will use a set of example parameters to illustrate some of the practical difficulties that arise in the septum magnet design process. For simplicity, these will be used for several different types of septa and that is why the values are not necessarily practical (acceptable) for all discussed septa types.

Example parameters*:

- Gap height: 20 mm
- Gap width: 40 mm
- Relative permeability of the magnetic core: 1000
- Septum thickness: 4 mm
- Septum current: 10 kA
- Septum field: 0.6 T.

*Caution! The chosen parameters are **only** for illustrative purposes and may **not** be practical for all discussed septa types.

Figure 13 shows the normalized absolute value of the magnetic flux density along line A. The main field homogeneity does not change significantly with a change in the gap g but the leakage field is strongly affected. The leakage field is relatively large (~0.5% of the main field) even when the gap g is zero. Additional measures could be necessary (magnetic screening, more complex septum shape) to bring the leakage field within acceptable limits.

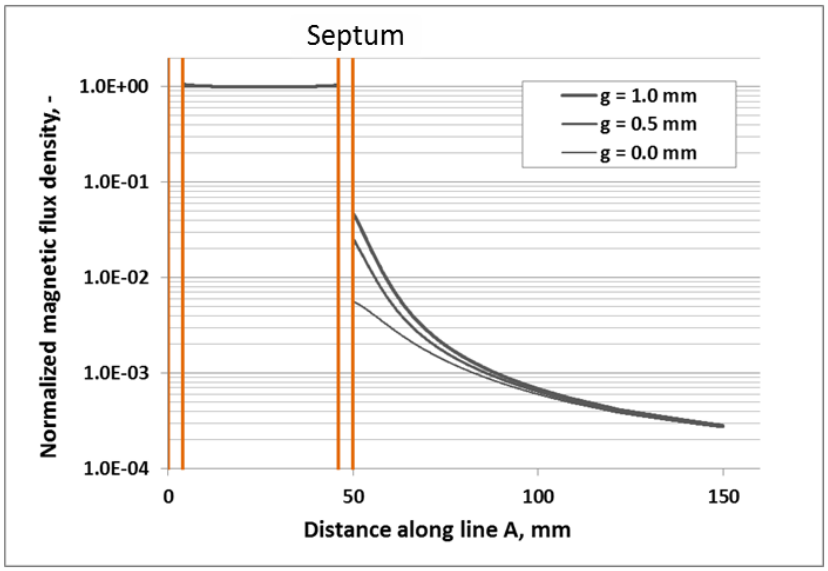


Fig. 13: Normalized absolute value of magnetic flux density along line A

For the given example:

- Septum current density: 125 A/mm^2
- Dissipated power in the septum conductor: 21 kW/m.

Keeping the septum conductor thin results in a reduced cross-section and increases its current density J expressed in A/mm^2 (Eq. (12)):

$$J = \frac{I}{A_s}, \tag{12}$$

where I is the septum current and A_s is the septum cross-section (in mm^2). Due to the high current density, direct drive DC septa have a significant thermal loading. The dissipated power of a copper septum per metre P_{Cu} is given below:

$$P_{Cu} \approx 1.68\times10^{-2}\,\frac{I^2}{A_s}. \tag{13}$$

Resistance of metallic conductors depends on their temperature. Equation (14) gives the dependency of the resistance R for non-cryogenic temperatures.

$$R = R_0[1 + \alpha(T - T_0)], \tag{14}$$

where R_0 is the initial resistance, α is the dimensionless coefficient, which depends on the particular conductor metal, T_0 is the initial conductor temperature, and T is the actual conductor temperature. Since the resistance is proportional to the conductor temperature, in constant current systems a thermal runaway could occur.

Note:

– Copper wire operating 50°C above its initial temperature has 20% more resistance.

To avoid overheating, adequate cooling should be provided. Often a fluid is pumped through channels in the conductor to remove the generated heat. Equation (15) gives the removed power P_{r} as a function of the cooling fluid parameters.

$$P_{\mathrm{r}} = \dot{m} c_p (T_{\mathrm{Out}} - T_{\mathrm{In}}) \,, \tag{15}$$

where $\dot{m}$ is the fluid mass flow rate, c_p is the specific heat capacity of the cooling fluid, T_{Out} is the fluid output temperature, and T_{In} is the fluid input temperature. Depending on its speed, the fluid flow could be laminar or turbulent. Turbulent flow removes heat much more efficiently but too high of a flow rate could cause vibrations and erosion. The Reynolds number R_e characterizes the fluid flow. If $R_e < 2000$ the flow is laminar and if $R_e > 4000$ the flow is turbulent.

$$R_e = \frac{\rho v D}{\mu} \,, \tag{16}$$

where ρ is the fluid density, v is the mean fluid velocity, D is the internal pipe diameter, and μ is the fluid dynamic viscosity. Another important parameter of the cooling system is the pressure drop over the cooled element and connecting pipes. Equation (17) gives the pressure drop Δp over a pipe with internal diameter $D$and Darcy friction factor f_{D}:

$$\Delta p = f_{\mathrm{D}} \frac{\rho v^2}{2D} \,. \tag{17}$$

Septa should be properly supported to withstand the electromagnetic force generated by the flowing electrical current. Equation (18) gives the electromagnetic force per metre of the septum.

$$F = \frac{BI}{2} \,, \tag{18}$$

where B is the magnetic flux density and I is the septum current.

For the given example:

– Required cooling water flow rate (‘in’ to ‘out’ temperature difference 40°C) per metre is 0.13 kg/s.m or 7.5 l/min.m
– For 4 cooling channels with diameter 3 mm and surface roughness 10 μm, the Reynolds number is 28 500 (definitively turbulent flow)
– Water pressure drop per metre is 1 bar
– The septum will be pushed out of the magnet’s gap with a force per metre as high as 3 kN/m or 320 kg/m.

Insulation of the conductors in septa is another important factor for reliable operation. Septa are often the places with the highest radiation levels across the entire accelerator. Organic insulating materials (in the magnet and in its proximity) are vulnerable to high levels of radiation. Radiation resistant insulating materials should be used wherever possible (see Ref. [19]). Depending on the radiation levels, vulnerable components (insulating oil, nearby cables) should be exchanged in a timely manner to avoid failures.

An example of direct drive DC septum ‘ISEP2’ in the rapid cycling synchrotron (RCS) ring at the Japan Proton Accelerator Research Complex (J-PARC) is shown in Fig. 14. Table 7 gives its basic design parameters (see Refs. [20−22]).

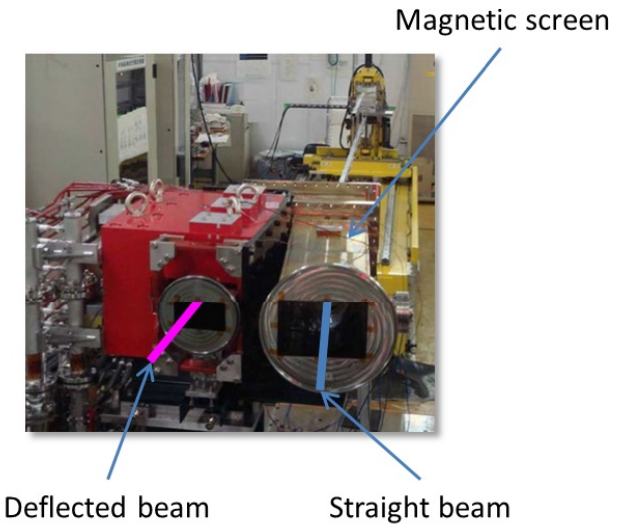


Fig. 14: Direct drive DC septum ISEP2 in RCS ring (J-PARC)

Table 7: Specification of direct drive DC septum ISEP2 in RCS ring (J-PARC)

Parameter	Value
Device design	"In-air"
Field length, mm	650
Gap height, mm	140
Gap width, mm	348
Beam momentum, MeV/*c*	181
Deflection angle, mrad	90.8
Septum thickness, mm	45
Current, kA	6
Magnetic flux density, T	0.475

5.2 Direct drive low frequency (LF) pulsed septum

The direct drive LF pulsed septum is very similar to the direct drive DC septum, excepting that the current does not flow all the time but it is pulsed. Pulsed operation reduces the overall consumption of the device and reduces the stored *beam* disturbance in circular accelerators. Moreover, it reduces the average dissipation in the septum and makes it possible to increase the septum current, and to operate at higher magnetic fields with reduced septum thickness. The ratio of the time period t when the septum is powered and the repetition rate period T is called the duty cycle δ:

$$\delta = \frac{t}{T}. \tag{19}$$

The root-mean-square current of a rectangular pulse is $\sqrt{\frac{1}{\delta}}$ times smaller than the maximum pulse current. For a half-sine pulse this coefficient is $\sqrt{\frac{2}{\delta}}$. Short pulses deposit the heat virtually instantaneously, creating large thermal cycling (see Ref. [23]).

Large currents produce large electromagnetic forces and if they are pulsed, strong dynamic mechanical stresses (known as 'hammering') are generated. Adequate mechanical support is needed to prevent septum displacement. If not properly designed, constantly vibrating conductors could 'cut

through' their isolating supports, leading to catastrophic failure. Some designs include dedicated shock absorbers to limit the mechanical wave transfer to the rest of the device.

Note:

- For short pulses there is no effective heat transfer and the case should be treated as if the pulses deposit the heat instantaneously.
- For a septum operating at 1 T with a 30 kA pulse, the maximum force per metre could be as large as 15 kN/m or 1500 kg/m.

A step-down transformer can be used to provide the required high current. Typically the transformation ratio is in the range 4:1 to 50:1. In septa having a sine-wave current pulse, a third harmonic circuit can be used to improve the flat top of the pulse. The fundamental and third harmonic current add together to widen the flat top. In some amplitude-sensitive applications, an active electrical regulation circuit may improve further the stability of the flat-top current (see Ref. [3]).

Figure 15 shows a typical example of an 'in-vacuum' direct drive LF pulsed septum PESMH16 used in the proton synchrotron (PS) at CERN. Its basic parameters are listed in Table 8 (see Refs. [24, 25]).

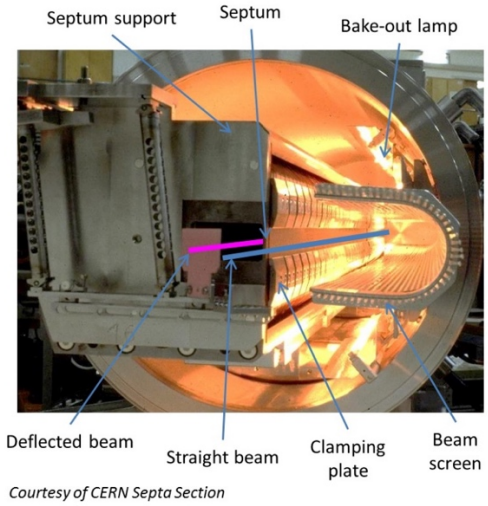


Fig. 15: Direct drive LF pulsed septum PESMH16 (PS, CERN)

Table 8: Specification of direct drive LF pulsed septum PESMH16 (PS, CERN)

Parameter	**Value**
Device design	'In-vacuum'
Vacuum, mbar	10^{-9}
Field length, mm	2180
Gap height, mm	30
Gap width, mm	65
Beam momentum, GeV/*c*	25.1
Deflection angle, mrad	30
Septum thickness, mm	3
Current, kA	28.5 (half-sine 3.5 ms)
Magnetic flux density, T	1.2

5.3 In-air Lambertson septum

Due to magnetic circuit symmetry, the zero-field region of the Lambertson septum has a very low leakage field and disturbs very little the stored *beam* in circular accelerators. This design allows for a thin septum. The device could be DC or LF pulsed. Lambertson septa have a more complex geometry and deflect the *beam perpendicular to the initial displacement*. Figure 16 shows a typical example of an 'in-air' DC Lambertson septum MSIA in the large hadron collider (LHC) at CERN (see Refs. [26,27]) and Table 9 summarizes its basic parameters. This particular design is characterized by two zero-field channels (for circulating *beam* and for counter rotating *beam*). Mu metal chambers are used for additional magnetic screening. A non-evaporable getter (NEG) coating is deposited on the inner surface of the chamber to improve the local vacuum. The top yoke side (with zero-field channels) extends 175 mm on each side of the device to screen the fringe magnetic fields.

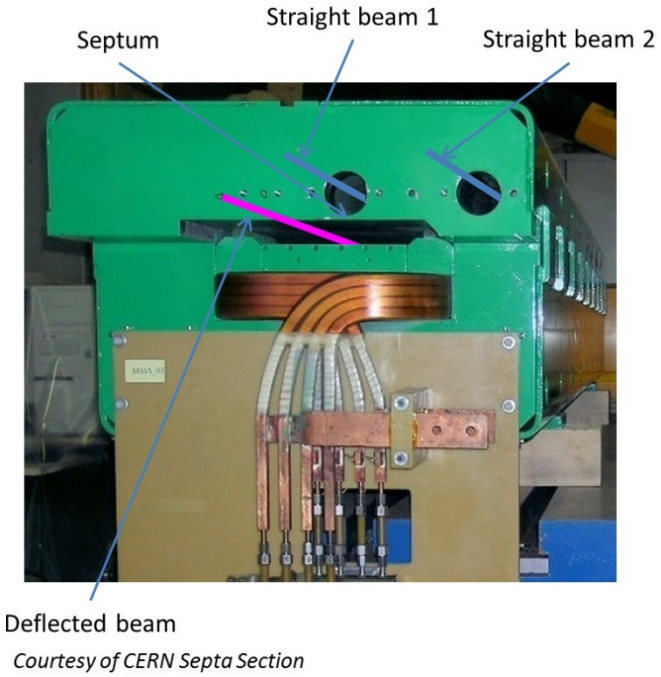


Fig. 16: Lambertson DC septum MSIA in LHC at CERN

Table 9: Specification of Lambertson DC septum MSIA (LHC, CERN)

Parameter	**Value**
Device design	'In-air'
Vacuum, mbar	10^{-7}
Field length, mm	3650
Gap height, mm	25
Gap width, mm	230
Beam momentum, GeV/*c*	450
Deflection angle, mrad	1.846
Septum thickness, mm	6
Current, kA	0.95 × 16 turns
Magnetic flux density, T	0.76

5.4 Half in-vacuum Lambertson septum

Another example of a DC Lambertson septum is the 'half-in-vacuum' septum for SwissFEL at the Paul Scherrer Institute (PSI). It benefits from a relative low vacuum volume and a very thin septum wall. The coil is on the air side, which makes it easily serviceable. The massive additional shorted electrical turn reduces HF current ripples from the power supply. The bottom yoke side (with zero-field channel) extends 150 mm on each side to screen the fringe magnetic fields. Figure 17 shows the 'half-in-vacuum' DC Lambertson septum AFS for SwissFEL switchyard and Table 10 summarizes its basic parameters (see Ref. [28]).

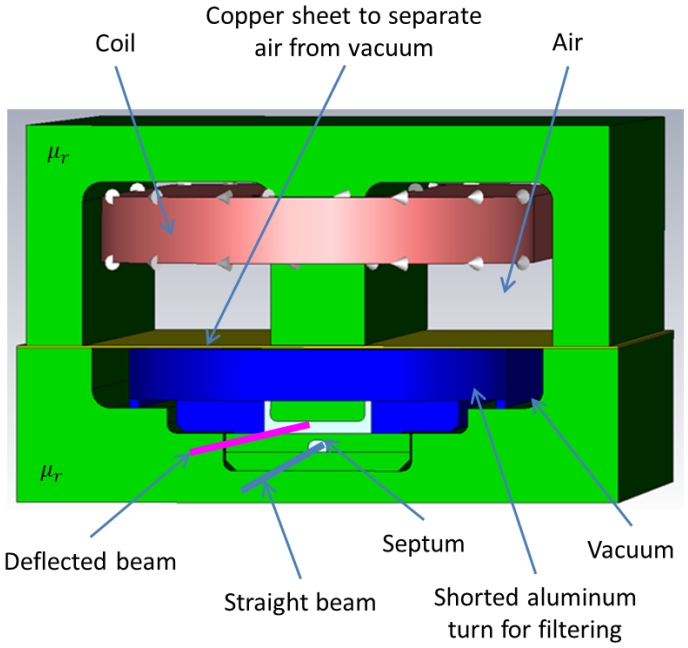


Fig. 17: Lambertson DC septum MSIA in LHC at CERN

Table 10: Specification of Lambertson DC septum MSIA (LHC, CERN)

Parameter	Value
Device design	'Half-in-vacuum'
Vacuum, mbar	10^{-7}
Field length, mm	760
Gap height, mm	6.8
Gap width, mm	61
Beam momentum, GeV/*c*	3.15
Deflection angle, mrad	35
Septum thickness, mm	2.5
Current, kA	0.1 × 41 turns
Magnetic flux density, T	0.51

5.5 Opposite field septum

Instead of a zero-field region the opposite field septa have a region with a magnetic field in the opposite direction (see Ref. [29]). In this way the electromagnetic forces at the septum cancel out and the mechanical aspects of septum support are relaxed. Both *beams* are deflected but in the opposite direction. The design could be DC or LF pulsed. An example of an opposite field 'in-air' design, LF

pulsed septum is the injection septum (Fig. 18) used at the Japan Proton Accelerator Research Complex (J-PARC) at The High Energy Accelerator Research Organization known as KEK. Table 11 gives its basic parameters (see Ref. [30]).

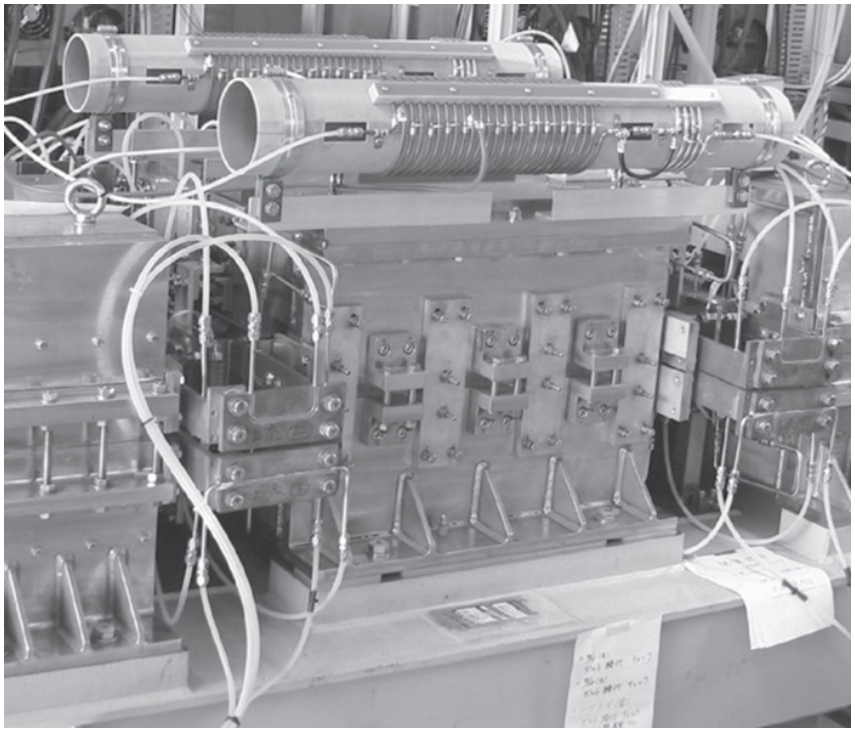


Fig. 18: Opposite field septum at J-PARK (KEK)

Table 11: Specification of opposite field LF pulsed septum at J-PARK (KEK)

Parameter	Value
Device design	'In-air'
Field length, mm	700
Gap height, mm	120
Gap width, mm	150/400
Beam momentum, GeV/*c*	3
Deflection angle, mrad	68
Septum thickness, mm	8
Current, kA	48 ×2 (half-sine 2.5 ms)
Magnetic flux density, T	0.6

5.6 Massless septum

A 'massless septum' idea is to create two distinctive magnetic field regions without using a physical barrier (physical septum) between them. A complex magnetic circuit and/or several excitation coils have to create the two field regions (Refs. [31,32]). An illustration of this approach is given by the design proposed at the Nuclear Science Research Facility (NSRF) at Kyoto University (Ref. [32]). The design can be DC or pulsed. The absence of a physical septum means there is no interaction with the *beam* and no *beam* losses and no radiation. The drawback of the design is the gradual transition between the two field regions, making the effective septum very thick. If the *beam* crosses through this transition region (with a field gradient) the accelerator optics can be compromised. As a 'rule of thumb' the gradient region is approximately the same as the magnetic gap height (or the distance between the closest current

conductors). A cross-section of a static magnetic simulation model and the field profile along line A is shown in Fig. 19.

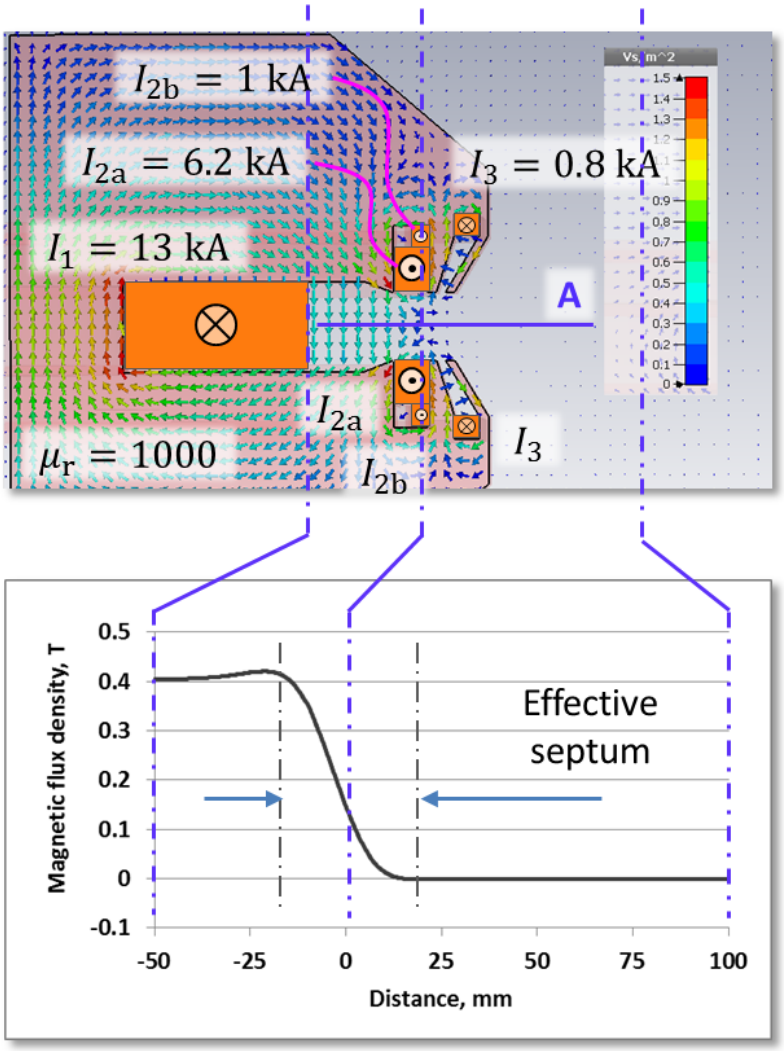


Fig. 19: Static magnetic simulation model of the proposed massless septum design and magnetic field profile along line A.

5.7 Eddy current septum

If the magnetic field changes inside of a solid conductor, (eddy) currents will be induced to oppose the magnetic field change. This effect is utilized to design very thin septa devices. The magnetic field in the field region is established relatively quickly and due to the eddy currents in the septum conductor the field does not penetrate through it. The septum acts as very efficient screen, keeping the magnetic field in the zero-field region at a very low level. Due to the fact that the eddy currents are produced by changing magnetic fields only, this scheme does not work for DC magnets. The designer of an eddy current septum needs to find a good engineering compromise between the thickness of the septum and the rate of change of the magnetic field. The higher the magnetic field change rate is, the thinner septum that is needed. Due to the dynamic behaviour of the eddy currents they attenuate well the high frequency components of the magnetic pulses but they screen poorly the low frequency ones. The low frequency pulse components, even though attenuated, propagate slowly through the septum and they appear on the other side with a large time delay. This delay could be several times longer than the magnetic pulse itself and could be easily overlooked in time domain simulations. Especially in circular accelerators, this

delayed leakage field could compromise the overall performance of the septum. Some additional magnetic screening could further improve the leakage field performance.

Eddy current septa use short (high frequency) current pulses and they are classified in a different group to emphasize the importance of the current pulse time profile for their proper operation. The need for a large magnetic field change rate requires a low magnet inductance and adds complexity to their pulsed power supplies. Usually, this type of magnet has one turn excitation coil to maintain the magnet's inductance and the required operating voltage low. Demanding electrical switches are needed to produce the short and large amplitude current pulses. Nevertheless, eddy current septa provide the thinnest magnetic septum and very good magnetic field screening, reaching main field to leakage field ratios of over 1000:1.

In the literature (see Refs. [33−35]) it is possible to find analytical models that describe the delayed propagation of the leakage field through a septum. Figure 20 illustrates its behaviour for a 5 mm thick copper septum, 1 T square main pulse field with duration 20 μs.

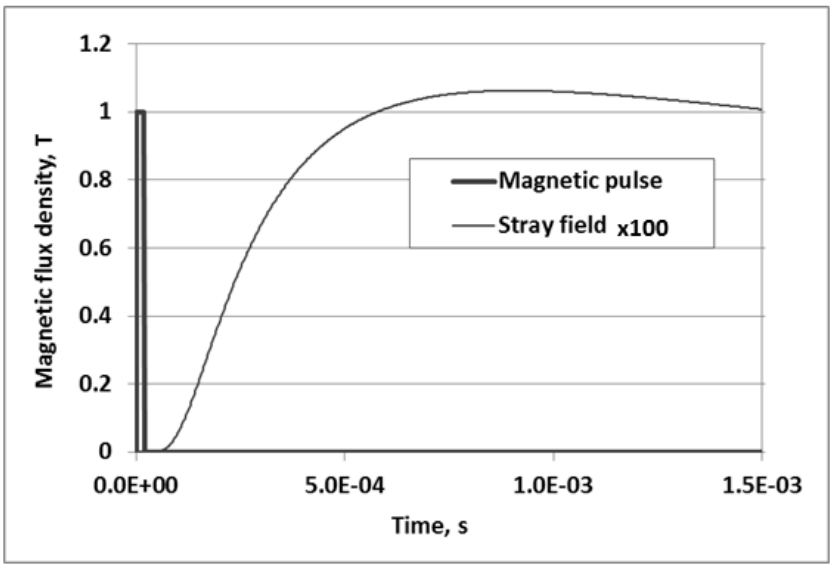


Fig. 20: Delayed leakage field from short unipolar magnetic pulse

Equations (20) and (21) describe analytically the delay time t_m at which the leakage field reaches its maximum value B_m (Ref. [34]):

$$t_\mathrm{m} = \frac{1}{2} {d_\mathrm{s}}^2 \sigma \mu \,, \tag{20}$$

$$B_\mathrm{m} = B_0 \frac{2\sqrt{2}\tau}{d_\mathrm{s}\sigma\mu\lambda_\mathrm{c}\sqrt{\pi e}}, \tag{21}$$

where d_s is the septum thickness, σ is the septum conductivity, μ is the septum permeability, B_0 is the amplitude of the main field, τ is the width of the pulse, and λ_c is the characteristic length of stray field decay.

In order to find an analytical solution, sometimes these methods simplify the problem using approximations and assumptions that deteriorate the result accuracy. They are often good for a qualitative understanding of the problem but show quantitative discrepancies with the real results. Other methods, such as 3D electromagnetic simulation or direct measurements, should be considered to gain confidence in the result.

The presence of a DC (or low frequency) component in the main field affects strongly the screening performance of eddy current septa. Figure 21 shows the measured leakage field (upper graph)

for different main field pulse waveforms (lower graph). Pulses with zero DC component yield the lowest leakage field.

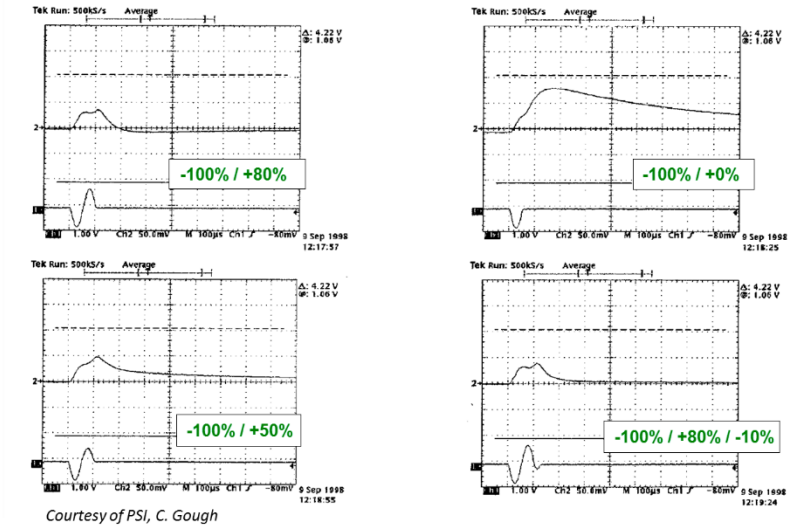


Fig. 21: Eddy current screening performance for different pulse shapes

A typical example of an eddy current septum is the 'in-vacuum' injection septum Y12 of the Swiss light source (SLS) at the Paul Scherer Institute (PSI), shown in Fig. 22 (Ref. [36]). Table 12 summarizes its basic parameters.

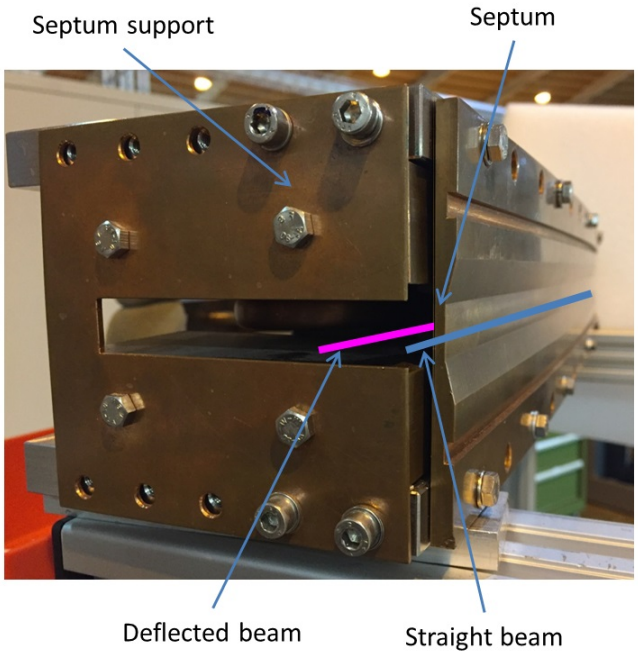


Fig. 22: 'In-vacuum' injection septum Y12 (SLS, PSI)

In this particular design, the eddy current septum reduces the leakage field to less than 0.1% of the main field. After adding an additional magnetic screening around the circulating beam (pipe of thin μm metal sheet) the leakage field is practically 'invisible' to the used measurement method ($< 10^{-5}$ of the main field) – see Refs. [36,37].

Table 12: Specification of 'In-vacuum' injection septum Y12 (SLS, PSI)

Parameter	Value
Device design	'In-vacuum'
Vacuum, mbar	10^{-7}
Field length, mm	600
Gap height, mm	6
Gap width, mm	20
Beam momentum, GeV/*c*	2.4
Deflection angle, mrad	70
Septum thickness, mm	2.5
Current, kA	4.3 (full-sine 0.16 ms)
Magnetic flux density, T	0.9

5.8 Practical considerations

(i) *High mechanical and thermal stress* – Adequate septum support should be used in order to handle the electromagnetic forces. Mechanical damping can reduce the effect of 'hammering'. Attention should be given to coping with large thermal cycling and adequate cooling should be provided.

(ii) *The maximum leakage field of eddy current septum is delayed* – An eddy current septum should be thick enough to effectively screen the main field. Attention should be given to the fact that some leakage field could appear on the other side of the septum (penetrate the septum) with a long delay. Additional screening might be needed to deal with the problem.

(iii) *Good field region* – Field homogeneity could be improved using magnetic poles geometry (shims etc.).

(iv) *Leakage dipole field* – Additional magnetic shielding might improve septum performance. Care should be taken that the shielding material does not saturate.

(v) *Beam impedance – Beam* impedance changes should be minimized using proper screening.

(vi) *Cooling* – Turbulent flow removes heat more efficiently but it increases the erosion rate and vibrations.

(vii) *Insulators degradation* – Use radiation-hard isolation materials whenever possible. In high radiation level areas, vulnerable materials (cable insulation, insulation oil) have to be exchanged in a timely manner to avoid failures.

(viii) *Alignment* – Use adequate (remote) positioning systems.

(ix) *Good vacuum* – Good vacuum septa should be bake-out compatible. Initial investment in integrating in situ bake-out systems could save long term costs and effort especially if the vacuum is often broken. Good vacuum conductivity and NEG coatings should be combined to maintain low background gas pressure.

(x) *Machine protection* – Septa are often the highest energy density elements within the entire accelerator. Adequate machine protection should be used to avoid damage in case of system failure or operator error.

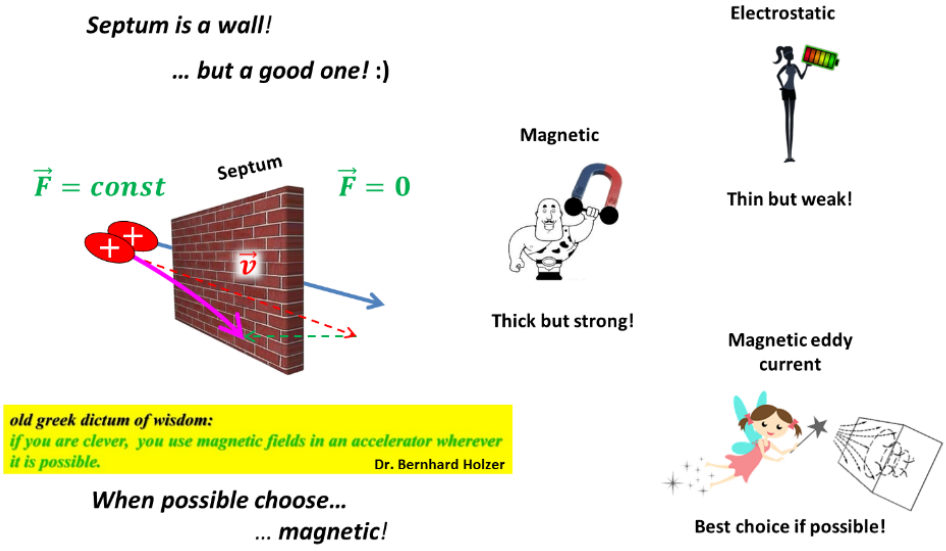
(xi) *Avoid brazed joints in vacuum as much as possible* – Brazed joints are often the most likely places for a failure to occur.

(xii) *Vacuum (cold) welding* – In the absence of air or other contaminants, metal surfaces (of the same metal) in contact could undergo 'cold welding', fusing permanently the two elements into one piece of metal. Use silver-plated bolts in steel threads to avoid bolts jamming.

(xiii) *Activation* – Care should be taken when handling activated materials. Maintenance should be done according to the radiation safety regulations.

(This list is not exhaustive.)

5.9 What to remember



Acknowledgements

Special thanks go to Dr Mike Barnes (CERN) for providing very useful material, Chris Gough (PSI) for providing very useful material and Dr Sladana Dordevic (PSI) for helping with the field simulations.

References

[1] A. Afonin, *et al.*, Proton Extraction from IHEP Accelerator Using Bent Crystals, International Workshop on Relativistic. Channeling and Related Coherent Phenomena, Italy, 2004.

[2] S. Baker *et al.*, *Nucl. Instrum. Methods* **A234** (1985), 602, https://doi.org/10.1016/0168-9002(85)91014-9.

[3] M. Barnes, Beam Transfer Devices Septa, CAS-CERN Accelerator School, Belgium, 2009.

[4] C. Bovet *et al.*, A selection of formulae and data useful for the design of A.G. synchrotrons (CERN, Switzerland, 1970).

[5] B. Holzer, Introduction to transverse beam dynamics, CAS-CERN Accelerator School, Slovakia, 2012.

[6] K. Wille, *The Physics of Particle Accelerators: An Introduction* (Oxford University Press, Oxford, 2000), p. 148.

[7] M. Reiser, *Theory and Design of Charged Particle Beams* (Wiley, New Jersey, 2008), p. 18.

[8] J. Borburgh *et al.*, Final results on the CERN PS electrostatic septa consolidation program (PAC, USA, 2003).

[9] J. Borburgh *et al.*, Consolidation project of the electrostatic septa in the CERN PS ring (PAC, USA, 2001).

[10] L. Sermeus *et al.*, The Design of the special magnets for PIMMS/TERA (EPAC, Switzerland, 2004).

[11] P. Bryant *et al.*, Proton-Ion Medical Machine Study (PIMMS) Part II, CERN-PS-2000-007-DR (CERN, Switzerland, 2000).

[12] J. Borburgh *et al.*, Design and construction of the LEIR injection septa, LHC Project Report 810 (CERN, 2005).

[13] J. Wenninger, Introduction to slow extraction to the North Targets, Training for operation (CERN, Switzerland, 2007).

[14] R. Barlow *et al.*, A new spark detection system for the electrostatic septa of the SPS North Targets, 14th ICALEPCS, USA, 2013.

[15] B. Goddard *et al.*, Beam loss damage in wire septum, EPAC 2000, Austria, 2000.

[16] H. Schonauer, Loss management studies for the PSB, Summary of the CERN Meeting on Absorbers and Collimators for the LHC, CERN, 2002.

[17] H. Schonauer, Loss concentration and evacuation by mini-wire-septa from circular machines for spallation neutron sources (PAC, USA, 1995).

[18] M. Traveria, *Nucl. Instrum. Methods* **A412** (1998), 183, https://doi.org/10.1016/S0168-9002(98)00582-8.

[19] A. Gabard *et al.*, Radiation hard magnets at the Paul Scherrer Institute, IPAC, USA, 2012, pp. 3518−3520.

[20] M. Yoshimoto *et al.*, Designs of Septum Magnet at 3 GeV RCS in J-PARC, EPAC, Scotland, 2006.

[21] M. Watanabe *et al.*, *IEEE Trans. Appl. Supercond.* **16** (2006).

[22] M. Yoshimoto *et al.*, *IEEE Trans. Appl. Supercond* **18** (2008).

[23] K. Fan *et al.*, *Nucl. Instrum. Methods* **A565** (2006), 439, https://doi.org/10.1016/j.nima.2006.06.031.

[24] J. Borburgh *et al.*, A new set of magnetic septa in the CERN PS Complex (PAC, USA, 1999).

[25] D. Manglunki *et al.*, Beam extraction, PS/SPS User Meeting 2015, CERN, Switzerland, 2015.

[26] O. Bruning *et al.*, LHC Design Report, Vol.1 The LHC Main Ring, Ch. 16 Injection System, (CERN, Geneva, 2004).

[27] M. Gyr, Expected magnetic field quality of the LHC septum magnets used for injection (MSI) and for extraction to the beam dump (MSD), LHC Project Note 129/rev. (CERN, Switzerland, 1999).

[28] S. Dordevic, SwssFEL Septum, Internal presentation, 2016.

[29] I. Sakai *et al.*, Operation of an opposite-field septum magnet for the J-Parc main-ring injection (EPAC, Scotland, 2006).

[30] K. Fan, *Nucl. Instrum. Methods* **A597** (2008), 142.

[31] S. Fartoukh, A semi-analytical method to generate an arbitrary 2D magnetic field and determine the associated current distribution, LHC Project Report 1012 (CERN, Switzerland, 2007).

[32] Y. Iwashita *et al.*, Massless septum with hybrid magnet (PAC, Sweden, 1998).

[33] W. Kang *et al.*, Development of an eddy-current septum magnet for the SSRF storage ring (APAC, China, 2001).

[34] W. Kang *et al.*, Study on the Stray field of the eddy-current septum magnet for the SSRF storage ring China.

[35] B. Kang *et al.*, *Nucl. Instrum. Methods* **385** (1997), 6, https://doi.org/10.1016/S0168-9002(96)01025-X.

[36] C. Gough, Minimizing leakage field from eddy current septum magnets, Presentation at SINAP 2005.

[37] C. Gough *et al.*, Septum and kicker systems for the SLS (PAC, USA, 2001), https://doi.org/10.1109/PAC.2001.988238.

Appendix A: Derivation of electrostatic deflection equation

All quantities are in SI units unless otherwise stated.

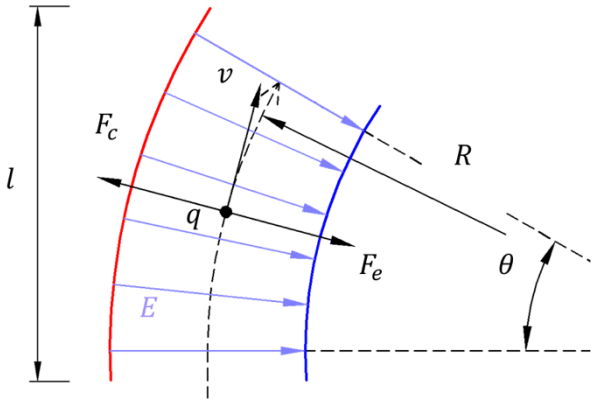


Fig. A.1: Charged particle travelling between curved electrodes that follow the particle's trajectory

Charged particle q with elementary charge q_e (positive) travels with instantaneous velocity v between curved electrodes that follow the particle's trajectory (as shown in Fig. A.1, where the red line represents positive and blue the negative electrode) and they create a constant electric field E along the particle's trajectory. The field is always perpendicular to the direction the particles travel, and the particle will experience a constant electric force F_e, aways perpendicular to the particles direction of travel, and it will follow an arc trajectory with radius R, balancing the electric field force F_e and the centrifugal force F_c. Equations (A.1) and (A.2) express the electric force and the centrifugal force respectively.

$$F_e = q_e E \,, \tag{A.1}$$

$$F_c = \frac{\gamma m_0 v^2}{R} \,, \tag{A.2}$$

where γ is a relativistic factor that takes into account the transverse relativistic effects on the particle dynamics with rest mass m_0. If the particle does not experience other forces the electric force F_e is equal and opposite to the centrifugal force F_c:

$$q_e E = \frac{\gamma m_0 v^2}{R} \,. \tag{A.3}$$

Expressing the arc trajectory radius R we have

$$R = \frac{\gamma m_0 v^2}{q_e E} \,, \tag{A.4}$$

$$p = \gamma m_0 \beta c \,, \tag{A.5}$$

$$v = \beta c \,. \tag{A.6}$$

Substituting the relativistic factor γ, using relativistic momentum p (Eq. (A.5)) and the velocity of the particle v expressed by the relativistic factor β (Eq. (A.6)) for the arc trajectory radius R we have

$$R = \frac{p\beta c}{q_e E}\,. \tag{A.7}$$

Changing the particle's momentum from SI units $[\frac{\text{kg·m}}{\text{s}}]$ to accelerator units $[\frac{\text{GeV}}{c}]$, using Eq. (A.8) yields

$$p[\tfrac{\text{kg·m}}{\text{s}}] = \frac{10^9 q_e}{c} p[\tfrac{\text{GeV}}{c}]\,, \tag{A.8}$$

$$R = \frac{10^9 \beta p}{E}\,. \tag{A.9}$$

For a bending radius R much larger than the length of the deflecting field l we can express the bending angle θ using the small angle approximation $\tan(\theta) \approx \theta$:

$$\theta \approx \frac{l}{R}\,. \tag{A.10}$$

Substituting R in Eq. (31) we obtain the final expression for the bending angle θ (note the momentum p is in GeV/c)

$$\theta \approx \frac{El}{10^9 \beta\, p}\,. \tag{A.11}$$

Geometrical length of a septum differs from the electric field length due to the mechanical construction and/or the fringe fields and an integral effective value l_{eff} is used to describe more precisely the effective interaction length.

To emphasize the parallel between (the duality of) the electric and magnetic deflection, curved electrodes are used to produce the deflecting electric field, always perpendicular to the particle's direction of travel. As we will see later, this is the case for magnetic deflection.

In practice, straight deflecting electrodes are more often used. It can be shown that if the deflection angle is small (gained transverse velocity v_x is much smaller than main velocity v_z), Eq. (A.11) gives still a good approximation of the deflection even for straight electrodes.

Appendix B: Derivation of magnetic deflection equation

All quantities are in SI units unless otherwise stated.

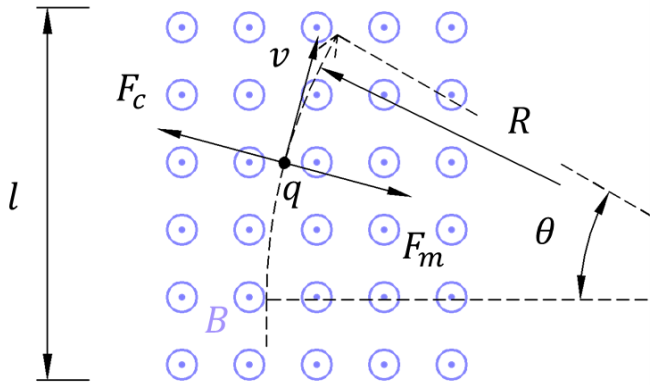


Fig. B.1: Charged particle travelling in transverse magnetic field

Charged particle q with elementary charge q_e (positive) travels with instantaneous velocity v through a constant transverse magnetic field B, going out of the plane of the paper (as shown in Fig. B.1). The particle will experience a constant magnetic force F_m, aways perpendicular to the particles direction of travel and to the direction of the magnetic field, and it will follow an arc trajectory with radius R, balancing the magnetic field force F_m and the centrifugal force F_c. Equations (B.1) and (B.2) express the scalar form of the magnetic force and the centrifugal force respectively.

$$F_m = q_e v B \,, \tag{B.1}$$

$$F_c = \frac{\gamma m_0 v^2}{R} \,, \tag{B.2}$$

where γ is a relativistic factor to take into account the transverse relativistic effects on particle's the dynamics with rest mass m_0. If the particle does not experience other forces the magnetic force F_m is equal and opposite to the centrifugal force F_c:

$$q_e v B = \frac{\gamma m_0 v^2}{R} \,. \tag{B.3}$$

Expressing the arc trajectory radius R we have

$$R = \frac{\gamma m_0 v^2}{q_e v B} \,, \tag{B.4}$$

$$p = \gamma m_0 \beta c \,, \tag{B.5}$$

$$v = \beta c \,. \tag{B.6}$$

Substituting the relativistic factor γ, using the relativistic momentum p (Eq. (B.5)) and the velocity of the particle v expressed by the relativistic factor β (Eq. (B.6)) for the arc trajectory radius R we have

$$R = \frac{p}{q_e B} \,. \tag{B.7}$$

Changing the particle's momentum from SI units $[\frac{\text{kg·m}}{\text{s}}]$ to accelerator units $[\frac{\text{GeV}}{c}]$ using Eq. (B.8) and using the approximate value of the speed of light $c \approx 3{\times}10^8$ m/s, yields

$$p[\tfrac{\text{kg·m}}{\text{s}}] = \frac{10^9 q_e}{c} p[\tfrac{\text{GeV}}{c}] \,, \tag{B.8}$$

$$R \approx \frac{10^9 p}{3{\times}10^8 B} \approx \frac{p}{0.3B} \,. \tag{B.9}$$

For a bending radius R much larger than the length of the deflecting field l we can express the bending angle θ using the small angle approximation $\tan(\theta) \approx \theta$:

$$\theta \approx \frac{l}{R} \,. \tag{B.10}$$

Substituting R in Eq. (B.10) we obtain the final expression for the bending angle θ (note the momentum p is in GeV/c)

$$\theta \approx \frac{0.3Bl}{p} \,. \tag{B.11}$$

Geometrical length of a septum differs from the magnetic field length due to the mechanical construction and/or the fringe fields and an integral effective value l_{eff} is used to describe more precisely the effective interaction length.

Proceedings of the CAS–CERN Accelerator School: Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018)

Measurement Techniques for Transfer Lines and Beam Instrumentation

P. Forck
GSI Helmholtz-Zentrum für Schwerionenforschung, Darmstadt, Germany

Abstract
The determination of beam parameters is essential for the operation and development of any accelerator facility. The working principle of beam instrumentation for electron and proton beams, with a focus on transfer line diagnostics and for injection and extraction matching, is discussed. This comprises beam instrumentation for beam current determination, beam loss monitors, methods for transverse profile and emittance diagnostics, and the usage of beam position monitors for bunched beams. Some examples of such measurements are also presented.

Keywords
Beam instrumentation; transformer; profile and position monitor; emittance.

1 Demands for beam diagnostics

Beam diagnostics form an essential constituent of any accelerator, as they make up the sensory organ showing the properties and the behaviour of the beam. Beam diagnostics deals with the real beam, including all possible imperfections of a real technical installation. There is a very wide range of diagnostic applications, resulting in quite different demands for the beam instruments and the related display of the measurement results.

- Reliable, quick measurements to determine the basic parameters of a machine setting are used as a fast check of the general functionality of the accelerator. The readings from the instrumentation give a single number or simple plots. These devices should be non-destructive for the beam, yielding online information. A prominent example is current measurement by a current transformer, as discussed in Section 2.
- Instruments are installed for a daily check of performance and stability, or for the control of wanted parameter changes in the accelerator setting. They are also used for solving simpler machine problems in the case of any device malfunction. An example is transverse profile determination, in most cases performed by a destructive device, as discussed in Section 4.
- Complex instrumentation for the commissioning of a new accelerator component, for the development of higher performance, and for solving more serious problems in the case of a malfunction must be available. The devices can be more complex to use and might be destructive for the beam. The importance is the creation of reliable information about the complex beam behaviour, allowing a clear interpretation. The determination of beam emittance is an important example, as discussed in Section 5.
- An additional application is the active control of device settings after the detection of the beam's properties, where the beam instrumentation serves as the sensor. This is generally called feedback of a given parameter to yield improved beam delivery. An example is reading of the beam position inside a synchrotron and correction of the orbit to its nominal value A second example is determination of the beam current for slow extraction followed by active control of the extraction parameters to achieve current values as close as possible to a set value.

There is a large variety of beam parameters to be measured. For good alignment of the beam, all relevant parameters should be controllable. In this contribution, we focus on the instrumentation of

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-005.395

transfer lines and instrumentation within a synchrotron to determine injection and extraction parameters. Some parameters are controlled at transfer lines, but the results of such manipulations are better determined within the synchrotron. Moreover, relevant measurements are shown to describe the possibilities, limitations, and data quality delivered by the instruments. For general articles and books on beam diagnostics, see Refs. [1–7].

Most of the examples originate from the GSI facility, because the author is most familiar with these realizations. This does not mean that GSI has the most advanced systems; at other facilities, better installations might exist. The GSI accelerator facility consists of two pulsed linacs running at 36 MHz and 108 MHz, respectively, to accelerate all ions from different ion sources up to 18 MeV/u, corresponding to a velocity of $\beta = v/c = 0.19$. This energy is well above the so-called Coulomb barrier, where nuclear reactions are possible and performed at fixed targets. To achieve higher energies, up to 2 GeV/u, corresponding to $\beta = 0.95$, a synchrotron with 218 m circumference is used. The beam can be transferred to a storage ring with a single-bunch transfer (so-called fast extraction). Alternatively, it can be sent to fixed-target experiments using slow extraction, lasting from 1 to 10 s. The scientific investigations at these targets are related to nuclear, atomic, or plasma physics, as well as human cancer therapy.

The outline of this contribution is oriented on beam quantities. First, measurement of beam current at transfer lines for pulsed or slowly extracted beams and within a synchrotron is discussed. Section 3 deals with determination of beam losses; this can be used for alignment, because the fraction of beam that is lost can sometimes be determined more sensitively than the transmitted current. Moreover, the signal from such beam loss monitors can serve as an interlock generation in the case of device malfunction. Section 4 describes several methods for transverse profile measurements. There are methods that can be operated for all beams, but also quite different methods are used for electrons (which are, in most cases, relativistic) and protons (which require a large facility to reach relativistic velocities). Some of these methods are destructive for the beam and cannot, therefore, be applied at a synchrotron to monitor the beam properties of the circulating beam. Section 5 deals with the determination of transverse emittance in transfer lines, where the statistical description of the ensemble of beam particles is relevant. Section 6 describes very frequently used beam position monitors that determine the transverse centre of mass of a bunched beam. In nearly all facilities, these instruments are used for steering, focusing, and matching alignment at transfer lines and synchrotrons. Section 7 deals with the determination of longitudinal parameters, such as average beam energy, longitudinal bunch profile, and phase space distribution. Different devices are used as required for the quite different bunch lengths at electron and proton facilities.

The special beam instrumentation required for a linac-based fourth-generation light source with free electron laser application is not discussed in this contribution. Those accelerators are operated with very short bunches (down to the femtosecond range) and sometimes short beam pulses (down to single-bunch operation), as well as very high demands in terms of resolution and reproducibility; design considerations and recent achievements are discussed in e.g., Refs. [8–10], and references therein. Moreover, diagnostics for the novel laser-based plasma wakefield accelerators for electrons or ions are not covered; owing to their different installation schemes and related beam parameters, the instrumentation is partly based on modified methods to cover a very large dynamic range.

2 Measurement of beam current

The total electrical current is one of the most important parameters for the operation of a particle accelerator. In the daily operation of almost all accelerator laboratories, a first check concerns the current; in most cases, it is measured using a beam current transformer. Several types of current transformer are available to determine the current at electron and proton linacs and synchrotrons, even for short pulses, such as the transfer between synchrotrons, as well as for de-bunched beams, such as in a storage ring. These devices are commercially available [11], even though quite different types are used; a general overview of current measurement devices is presented in Ref. [12]. Current transformers are generally non-intercepting devices. Their principle is the detection of the magnetic field carried by the beam. For

currents below about 1 µA, transformers cannot be used, owing to noise limitations.

From the first days of accelerators, Faraday cups were used. The cup gives a direct measurement of the particle beam's charge because the particles are stopped in the cup. For large currents, this destructive method cannot be applied, because the total energy carried by the beam can destroy the intercepting material. For higher energetic particles, with energies above some 100 MeV/u for ions, the penetration depth reaches more than several centimetres and Faraday cups are no longer useful.

For low currents but higher energies, above about 100 MeV/u for ions, particles detectors are used. Here, the energy loss of the charged particle travelling through matter is detected. Either the particles are counted directly or the particle flux can be calculated with the help of a calibration. For most parameters, this is an intercepting method. Slow extraction from a synchrotron within a typical duration of some seconds is an example for such beam parameters.

Pulsed linacs or pulsed cyclotrons can be used as an injector to a synchrotron, where only a typical pulse length of 100 µs is needed for a multiturn injection. For a single-turn injection, the time is of the order of 1–10 µs. For other applications, such as atomic or nuclear physics, a continuous wave accelerator is used. Here, the bunches are delivered for an 'infinitely' long time, but the bunch structure still remains, owing to the RF acceleration by a linac or a cyclotron. There also exist types of accelerator that do not produce bunched beams, examples are the van de Graaff and Cockcroft–Walton types, based on electrostatic acceleration.

2.1 Current transformer for pulsed beams

2.1.1 General considerations and passive or fast current transformer

In an accelerator, the current is formed by N_{part} particles of charge state q per unit of time t or unit of length l and velocity $\beta = v/c$. The electrical current passing a given location is

$$I_{\mathrm{beam}} = \frac{qeN_{\mathrm{part}}}{t} = \frac{qeN_{\mathrm{part}}}{l} \cdot \beta c \,, \tag{1}$$

where e is the elementary charge. The magnetic field B of a current can be calculated according to the Biot–Savart law

$$\mathrm{d}\vec{B}(\vec{r}) = \mu_0 I_{\mathrm{beam}} \cdot \frac{\mathrm{d}\vec{l} \times \vec{r}}{4\pi r^3} \,, \tag{2}$$

where $\mu_0 = 4\pi \times 10^{-7}$ Vs/Am is the permeability of the vacuum, $\mathrm{d}\vec{l}$ is the length in the direction of the beam and $\vec{r}$ is the distance between the centre of the beam and the location of the field determination. Owing to the cylindrical symmetry outside of the beam, only the azimuthal component has to be considered along the unitary vector $\vec{e}_{\varphi}$, as shown in Fig. 1:

$$\vec{B} = \mu_0 \frac{I_{\mathrm{beam}}}{2\pi r} \cdot \vec{e}_{\varphi} \,. \tag{3}$$

For a beam current of 1 µA and a distance of 10 cm, the magnetic field has a value of only 2 pT. To put this into perspective, the Earth's constant and homogeneous magnetic field has a value of about 50 µT.

The beam current can be determined by monitoring the accompanied magnetic field with a current transformer, as shown schematically in Fig. 2. The beam passes through a highly permeable torus as the 'primary winding'. An insulated wire, wound around the torus with N turns, serves as the 'secondary winding' of the transformer, with inductance L. The inductance L for a torus material of length l in the beam direction, inner radius r_{i} and outer radius r_{o} having a relative permeability μ_{r} and N windings is given by

$$L = \frac{\mu_0 \mu_{\mathrm{r}}}{2\pi} \cdot lN^2 \cdot \ln \frac{r_{\mathrm{o}}}{r_{\mathrm{i}}}. \tag{4}$$

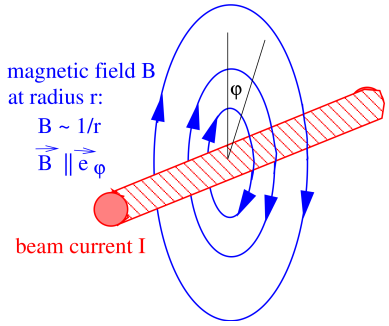


Fig. 1: Magnetic field of a current

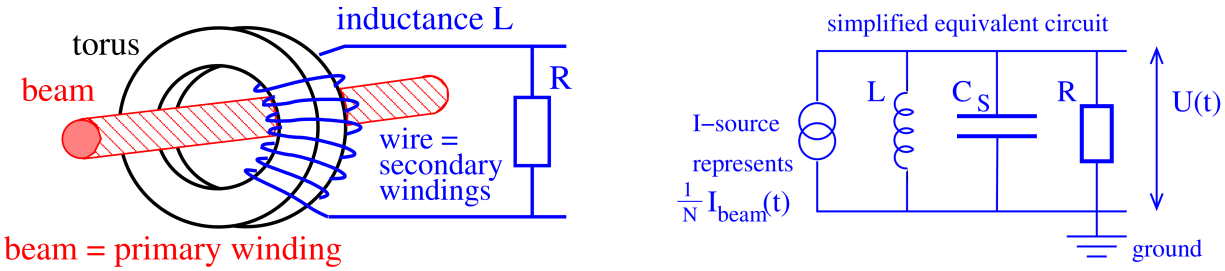


Fig. 2: Left: Scheme of a current transformer built as a ring-core (torus) around the beam. Right: Simplified equivalent circuit.

The value of the induction can be chosen by the core size (as given by r_i and r_o) and the number of windings N. A further reason for the torus is to guide the field lines, so only the azimuthal component is measured and the signal strength is nearly independent of the beam position inside the vacuum pipe.

Generally, for an ideal current transformer loaded with a low value of ohmic resistance R, the ratio between the primary current I_prim conducted in N_prim windings and the secondary current I_sec conducted in N_sec is given by

$$I_\mathrm{sec} = \frac{N_\mathrm{prim}}{N_\mathrm{sec}} \cdot I_\mathrm{prim} \quad \Longrightarrow \quad I_\mathrm{sec} = \frac{1}{N} \cdot I_\mathrm{prim} \quad \text{as} \;\; N_\mathrm{prim} = 1 \tag{5}$$

$N_\mathrm{prim} = 1$ owing to the single pass of the beam through the torus. N_sec is the winding number on the secondary side; for simplicity, it is called N in the following.

For most practical cases, a measurement of a voltage U is preferred; therefore, the resistance R is introduced, leading to

$$U = R \cdot I_\mathrm{sec} = \frac{R}{N} \cdot I_\mathrm{beam} \quad . \tag{6}$$

The ratio between the usable signal voltage U and the beam current I_beam is called the sensitivity S (or transfer impedance; for a more stringent discussion, see Section 6.1)

$$U = S \cdot I_\mathrm{beam} \quad . \tag{7}$$

The properties of a transformer can be influenced using different external electrical elements to match the response to a given time structure of the beam. We first consider the characteristics of a so-called passive transformer or fast current transformer, where the voltage at a 50 Ω resistor is recorded. The equivalent circuit of the secondary transformer side is depicted in the right-hand side of Fig. 2. The beam current is modelled by a current source with a reduction given by the number of windings N, according to Eq. 5. One also has to take some stray capacitance C_S into account, which is caused by the capacitance between the windings and the torus and along the shielded cable to the resistor R. Generally,

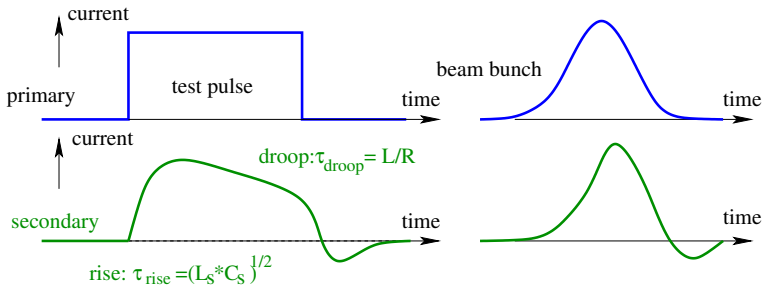


Fig. 3: Response of a.c. transformer to a rectangular pulse and a more realistic beam pulse

the voltage $U(t)$ of the parallel shunt of the three elements is measured. The response is influenced by the choice of elements, as discussed in Refs. [1–7, 12]; here, we state only the results.

For the following discussion, we are interested in the time response of the measurement device to a given beam pulse. Therefore, one frequently uses the rise time constant τ_{rise} and the droop time constant τ_{droop}, as depicted in Fig. 3. If the excitation is given by a step function, the signal amplitude A increases as $A \propto (1 - \mathrm{e}^{-t/\tau_{\mathrm{rise}}})$ and τ_{rise} corresponds to the time for an increase by $\mathrm{e}^{-1} = 37\ \%$. It is linked to the upper cut-off frequency by

$$\tau_{\mathrm{rise}} = \frac{1}{\omega_{\mathrm{high}}} = \frac{1}{2\pi f_{\mathrm{high}}} \quad . \tag{8}$$

Correspondingly, the droop time constant is linked to the lower cut-off frequency, as

$$\tau_{\mathrm{droop}} = \frac{1}{2\pi f_{\mathrm{low}}} \quad . \tag{9}$$

Both equations have a general meaning and will be used to transfer quantities, such as bandwidth, given in the frequency domain, into the corresponding description in the time domain

For the passive current transformer described by the equivalent circuit of Fig. 2, the rise and droop time constants are given by

$$\tau_{\mathrm{rise}} = RC_{\mathrm{S}} \quad \text{and} \quad \tau_{\mathrm{droop}} = \frac{L}{R} \quad . \tag{10}$$

A more realistic schematic diagram of the passive transformer is shown in Fig. 4. The main difference is the additional loss resistivity in the cables, which is represented by a serial resistor R_{L}. Additionally, there is a stray inductance between the windings, which is best modelled by a serial insertion of an inductance L_{S}. With these two modifications, the rise and droop times are modified to yield

$$\tau_{\mathrm{rise}} = \sqrt{L_{\mathrm{S}} C_{\mathrm{S}}} \quad \text{and} \quad \tau_{\mathrm{droop}} = \frac{L}{R + R_{\mathrm{L}}} \quad . \tag{11}$$

The specifications of the GSI device are listed in Table 1. Careful matching is necessary between the torus and the $50\ \Omega$ resistor R, where the voltage drop is measured.

An additional general point is that close to the transformer the electrical conductivity of the beam pipe has to be interrupted, as shown schematically in Fig. 5. This is done with an insulator, either a ceramic gap or a plastic vacuum seal. The reason is to prevent a flow of image current inside the transformer torus. This image current has the opposite sign; without the gap, the fields of the image current and beam current add up to zero. The image current must be bypassed outside of the transformer torus by some metallic casing. It is surrounded by high-permeability μ-metal, also used for the shielding of the transformer against external magnetic fields. A general review of transformers is presented in Refs. [12, 14].

Passive transformers are mainly used when beam pulses in a time range from nanoseconds to microseconds are to be observed. The observables for the beam are the actual bunch shape, the arrival

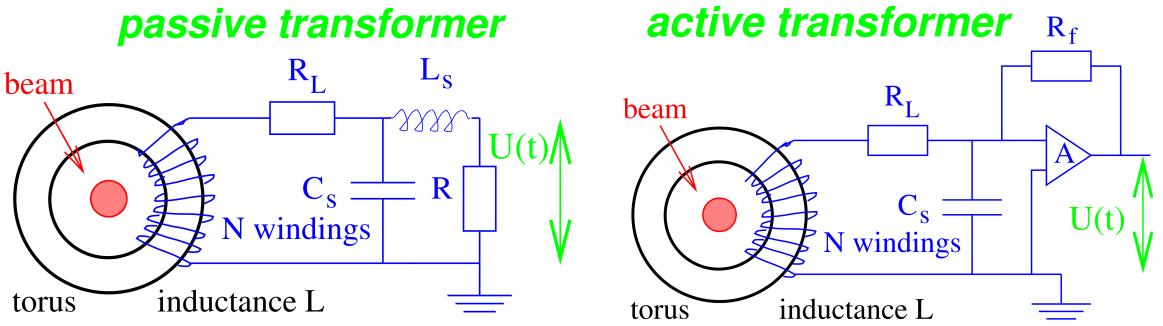


Fig. 4: Equivalent circuits. Left: Passive beam transformer. Right: Active transformer

Table 1: Basic specifications of GSI passive transformer [13]

Parameter	Value
Torus radii	$r_\mathrm{i} = 70$ mm, $r_\mathrm{o} = 90$ mm
Torus thickness, l	16 mm
Torus material	Vitrovac 6025: $(\mathrm{CoFe})_{70\%}(\mathrm{MoSiB})_{30\%}$
Torus permeability, μ_r	$\mu_\mathrm{r} \simeq 10^5$ for $f < 100$ kHz, $\mu_\mathrm{r} \propto 1/f$ for $f > 100$ kHz
Number of windings, N	10
Sensitivity, S	4 V/A at $R = 50\ \Omega$ (10^4 V/A with amplifier)
Current resolution, I_min for $S/N = 1$	40 $\mu\mathrm{A}_\mathrm{rms}$ for full bandwidth
Droop time constant,$\tau_\mathrm{droop} = L/R$	0.2 ms, corresponding to 5% per 10 µs pulse length
Rise time constant, $\tau_\mathrm{rise} = \sqrt{L_\mathrm{S} C_\mathrm{S}}$	1 ns
Bandwidth	$f_\mathrm{low} = 0.75$ kHz to $f_\mathrm{high} = 660$ MHz

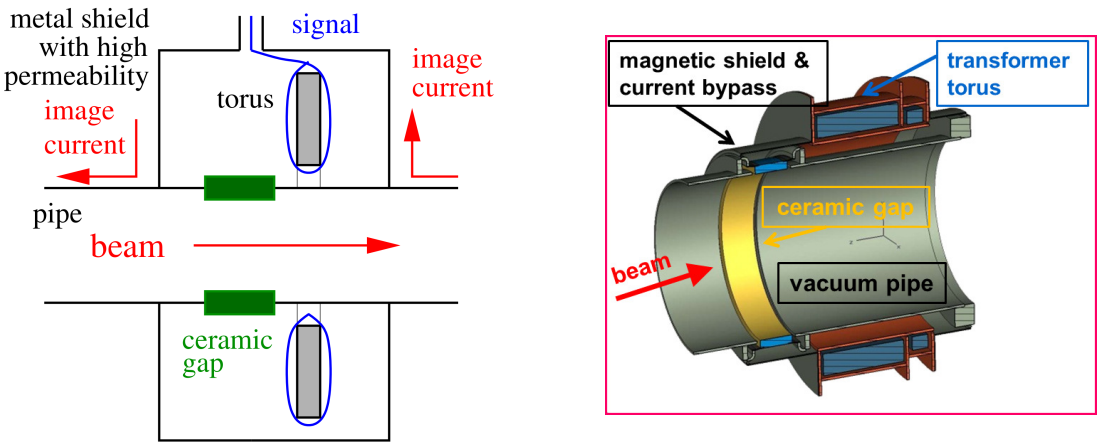


Fig. 5: Left: General scheme of beam pipe intersected by a ceramic gap and a transformer casing, as used as an image current pass and a shielding against external magnetic fields. Right: Technical drawing for a transformer arrangement surrounding a beam pipe, including the gap and magnetic shield.

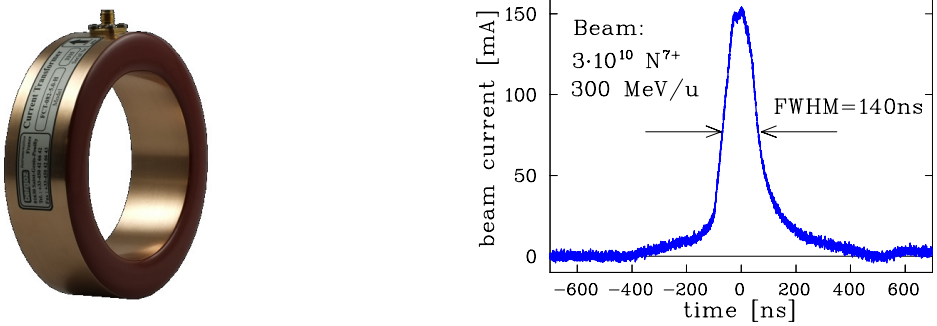


Fig. 6: Left: Commercially available passive transformer [11]. Right: Signal from passive transformer for single beam bunch extracted from GSI synchrotron at 300 MeV/u.

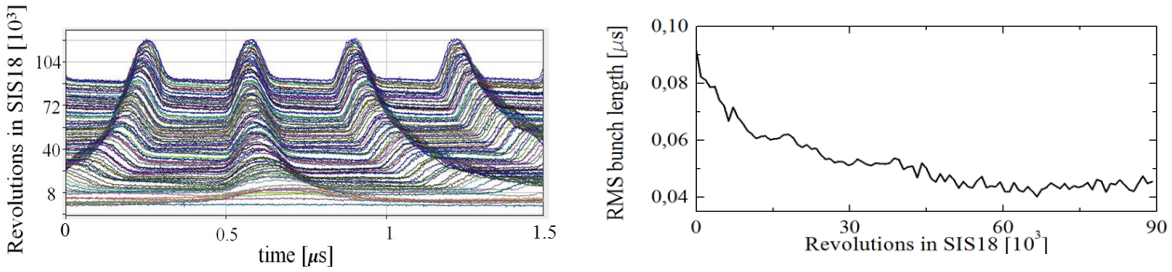


Fig. 7: Left: Signals from a passive transformer installed in the GSI synchrotron during the acceleration of a U^{73+} beam. The beam is injected at 11 MeV/u, corresponding to a velocity of $\beta = 15\%$; then four bunches are built and accelerated to 350 MeV/u, corresponding to $\beta = 69\%$, with an accelerating frequency swing from 0.85 MHz to 3.80 MHz within 0.3 s. A trace of the circulating bunches is shown for each 0.15 ms. Right: Width (one standard deviation) of the bunches during acceleration.

time with respect to an external reference (e.g., the acceleration frequency), and the total number of particles, gained by integration of the signal. A first example is shown in Fig. 6 for one bunch extracted from a synchrotron within one revolution, so-called fast extraction. For this case, the passive transformer is installed in the beam pipe outside the synchrotron. The beam pulse length is typically between 1 ns and 10 µs. Moreover, with the passive transformer, the bunch structure can be observed in the time domain and a bandwidth, typically 1 GHz, corresponding to a rise time constant of $\tau_{\text{rise}} = 160$ ps, can be achieved [11, 13].

As a second example, the observation of the bunch structure during acceleration inside a synchrotron is shown in Fig. 7 for the acceleration of an ion beam injected at a non-relativistic velocity of $\beta = 15\%$ of the velocity of light and accelerated to $\beta = 69\%$. Owing to the increase of the velocity, the bunches approach each other and get smaller, as expected from the conservation of the normalized longitudinal emittance (see Section 7). Generally, observation of bunches during acceleration and of possible bunch merging or splitting is done by the fast current transformer and, eventually, a feedback system to control the acceleration frequency, amplitude, and phase is generated from the fast current transformer signal. For injection alignment and extraction timing, the bunch shape is recorded in connection with the kicker signal and eventually corrected to prevent for varying kicker angle during the bunch passage; some examples are discussed in other contributions to these proceedings.

2.1.2 Active a.c. transformer

To observe beam pulses longer than several microseconds, the droop of a passive transformer leads to a significant deformation of the measured signal, as shown schematically in Fig. 3. The droop time constant

Table 2: Some basic specifications of GSI linac active transformer

Parameter	Value
Torus radii	$r_\mathrm{i} = 30$ mm, $r_\mathrm{o} = 45$ mm
Torus thickness l	25 mm
Torus material	Vitrovac 6025: $(\mathrm{CoFe})_{70\%}(\mathrm{MoSiB})_{30\%}$
Torus permeability, μ_r	$\simeq 10^5$
Number of windings, N	2×10, with opposite orientation
Maximal sensitivity, S	10^6 V/A (with operational amplifier)
Ranges of beam current	1 µA to 100 mA
Current resolution, I_min for $S/N = 1$	0.2 $\mu\mathrm{A}_\mathrm{rms}$ for full bandwidth
Droop	<0.5% for 5 ms pulse length
Upper cut-off frequency, f_high	1 MHz (as given by operational amplifier property)
Rise time constant, τ_rise	160 ns

can be made much longer by using an operational amplifier, with feedback resistor R_f and open-loop gain A, instead of a constant resistor to measure the voltage. The equivalent circuit is shown in the right-hand panel of Fig. 4. The effect is the reduction of the operational amplifier input resistance by a factor of R_f/A. This is the set-up of a so-called transimpedance amplifier or current-to-voltage converter. The remaining resistance is, in most cases, only given by the cable resistivity R_L, since $R_\mathrm{f}/A \ll R_\mathrm{L}$. The droop is now

$$\tau_\mathrm{droop} = \frac{L}{R_\mathrm{f}/A + R_\mathrm{L}} \simeq \frac{L}{R_\mathrm{L}} \,. \tag{12}$$

Droop time constants up to 1 s are possible. (Typical values of the cable resistance, R_L are of the order of 1 Ω.) An additional winding (not shown in the right-hand panel of Fig. 4) must be used to compensate for the 'natural', known, droop. In addition, the feedback resistor R_f of the operational amplifier can be used for range switching. A transformer that utilizes this principle is called an active transformer or alternating current current transformer.

The rise time τ_rise for this active type is much larger than for the passive type. This is mainly because of the choice of a low bandwidth (low value of the upper cut-off frequency) of the operational amplifier to obtain a significant reduction of the amplifier's high-frequency noise. This results in the high sensitivity of an active transformer.

The specifications of a typical device [13] used at the heavy-ion linac at GSI with pulse lengths of 100 µs (used when filling a synchrotron) and 5 ms (used directly in experiments) are given in Table 2; a photograph of the hardware is shown in Fig. 8. For a specification of the parameters of the commercially available type, see Ref. [11].

A measurement made with these types of transformer is shown in the left-hand panel of Fig. 9; one macroscopic pulse is recorded at three locations, starting at the ion source and ending at the exit of the first linac tank. A high-current ion beam is generated in the source with ∼0.5 ms length. The slow rise of the beam current before the linac results from the build-up of the space charge compensation of the unbunched beam. A chopper in front of the linac is used to generate the necessary pulse length. Transformers are essential to determine and maximize the transmission through the accelerator, in particular if the ion source conditions are not very stable. A non-destructive method is needed to follow the beam pulse for its pass through the full accelerator. During the operation of the accelerator, these transformers are the most frequently used diagnostics tools.

To control the filling of a synchrotron by a pulse from a linac, transformers of the described type are used; one is mounted in the external beam line, measuring the delivered current, and the other is installed inside the synchrotron to determine the stored current. The used multiturn injection is done

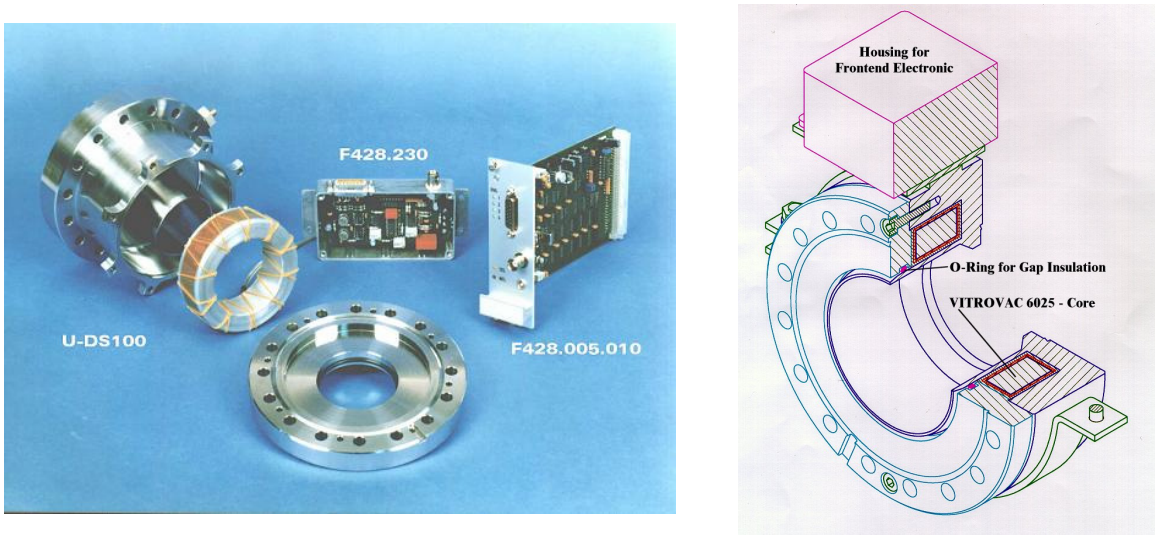


Fig. 8: Left: An a.c. current transformer used at the pulsed heavy-ion linac at GSI. The torus with the differential winding is shown, together with the electronic modules. Right: Technical drawing of same transformer.

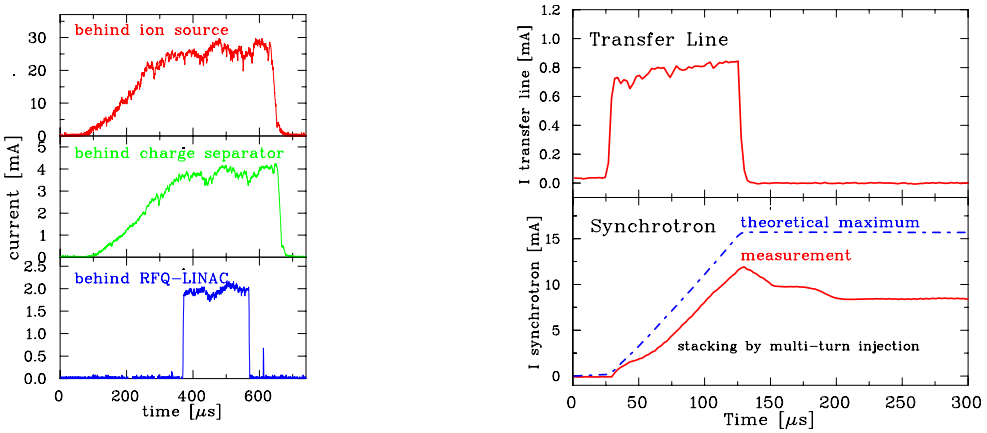


Fig. 9: Left: Transmission determination by a.c. transformers at three different locations at the GSI linac using a Ni^{2+} beam from a high-current source. The upper curve shows the current after the ion source, where all charge states are present. The middle curve shows the current behind a magnetic bend to filter the Ni^{2+} ionic state. The third curve shows the current after the first linac module. Right: Multiturn injection at GSI measured with two a.c. transformers. The upper curve is the current delivered by the linac and the lower curve is the current stored in the synchrotron. In this case, 20 turns are used and 70% efficiency is achieved for this Ni^{26+} beam.

with some bumper magnets to fill the large horizontal acceptance of the synchrotron by the linac pulse, which is much longer than the revolution period. Such a measurement at the GSI facility is shown in the right-hand panel of Fig. 9: the injection takes place over 20 turns and an accumulation by a factor of $\sim$15 is achieved. For each beam setting, optimization of the injection process is an important task, in order to achieve maximal matching.

2.2 The d.c. transformer

An important task for beam diagnostics is the measurement of a coasting (or so-called direct current) beam current. The application is either a linac that permanently produces a beam (so-called continuous wave mode) or a synchrotron with storage times from seconds to many hours. The transformers discussed in the previous sections only work for a pulsed beam, even though the droop time constant can be made to be of the order of a second.

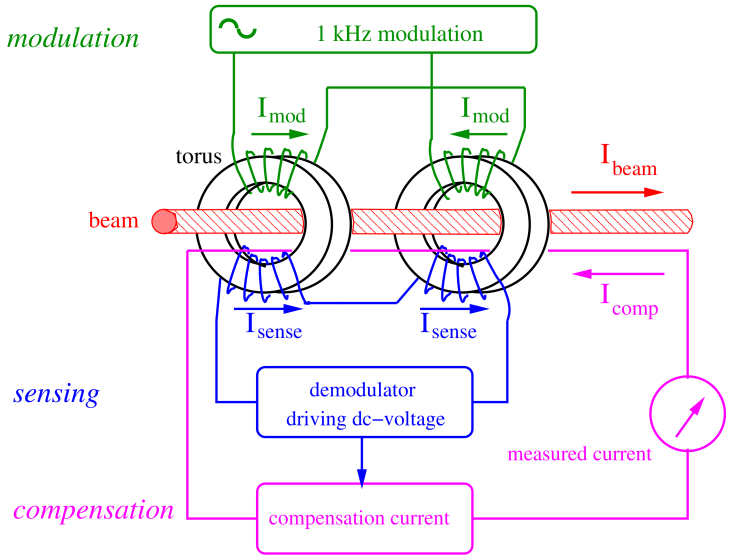


Fig. 10: Schematic of d.c. transformer (see text)

The principle of a d.c. transformer, the so-called direct current current transformer [11, 13, 15], is shown schematically in Fig. 10. It consist of two tori with three types of winding each. The first windings of each torus, with opposite orientation, are used as a modulator. The modulation frequency is typically 1–10 kHz. The amplitude of the modulation current is high enough to force the torus into a magnetic saturation of $B_{\mathrm{sat}} \simeq 0.6$ T, for positive and negative azimuthal orientation each period. The secondary windings, with equal orientation, act as a detector for the modulated signal, see Fig. 11. Assuming perfectly identical magnetic characteristics of both tori, the detector signal, as shown in the scheme, should be exactly zero if there is no beam current flowing through the tori. However, an asymmetric shifting of the hysteresis curve results if a d.c. beam is fed through the tori, because of the additional magnetic field from the beam. The sum signal U_{S} is different from zero, with a modulation twice the modulation frequency. In the demodulator stage, this signal is rectified. The d.c. current is measured by means of the current generated in the feedback compensation circuit, which forces the output signal back to zero, flowing through the third windings of both tori. The detector can even be used at even harmonics of the modulation frequency, resulting in higher sensitivity and an improvement in the signal-to-noise ratio. Owing to the extremely specific requirements concerning the matching of the magnetic characteristics for a pair of tori, the design of a magnetic modulator with high resolution and d.c. stability is rather complex and depends very much on the selection and treatment of the core-material, as discussed in more detail in Ref. [15]. The applied feedback circuit for the zero flux compensation makes the device very sensitive and linear. To get a fast response, the signal from an a.c. transformer is added to the feedback circuit. With this enhancement, the time resolution of the full device is of the range of 20 µs.

The specifications of a typical d.c. transformer developed for the heavy-ion synchrotron at GSI are given in Table 3 and a photograph is given in Fig. 12. The resolution is about 1 µA. The offset drift, mainly caused by the magnetic properties of the tori, are of the order of 20 µA per day. The offset can be improved by an automatic zero compensation when the synchrotron does not contain any beam, e.g., after the ramp-down phase of the magnets. For the parameters of the commercially available type, see Ref. [11].

An example of a d.c. transformer measurement is shown in Fig. 13. The electrical current and the number of stored particles are shown. A $^{238}\mathrm{U}^{73+}$ beam is injected in the synchrotron and acceleration starts shortly after the injection. A loss of particles is seen during the RF bunching process, owing to some misalignment. During the acceleration, the slope of the magnetic field $\mathrm{d}B/\mathrm{d}t = 1.3$ T/s is constant, resulting in a constant increase in the particle momentum. The current increases linearly only for non-

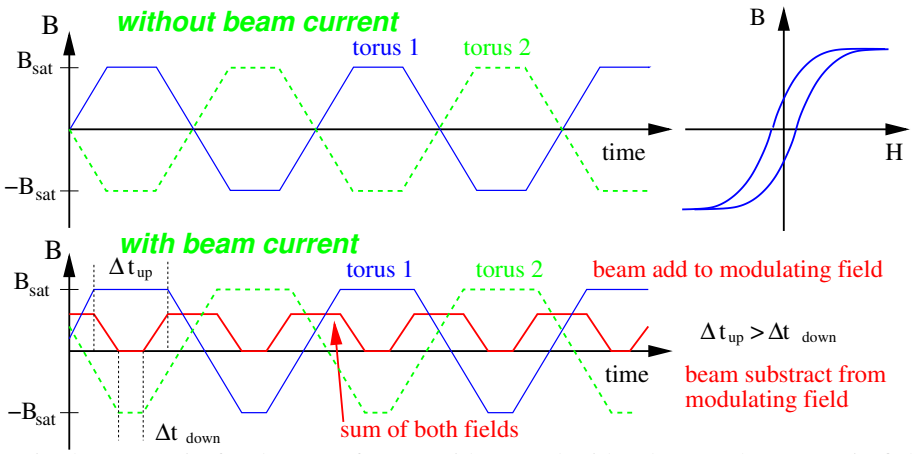


Fig. 11: Fluxes in the two tori of a d.c. transformer without and with a beam. The magnetic field of the beam adds to the modulated field for one modulation phase and 180° later it is subtracted. The sum of the maximum magnetization gives the measured signal. The scheme is drawn for a triangular-modulation; for a realistic sine modulation, the edges are smoother.

Table 3: Specifications of d.c. transformer installed at GSI synchrotron

Parameter	Value
Torus radii	$r_\mathrm{i} = 135$ mm, $r_\mathrm{o} = 145$ mm
Torus thickness, l	10 mm
Torus material	Vitrovac 6025: $(CoFe)_{70\%}(MoSiB)_{30\%}$
Torus permeability	$\mu_\mathrm{r} \simeq 10^5$
Isolating gap	Al_2O_3
Number of windings, N	16 for modulation and sensing
	12 for feedback
Ranges for beam current	300 µA to 1 A
Current resolution, I_min for $S/N = 1$	2 µA
Bandwidth	d.c. to 20 kHz
Rise time constant, τ_rise	20 µs
Offset compensation	±2.5 µA in auto mode
	<15 µA/day in free run
Temperature coefficient	1.5 µA/°C

relativistic velocities, owing to the decrease in the revolution time, which is inversely proportional to the velocity. For the maximal energy of 750 MeV/u, corresponding to $\beta = 84\%$, the relativistic increase of the mass starts to be significant. After reaching the maximum energy, a ∼0.5 s flat top is seen for the de-bunching phase. The slow extraction of several seconds is achieved by changing the tune using a fast quadrupole to get close to a 1/3 resonance, to blow up the transverse beam dimension. The resonant beam particles are then separated from the circulating stable beam by an electrostatic septum and are extracted.

2.3 Energy loss and ranges of particles in matter

2.3.1 Material interaction of protons and ions

For the discussion for all intercepting diagnostics, such as Faraday cups, ionization chambers, wire scanners, etc., the energy loss, with the related range of the particles penetrating the matter, serves as the basic physical mechanism. The energy loss of a proton or an ion is mainly due to the collision of the projectile with the electrons of the stopping target, so-called electronic stopping. Owing to the different masses

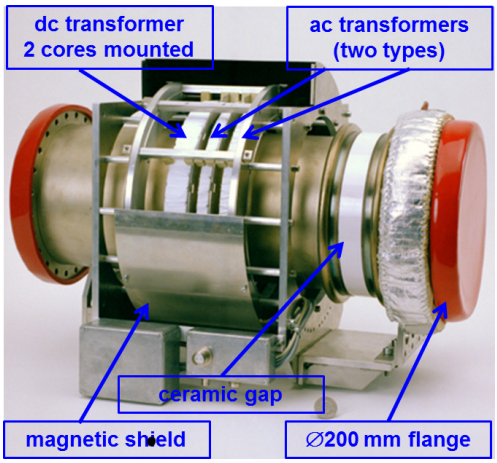


Fig. 12: The d.c. transformer installed at the GSI synchrotron. The left torus is the d.c. transformer (two tori mounted closely together), the middle is the a.c. transformer used for the feedback of the d.c. type and the right one is the a.c. transformer for injection control.

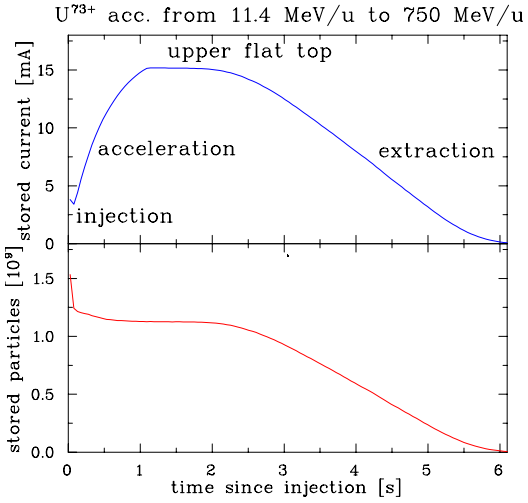


Fig. 13: Signal of a d.c. transformer for a $^{238}U^{73+}$ beam at the GSI synchrotron. The acceleration must be about 1 s. After a delay of about 0.5 s, the slow extraction for 4 s follows.

of the ion and the electron, the energy transfer to the electron per collision is, in most cases, <100 eV. The electronic stopping power $\mathrm{d}E/\mathrm{d}x$ can be approximated by the semi-classical Bethe–Bloch formula, which is written in its simplest form as [16–18]

$$-\frac{\mathrm{d}E}{\mathrm{d}x} = 4\pi N_\mathrm{A} r_\mathrm{e}^2 m_\mathrm{e} c^2 \cdot \frac{Z_\mathrm{t}}{A_\mathrm{t}}\rho \cdot \frac{Z_\mathrm{p}^2}{\beta^2}\left[\ln\frac{2m_\mathrm{e}c^2\gamma^2\beta^2}{I} - \beta^2\right] \tag{13}$$

where N_A is the Avogadro constant, m_e and r_e are the mass and classical radius of an electron, respectively, and c is the velocity of light. The target parameters are ρ, the density of the target with nuclear mass A_t and nuclear charge Z_t; the quantity $Z_\mathrm{t}\rho/A_\mathrm{t}$ corresponds to the electron density. I is the mean ionization potential for removing one electron from a target atom; a rough approximation for a target with nuclear charge Z is $I \simeq Z \times 10$ eV; more precise values are given, e.g., in Refs. [19, 20]. The projectile parameters are Z_p, the nuclear charge of the ion, with velocity β and $\gamma = (1-\beta^2)^{-1/2}$. This formula must be modified because ions travelling through matter are not bare nuclei, but have some inner electrons. An effective charge is used instead of Z_p, calculated by, e.g., semi-empirical methods described for the

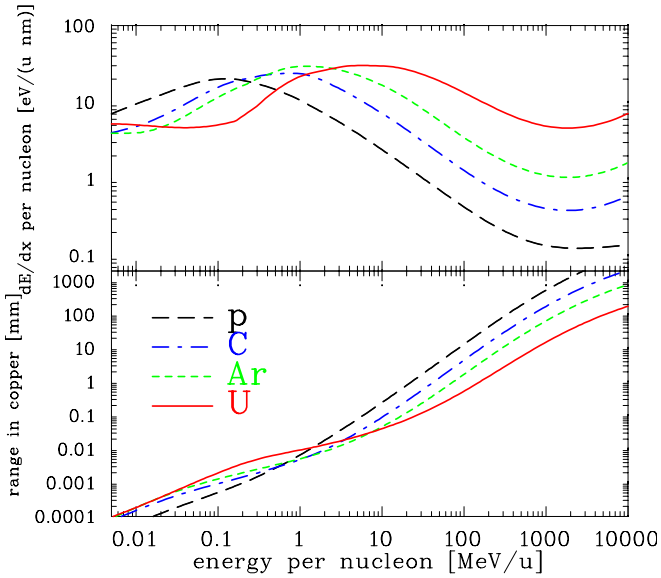


Fig. 14: Energy loss per nucleon at the surface and range in copper as a function of the kinetic energy for several ions. The energy range is plotted from 5 keV/u to 10 GeV/u and the range from 100 nm to 1 m. The calculation of the electronic and nuclear stopping uses the semi-empirical code SRIM [19, 20].

codes SRIM [19, 20] or LISE++ [21]. The result of such a semi-empirical calculation for the energy loss is shown in Fig. 14 for different ions into copper. The energy loss is maximal for ions with a kinetic energy around 100 keV/u to 7 MeV/u (corresponding to velocities $\beta \sim 1.5\%$ to 12%), depending on the ion species. These are typical energies of a proton or heavy-ion linac. Below 100 keV/u, the energy loss decreases and nuclear stopping becomes significant. Energies below 10 keV/u are typical for proton or heavy-ion sources mounted on a high-voltage platform. For relativistic energies above 1 GeV/u, the energy loss is nearly constant; these are typical energies of particles extracted from a synchrotron.

For the consideration of a Faraday cup (see Section 2.4), the range in material is important. Copper is often used for cups, owing to its high heat conductivity. For a particle accelerated to E_{kin}, the range R is calculated numerically from the stopping power via

$$R = \int_0^{E_{\mathrm{kin}}} \left(\frac{\mathrm{d}E}{\mathrm{d}x} \right)^{-1} \mathrm{d}E \tag{14}$$

and has an approximately scaling for ions above $\simeq$10 MeV/u [16] of

$$R \propto E_{\mathrm{kin}}^{1.75} \ . \tag{15}$$

The results are shown in Fig. 14. This range should be shorter than the mechanical dimension for a practical Faraday cup design; for energies below 100 MeV/u, the range is less than 1 cm. Or, in other words, cups are only useful at linacs or cyclotrons. For higher energies, the range is too large and other techniques, e.g., particle detectors, as described in Section 2.5, are applied. In addition, for particles above the nuclear Coulomb barrier, nuclear reactions are possible and charged particles might leave the material, resulting in an incorrect reading.

2.3.2 Material interaction of electrons

The stopping of electrons in matter differs from protons and ions; see Fig. 15, where the sum of the two relevant processes of collisional loss $\mathrm{d}E/\mathrm{d}x|_{\mathrm{col}}$ and radiation loss $\mathrm{d}E/\mathrm{d}x|_{\mathrm{rad}}$ is plotted. The collisional loss for electrons due to electronic stopping $\mathrm{d}E/\mathrm{d}x|_{\mathrm{col}}$ is also described by a modified Bethe–Bloch

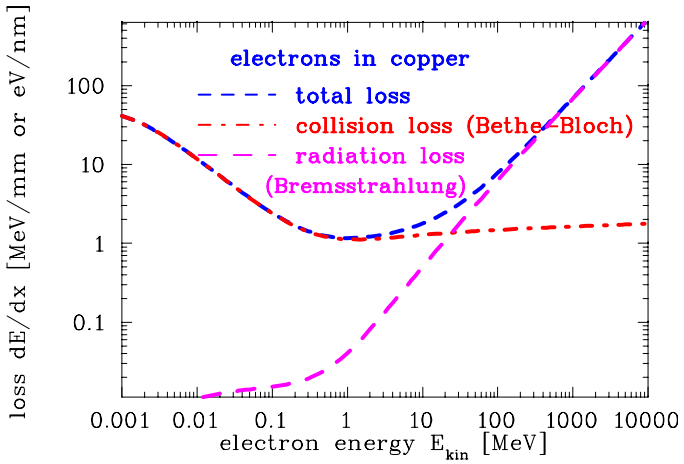


Fig. 15: Energy loss of electrons in copper for collision loss, as given by the Bethe–Bloch equation, and radiation losses, as dominated by the emission of Bremsstrahlung (data from Ref. [22]).

formula; the modification is caused by the equal mass of the projectile and the target electrons and their indistinguishability. This regime dominates only for energies below 1 MeV, as corresponding to a velocity of $\beta = v/c = 94\%$ and a Lorentz factor of $\gamma = (1 - \beta^2)^{-1/2} = 2.96$. Even a small electron accelerator attains larger final energies. For energies above a few tens of megaelectronvolts, the radiation loss by Bremsstrahlung, i.e., the emission of photons through the acceleration or deceleration of electrons in the vicinity of the target nucleus, is dominant, see, e.g., Ref. [16]. This radiation loss scales roughly linear to the electron energy and quadratically to the target charge Z_t as $\mathrm{d}E/\mathrm{d}x|_\mathrm{rad} \propto E \cdot Z_\mathrm{t}^2$. The trajectories of the primary electrons in the target are more curved than for ions, owing to the possible high energy and momentum transfer in a single collision; therefore, electrons have much larger lateral straggling than ions. Moreover, the longitudinal straggling is larger than for ions, resulting in a wider range distribution.

At electron accelerators, Faraday cups are mainly used behind the electron gun for electrons with typically 100 keV energy. In this case, the construction is comparable to the Faraday cups used for proton beams, see Section 2.4. At higher energies, the required amount of material to stop the electrons and to absorb secondary charged particles (e.g., e^+–e^- pair production from the Bremsstrahlung) increases significantly, owing to this long range. An example of a Faraday cup used for high energies up to some gigaelectronvolts, including design considerations, can be found in Ref. [23].

2.3.3 Secondary electron generation

When a charged particle, either an ion or an electron, travels through matter, it interacts with the target electrons, as schematically depicted in the left-hand panel of Fig. 16. In a close primary collision, the target electron can be accelerated to high energies much larger than 100 eV; sometimes these electrons are called δ-rays. The energy distribution depends on the impact parameter and the energy of the incoming particle. In a metallic solid-state material, the fast δ-electrons collide with the surrounding electrons; because both collision partners have equal masses, the energy is transferred efficiently and more electrons are liberated, see Fig. 16. This so-called thermalization of the electrons inside the conduction band is performed in a typical scattering length of $L_\mathrm{s} \simeq 10$ nm, i.e., within the distance of some 100 lattice planes. Owing to these many collisions, there is only a weak correlation between the direction of the incoming particle and the scattered electrons. If these electrons reach the surface and still have an energy greater than the work function (typically 1 to 5 eV), they have a high probability of escaping from the metal. The current of the secondary electrons is given by the so-called Sternglass formula [24]

$$I_\mathrm{sec} = Y \cdot \frac{\mathrm{d}E}{\rho \mathrm{d}x} \cdot I_\mathrm{beam} \tag{16}$$

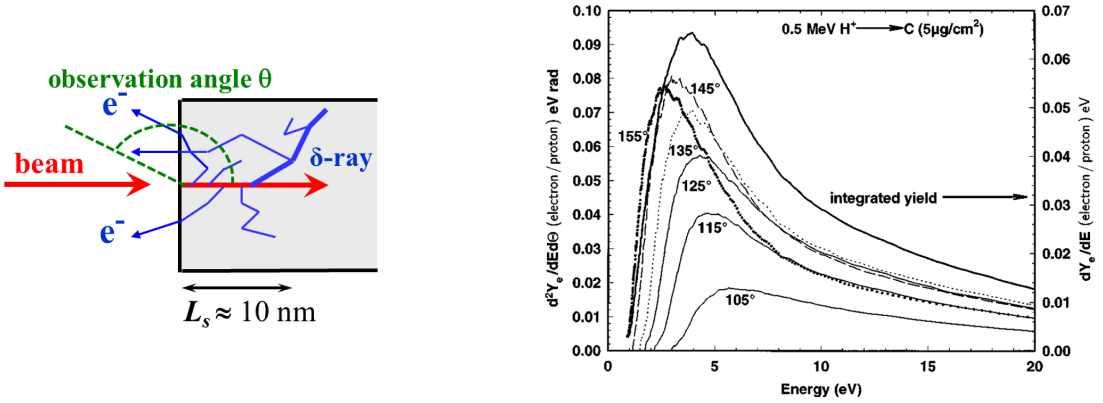


Fig. 16: Left: Secondary electron generation from a surface by charged particle impact; L_s is the scattering length from which electrons are able to reach the surface. Right: Typical measurement of electron energy spectrum for different emission angles and total yield in backward direction by 0.5 MeV proton impact on a carbon foil [25].

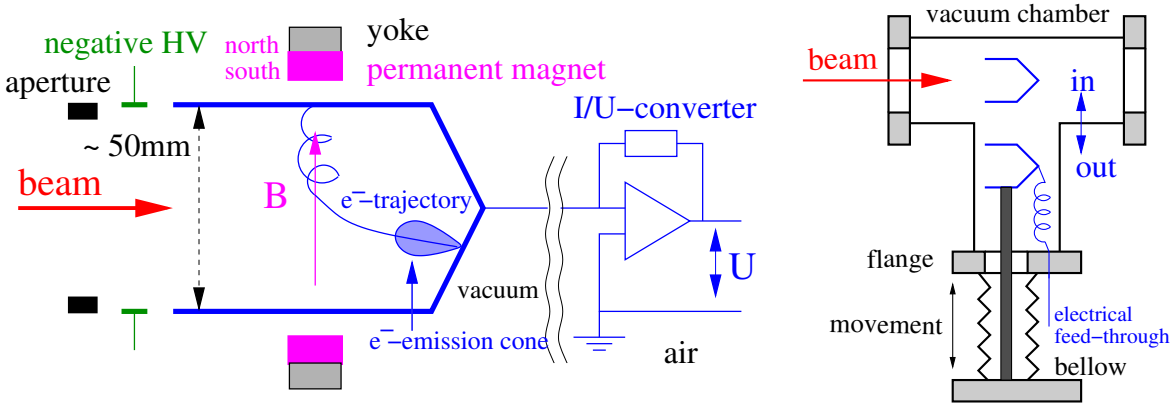


Fig. 17: Left: Scheme of an uncooled Faraday cup with magnetic and electric secondary electron suppression. Right: Scheme for inside and outside position of destructive diagnostics devices.

where I_{beam} is the current of incoming particles and Y is the yield factor describing the amount of secondary emission per unit of energy loss $\mathrm{d}E/\rho\mathrm{d}x$ at the surface of the metal. The yield Y depends on the metal (via the work function) and the surface quality, e.g., modified by possible adsorbed gas contamination. The mean kinetic energy of the escaping secondary electrons is of the order of 10 eV and has some dependence on the emission angle; an example for the energy spectrum is shown in Fig. 16 and discussed in more detail in Ref. [25]. To first order, the angular distribution $P(\theta)$ of the electron emission can be approximated by a $P(\theta) \propto \cos\theta$ law, where θ is the angle of the trajectory with respect to the surface ($\theta = 180^\circ$ indicates back scattering).

2.4 Faraday cups for low currents

A Faraday cup is a beam stopper that is designed to measure the electrical current of the beam. The basic cup design is shown in Fig. 17 and a photograph is given in Fig. 18. An isolated metal cup is connected to a current-sensitive pre-amplifier. As shown for an active beam transformer, the pre-amplifier consists of a low-impedance input and a conversion to a voltage. Range switching is achieved by using different feedback resistors for the operational amplifier. Much smaller currents can be measured with a Faraday cup than with a transformer: a measurement of 10 pA for a d.c. beam is possible with a low-noise current-to-voltage amplifier and careful mechanical design; this is five orders of magnitude more sensitive than a d.c. transformer. Low-current measurement is important, e.g., for the acceleration of radioactive beams.

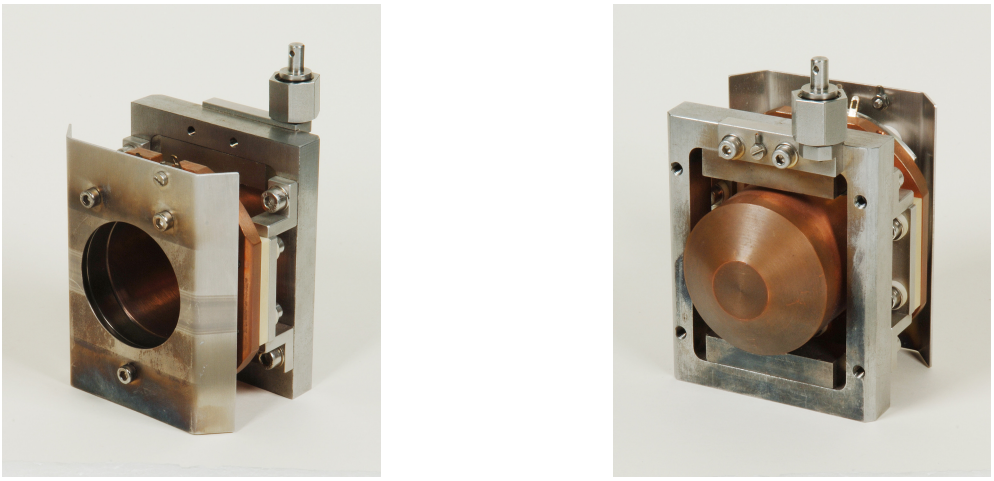


Fig. 18: Uncooled Faraday cup (⌀50 mm) with magnetic and electric secondary electron suppression

When an accelerated particle hits a surface, secondary electrons are liberated, see also Section 2.3. The flux of these electrons is proportional to $\cos\theta$, where θ is the angle of the electron trajectory with respect to the surface; their average energy is less than $\simeq$10 eV. If these electrons leave the insulated cup, the reading of the beam current is wrong by this amount. A secondary electron suppression has to be foreseen. It can be realized as follows.

- By using a high-voltage suppression close to the entrance of the cup. By applying a voltage well above the mean energy of the secondary electrons, they are pushed back to the cup surface, see the left-hand panel of Fig. 19. The disadvantage of this method is related to the fact that the electrical field on the beam axis is smaller than that at the edges; in the depicted case, the maximum potential is about 35% of the potential applied to the electrode. The emission of the maximal energetic electrons occurred opposite to the beam direction and the potential must be chosen to be greater than the kinetic energy of at least 99% of the electrons. A measurement with varying suppression voltage is shown in the right-hand panel of Fig. 19.
- By using a magnetic field created by permanent magnets. In this field B, the secondary electrons spiral around the magnetic field lines with the cyclotron radius r_c

$$r_\mathrm{c} = \frac{\sqrt{2 m_\mathrm{e} E_\mathrm{kin}}}{eB} \tag{17}$$

  where m_e is the electron's mass, e is its charge, and E_kin is the kinetic energy component perpendicular to the field lines. For $E_\mathrm{kin} = 10$ V and a field of 10 mT, $r_\mathrm{c} \sim 1$ mm. With permanent magnets, field lines perpendicular to the beam axis can be created relatively easily, see Fig. 20.

2.5 Low-current measurement used for slow extraction

Many experiments in particle, atomic, or nuclear physics use protons or ions in the energy range as reached in synchrotrons, i.e., higher than some 100 MeV/u. A typical beam current during the acceleration inside the synchrotron followed by a slow extraction recorded by a d.c. transformer is shown in Fig. 13. Typical beam currents, as slowly extracted from a synchrotron range, from only 10^3 to 10^{12} particles per second, which corresponds to an electrical current from 10^{-15} to 10^{-6} A. This is well below the resolution of a d.c. transformer. Owing to the high energy, the range of the particles is too large for the use of Faraday cups, as shown in Fig. 14 for copper. In addition, a lot of secondary charged particles would be created in the cup material. The techniques of particle detectors are used instead. An overview of the typical dynamic range of detectors is presented in Fig. 21.

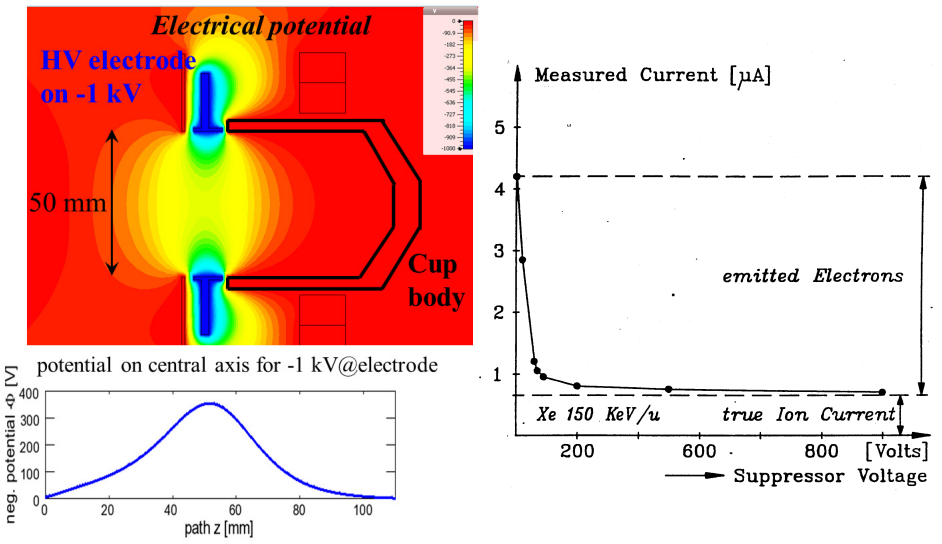


Fig. 19: Left: The electric potential of the Faraday cup shown in Fig. 18 with the high-voltage electrode biased by -1 kV is shown at the cylindrical symmetry plane. The beam enters from the left. The negative potential along the central axis is depicted at the bottom. Right: The effect of secondary electron suppression inside a Faraday cup as determined by the beam current determination as a function of applied voltage.

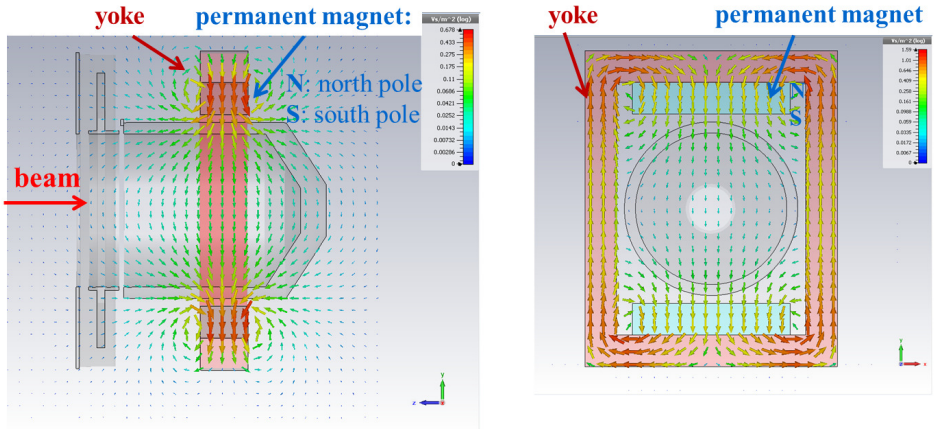


Fig. 20: Magnetic field lines of the arrangement of Co-Sm permanent magnets within the yoke for the Faraday cup of Fig. 18. The homogeneous field strength is $B \sim 60$ mT.

– For an ion rate below 10^6 s^{-1}, the individual particles can be counted by scintillators.
– For the medium range, from about 10^4 to 10^9 s^{-1}, the energy loss in a gas is measured by an ionization chamber.
– For the higher range, from about 10^8 s^{-1}, the emission of secondary electrons from a metal surface forced by the primary ion's energy loss is determined by secondary electron multipliers (SEMs).

The design of such a system and the combination of the three detector is described in, e.g., Refs. [26–28].

2.5.1 Scintillation counter

When a charged particle penetrates a scintillating material, the electronic energy lost by the collision of the beam particles with the target electrons creates atomic excitation; the excited atoms might then decay via emission of fluorescence photons. These photons can be detected and amplified by a photomultiplier,

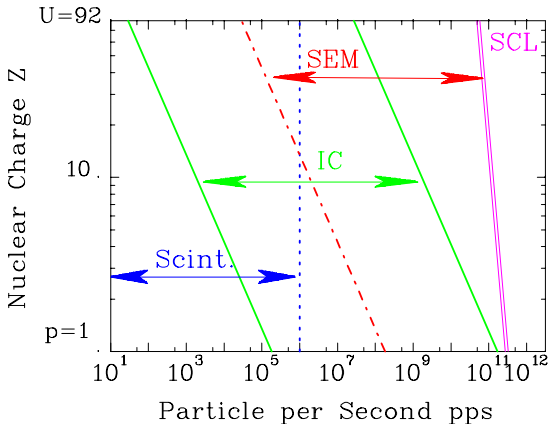


Fig. 21: Overview of the different detector systems used for slow extraction at the GSI synchrotron. The numbers are given for the different ions with a kinetic energy of 1 GeV/u, extracted within 1 s, and a spot size of 1 cm^2. With a scintillator, up to 10^6 particles can be counted. The lower limit of the ionization chamber and SEM correspond to a secondary current of 1 pA, the upper threshold of the ionization chamber is 30 Gy/s. SCL: incoherent space charge limit of the synchrotron at injection energy.

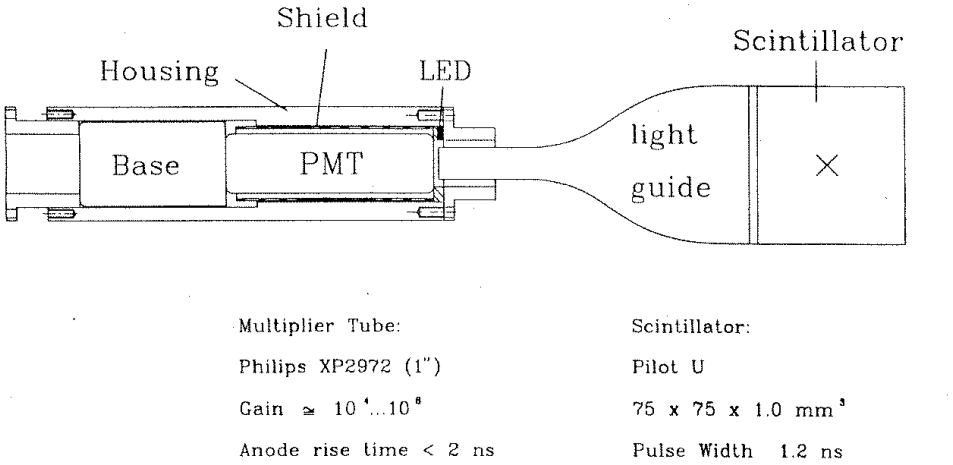


Fig. 22: Plastic scintillation counter connected to photomultiplier via plastic light guide

converted to a logical pulse by a discriminator and finally counted by a scaler; for more details see Refs. [16, 17].

In most cases, plastic scintillators are used. They are cheap and easy to produce in nearly every mechanical shape. As an example, the arrangement used at GSI is shown in Fig. 22 and a photograph, together with an ionization chamber, is shown in Fig. 23. A large size of 75 mm $\times$ 75 mm has been chosen with a thickness in the beam direction of 1 mm. The plastic scintillator (in this case BC 400) has a short decay time in the range of some nanoseconds, see, e.g., Ref. [17]. Typical pulses are displayed in Fig. 24, recorded with a medium-energy heavy-ion beam. The energy resolution is quite good, which is important because it yields good accuracy of this absolute count rate measurement. Owing to the long cables, several hundreds of metres long, used between the detector in the beam pipe and the electronic processing, cable dispersion broadens the peak. Before pile-ups start to play a role, count rates up to several 10^7 s^{-1} can be accepted. A typical measurement of an extracted current is shown in Fig. 25, where the particle rate as determined by a scintillator is compared with the particle rate as determined by an ionization chamber. This is corroborated with the signal of the d.c. transformer measuring the stored current in the synchrotron.

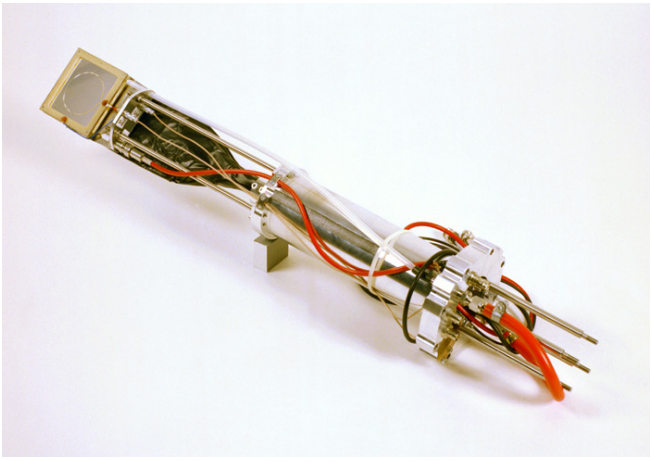


Fig. 23: The hardware of the scintillators (wrapped in black tape) and the ionization chamber installed at GSI

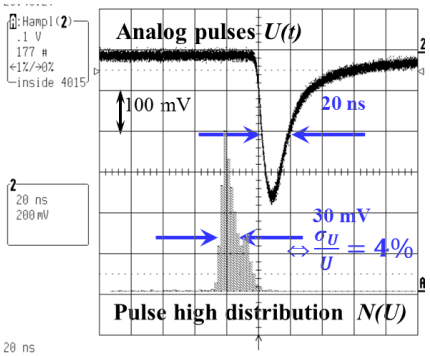


Fig. 24: Analogue pulses from a plastic scintillator recorded with a low-current 300 MeV/u Kr beam; the scaling is 20 ns per division and 100 mV per division. The lower curve is the pulse height distribution.

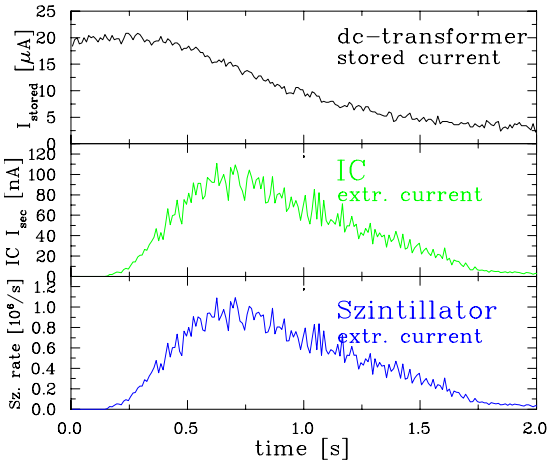


Fig. 25: Extracted current measured by an ionization chamber (middle, given in nanoamps for the secondary current) and a scintillator (bottom, given in particles per second) as compared with the current inside the GSI synchrotron determined by a d.c. transformer (top, given in microamps) for one extraction of a 250 MeV/u $^{208}Pb^{67+}$ beam with a total amount of 10^6 particles.

Table 4: Ionization potential and W values

Gas	H_2	He	N_2	O_2	Ar	CH_4	CO_2
Ionization potential [eV]	15.6	24.5	15.5	12.5	15.7	14.5	13.7
W value [eV]	36.4	42.7	36.4	32.2	26.3	29.1	33.0

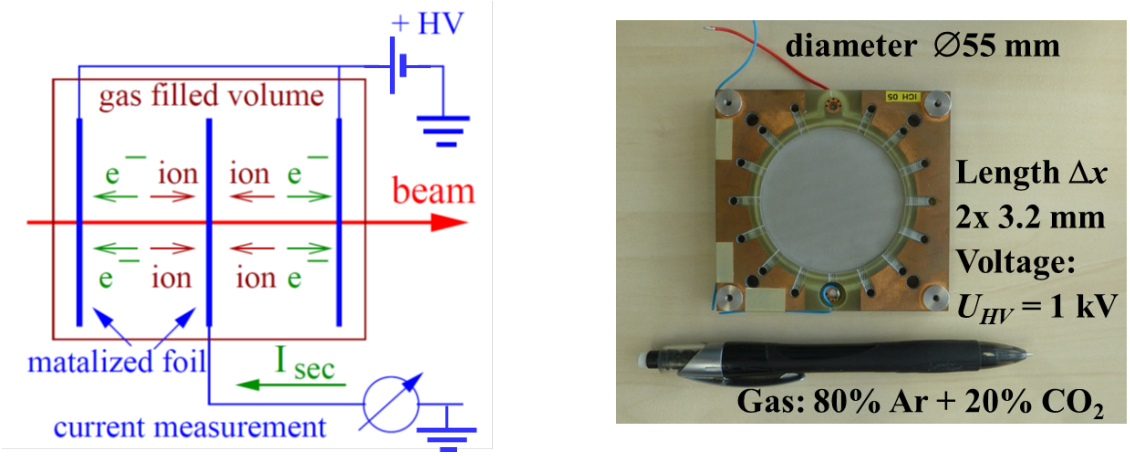


Fig. 26: Left: Scheme of an ionization chamber. Right: Its realization, with a 2 mm $\times$ 3.2 mm active length in the beam direction.

The disadvantage of the plastic scintillators is the very low radiation hardness, owing to their complex organic molecular structure. Inorganic crystals have a much higher radiation hardness. In most cases, Ce-activated materials are used [17, 29]. However, these inorganic single-crystal scintillators are difficult to produce in larger sizes.

2.5.2 Ionization chamber

For the medium particle range, the current is determined by the energy loss of charged particles in a gas inside an ionization chamber. A general description of ionization chambers can be found in Refs. [17,30]. Owing to the large statistics, the average energy for the production of an electron–ion pair is a well-known number, the so-called W values are given in Table 4. The arrangement of an ionization chamber is displayed in Fig. 26 and a photograph is displayed in Fig. 27. For this realization, the beam passes a 6.4 mm thick active gas volume and creates electron–ion pairs. In most cases, pure Ar is used, sometimes also mixed with about 10% of molecular gases, such as CH_4 or CO_2. The active gas volume is confined by metallized plastic foils or networks. These metallized electrodes are biased with about 1 kV, to separate the charges. The middle electrode measures the amount of secondary charge using a sensitive current amplifier. A secondary current down to the picoamp region can be measured precisely, see, e.g., Ref. [31]. The ionization chamber, filled with gas at atmospheric pressure, is separated from the vacuum by a metallic foil of typically 100 µm thickness. The foil thickness must be low to prevent for significant energy loss in this foil. The specifications of the ionization chamber at GSI are given in Table 5.

With the help of the W values and the calculated energy loss $\mathrm{d}E/\mathrm{d}x$ within an active length Δx, the number of primary ions I_{beam} is determined from the secondary current I_{sec}, as

$$I_{\mathrm{sec}} = \frac{1}{W} \cdot \frac{\mathrm{d}E}{\mathrm{d}x} \Delta x \cdot I_{\mathrm{beam}} \quad . \tag{18}$$

The precision of such a measurement depends mainly on the accuracy of the energy-loss calculation. For energies above 1 GeV/u, the value of $\mathrm{d}E/\mathrm{d}x$ is only weakly dependent on the energy E; see Fig. 14. When the energies of the beam particles are less than this, calibrations have to be made to reach an

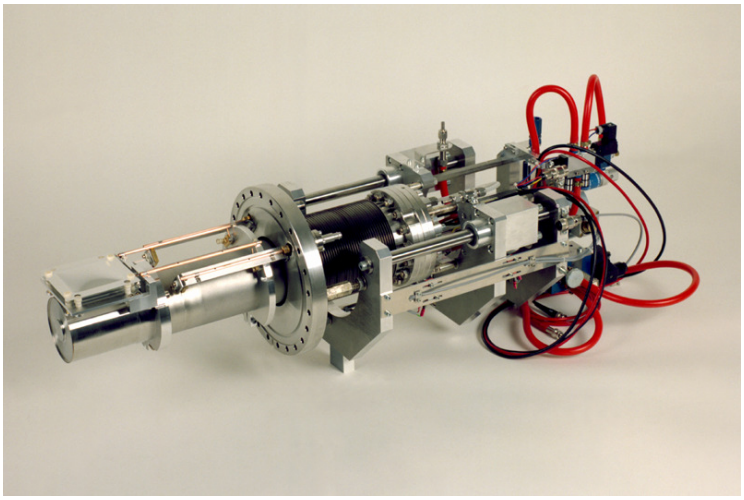


Fig. 27: Pneumatic drive mounted on a ⌀200 mm flange containing a scintillator and ionization chamber inside a pocket at atmospheric pressure and an SEM in a vacuum.

Table 5: Specifications of the ionization chamber at GSI

Parameter	Value
Active surface	⌀55 mm
Active length	2 mm × 3.2 mm
Electrode material	Plastic network
Coating	100 µg/cm^2 silver
Gas (flowing)	80 % Ar + 20 %CO_2
Pressure	1 bar
Voltage	500–2000 V

accuracy in the range of 1%. The ionization chamber can be calibrated with respect to a scintillator, where an absolute measurement is performed, as shown in Fig. 25. From the number of counted particles and the secondary current of the ionization chamber, a calibration factor can be determined, see, e.g., Ref. [28].

The lower limit of an ionization chamber is given by the minimum measurable secondary current, which is about 1 pA. Taking this current, the line in Fig. 21 is determined. The upper limit is based on the recombination of the electrons with the ions from the gas—for the case of Ar working gas, the reaction is $\mathrm{Ar}^+ + \mathrm{e}^- \rightarrow \mathrm{Ar}$—which is a sink for the generated charges. This is demonstrated in Fig. 28. We first discuss the characteristic line of the ionization chamber with a moderate primary current. The proportional region is reached after applying a moderate voltage of 500 V, corresponding to 1 kV/cm. The output is then independent of the voltage (or the applied electric field). The characteristic line is determined by comparison with the SEM (see next section) to obtain the efficiency. For a higher primary current, the plateau is no longer reached. A higher voltage cannot be applied without the risk of sparks. The drift time of the ions towards the electrodes depends on the applied electric field, while the drift time for the electrons is nearly constant for the given parameters [17]. The collision frequency of electrons and ions is proportional to their concentration n_e and n_i. The recombination rate is

$$\frac{\mathrm{d}n_\mathrm{e}}{\mathrm{d}t} = \frac{\mathrm{d}n_\mathrm{i}}{\mathrm{d}t} = \alpha \cdot n_\mathrm{e} n_\mathrm{i} \propto I_\mathrm{beam}^2 \tag{19}$$

where α is the rate coefficient. The loss rate is therefore proportional to the square of the primary beam current. Typical values are summarized in Table 6. One can conclude that the secondary current for the

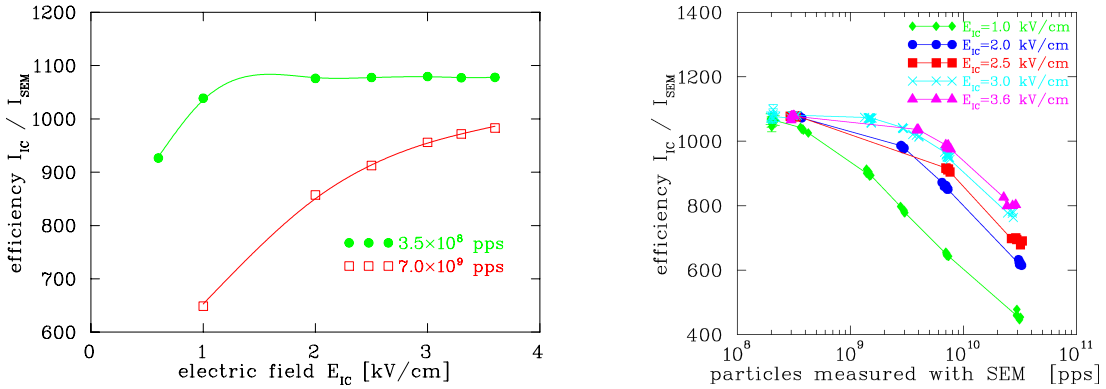


Fig. 28: Left: Characteristic line for ionization chamber, as determined for the 300 MeV/u Ne^{10+} beam. The efficiency is measured by the ratio of the secondary current of the ionization chamber and the SEM as a function of the applied electric field [27]. For a low primary current, a plateau is reached; for a high current, the recombination leads to a decrease of the secondary charges. Right: Comparable data shown as a function of transmitted ions per second for different applied electric fields.

Table 6: Properties of the ionization chamber close to the saturation point, see text

Parameter	Ions	Electrons
Drift velocity, w^{drift}	0.1 mm/ µs	20 mm/ µs
Drift time, t^{drift}	50 µs	0.1 µs
Saturation density, n^{sat}	2×10^9 cm^{-3}	4×10^6 cm^{-3}
Secondary current, I_{IC}	1 µA	
Dose rate, D_{IC}	30 Gy/s	
Average rate coefficient, α	10^{-9} cm^3/s	

used ionization chamber should be below ∼1 µA for typical parameters, corresponding to a dose rate of about 30 Gy/s [27]. To obtain the value of the primary beam current, the conversion from the energy loss as given in Eq. 18 must be used.

To get a large upper threshold, the use of gases with high electron affinity, such as O_2 and H_2O, must be avoided [17]. The electron affinity describes the probability of an electron capture, e.g., $O_2 + e^- \rightarrow O_2^-$. The presence of negative ions enhances the recombination with positive ions, e.g., $O_2^- + O^+ \rightarrow O_3$, resulting in a smaller amount of measured secondary charge. In particular, an ionization chamber for high count rate applications should not contain air.

An ionization chamber can serve as a permanently installed monitor in front of the target for online beam current determination. The output signal from this ionization chamber can serve as the input for beam current control via a feedback circuit to control the amount of slowly extracted current. In particular, in a medical synchrotron delivering ions to a patient, the control of beam current by varying the amplitude of the transverse excitation power for a slow knock-out extraction has been realized [32–34].

2.5.3 Secondary electron multiplier

To cover the large current range from, e.g., slow extraction from a synchrotron, the emission of secondary electrons from a metallic surface can be determined using a secondary electron multiplier (SEM). Such a set-up is shown in Fig. 29, where three 100 µm thick aluminium foils are visible, installed on the vacuum side of the feed-through. The specifications of the SEM used at GSI are given in Table 7. The outer two foils are biased by typically +100 V to sweep out the free electrons. The middle foil is connected to a sensitive current amplifier. The secondary emission current depends on the energy loss at the surface,

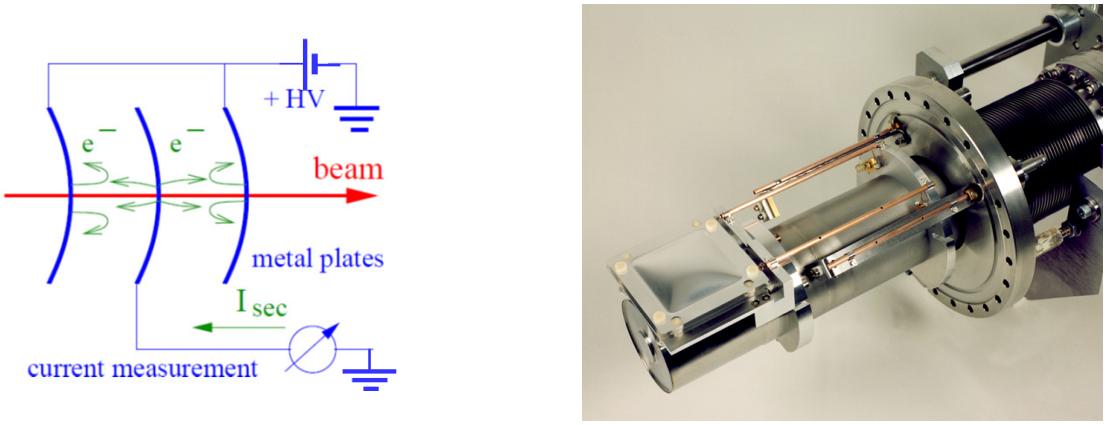


Fig. 29: Left: Scheme of a scanning electron monitor (SEM). Right: Photograph of the SEM part, made of three 100 µm thick aluminium foils, of the feed-through used for slow extraction current measurements at GSI.

Table 7: Specification of the SEM at GSI

Parameter	Value
Material	Pure Al (≃99.5%)
Thickness	100 µm
Number of electrodes	3
Active surface	80×80 mm^2
Distance between electrodes	5 mm
Voltage	100 V

$\mathrm{d}E/\rho\mathrm{d}x$, and is given by the Sternglass formula [24]

$$I_{\mathrm{sec}} = Y \cdot \frac{\mathrm{d}E}{\rho \mathrm{d}x} \cdot I_{\mathrm{beam}} \,,$$

where Y is the yield factor describing the amount of secondary emission per unit of energy loss at the surface of the aluminium foil, see Eq. 16. Experimental calibration is required to determine the value of Y, yielding a precision of 5%, see, e.g., Ref. [28]. The accuracy is less than for an ionization chamber, because the actual value of the yield is determined by the surface structure of the material, which might vary, depending on the production and cleaning method. Equation 16 reflects the fact that the only electrons that are knocked out are those that are excited to the conducting band close to the surface of roughly ≃10 nm; see Section 2.3.3 for a discussion of secondary electron emission. The physics of the secondary electron emission is also used for profile measurements, see Section 4.2.

Aluminium is often used because of its well-suited mechanical properties. In particular, it is stiff enough to prevent microphonic pick-up by external vibrations caused, e.g., by turbo vacuum pumps. However, the emission yield Y might change with irradiation. This is caused by modifications of the surface, which changes the work function and therefore the average yield Y. Measurements have shown a significant degradation, by a factor of two, after 10^{18} proton/cm^2 with 450 GeV protons at CERN SPS [35]. Here Ti foils have a much lower sensitivity to radiation. Radiation hardness is important, because these devices are often used in front of the target as a permanent current monitor.

3 Beam loss detection

In a real accelerator, the transmission from the source to the target is never 100%. The fraction of beam particles lost must be controlled carefully to achieve optimal transmission. The lost beam particles cause some activation of the accelerator components by nuclear reactions. Moreover, the surrounding material

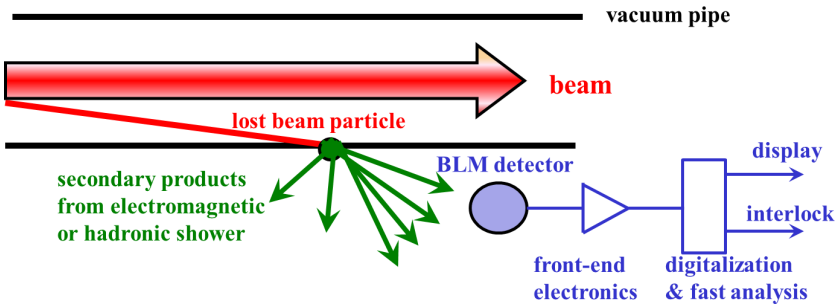


Fig. 30: Scheme for the installation of beam loss monitors outside a vacuum pipe

can be destroyed by the radiation, as well as by the heating due to the particles' energy loss. To detect the shower of secondary particles as originated by nuclear interaction processes, a large variety of beam loss monitors exists. Most of these are particle counters, detecting secondary reaction products. At nearly every high-current accelerator facility, these monitors are installed for the protection of the accelerator components. The relatively cheap beam loss monitor instruments are mounted outside of the vacuum pipe at crucial locations, as depicted schematically in Fig. 30. Their signals give important information to prevent unwanted loss during the operation, e.g., caused by malfunctions of components. A careful analysis of the location and time structure of possible losses must be performed before suitable types of beam loss monitors can be chosen; see Refs. [36–39] for reviews. Beam loss monitors can be sensitive devices, which detect even a very small fraction of losses; therefore, they are also used for machine tuning. Two examples are presented in Section 7.

3.1 Secondary particle production

When a high-energy particle hits a vacuum pipe, or any other material, secondary particles are generated. The relevant processes are described, e.g., in Refs. [40, 41]. Here only a brief overview is given.

- *Interaction of electrons.* For electron energies above ∼100 MeV, Bremsstrahlung dominates the slow-down process in materials, as shown in Fig. 15. The created high-energy γ-photons give rise to further particles via e^+–e^- pair production. If the energies of the γ-photons are high enough, other particles, e.g., $\mu^{\pm}$, $\pi^{\pm}$, ..., can also be produced, an electromagnetic shower is generated. Moreover, the nucleus can be excited to so-called giant resonances. This is a collective nuclear excitation, in which the neutrons oscillate against the protons. The dipole mode of these giant resonances has a threshold of about 6 MeV for typical materials. The de-excitation proceeds with high probability via neutron emission as a (γ, n) reaction. For higher energies, (γ, p) and (γ, np) channels are also open. When the electron is slowed down to below ∼10 MeV, ionization loss by electronic stopping dominates, i.e., an electron vacancy is created at the atomic shell, which is then filled by an outer electron, leading to a soft X-ray emission. Those X-ray photons are absorbed within a short distance, from millimetres to centimetres in typical metals.
- *Interaction of protons.* Beside electronic stopping, as shown in Fig. 14, nuclear interactions are possible. First, we define the term 'thick target', when the interaction length is comparable to the range as given by the electronic stopping. In these thick targets, the probability of a nuclear reaction rises to nearly 100% for energies above 1 GeV, see Fig. 31. Most of the final channels of these nuclear reactions include neutron emission. As an example, for the stopping of 1 GeV protons in copper or iron, ∼10 fast neutrons are liberated. The neutron yield scales approximately with E_{kin} for energies above 1 GeV. In addition, hadron showers (by the strong interaction at the quark level) are possible, resulting in various species of 'elementary' particles.
- *Interaction of ions.* Beside the interactions also valid for the protons, an ionic projectile can be induced by fragmented or fission, leading to radioactive nuclei and fast proton and neutron emission.

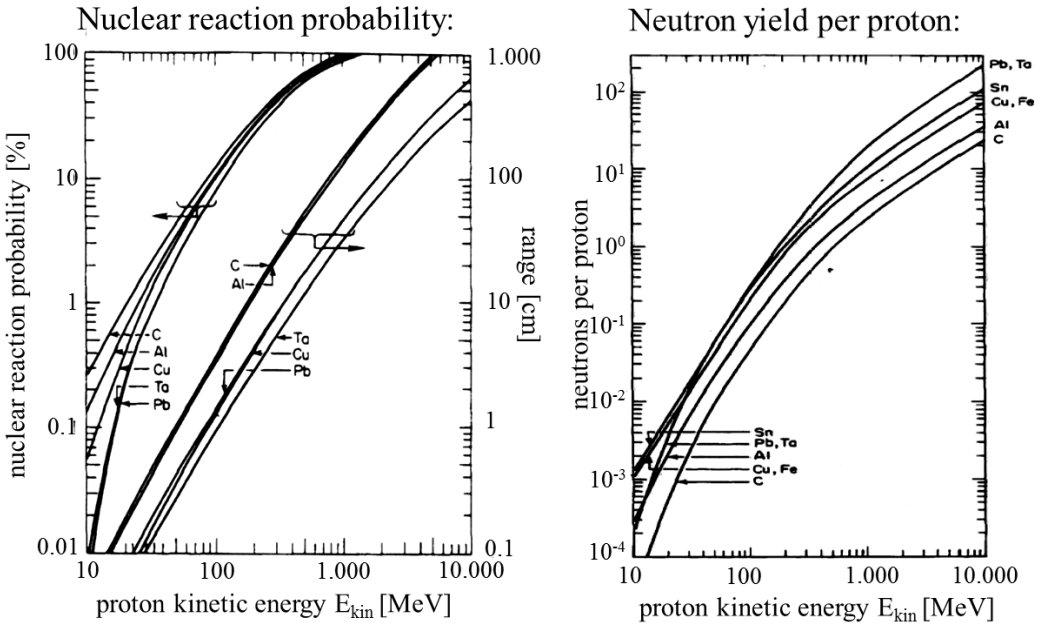


Fig. 31: Left: Ranges of protons in various metals and the probability of an inelastic nuclear reaction within the range as a function of the proton energy, i.e., within a thick target. Right: Total number of neutrons per proton incident on a thick target of various metals as a function of the proton energy [41].

Simple scaling laws do not exist, owing to the complex nuclear physics involved.

Common to all interactions is the production of radioactive nuclei, leading to activation of the accelerator components. The emitted charged particles are stopped relatively quickly by the surrounding material but the neutrons, produced by most primary interactions, can travel long distances. Some beam loss monitors are therefore sensitive to these neutrons. Except for the production of radioactive nuclei, all processes are fast, compared with the timescale of interest in accelerator physics, i.e., faster than $\sim$10 ns. In this sense, a beam loss monitor reacts promptly to particle loss. Owing to the kinematics of the primary interaction, the secondary particles are emitted in a (more or less) forwardly peaked angular distribution. This leads to a spatial resolution of the loss detection by the monitor position close to the loss point.

3.2 Types of beam loss monitor

The task for beam loss monitors is to determine the position and time of the loss. Their signals should be proportional to the amount of loss at this position. A high sensitivity is needed to measure small losses. A high dynamic range is required to deal with a sudden loss of a sizeable fraction of the beam. Depending on the application, a bunch-to-bunch resolution on a $\sim$10 ns time scale is needed, as well as up to 100 ms for a slow detection. All loss monitors are installed outside of the vacuum pipe, detecting mostly secondary particles. These can be neutrons, which are not much absorbed by the surrounding material, charged particles, such as protons, e^- or e^+, or γ-rays. Figure 32 shows some frequently used types of beam loss monitor, as tested at the GSI synchrotron. An overview of the different types is given in Ref. [39].

3.2.1 Plastic scintillators

Plastic scintillators detect charged particles through electronic stopping, as discussed in Section 2.5.1. γ-rays are detected as they liberate electrons from the molecules via the photo effect or Compton scattering; those electrons generate optical photons through electronic stopping. Moreover, plastic scintillators are also sensitive to neutrons, owing to their elastic scattering on the hydrogen atoms of the polymers [17]. Owing to the elastic scattering process generated by the relativistic neutrons, a large momentum transfer

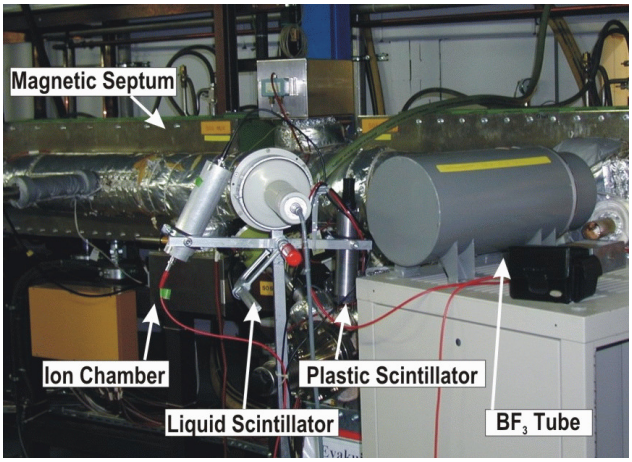


Fig. 32: Tested beam loss detectors installed at the extraction from the GSI heavy-ion synchrotron [42]

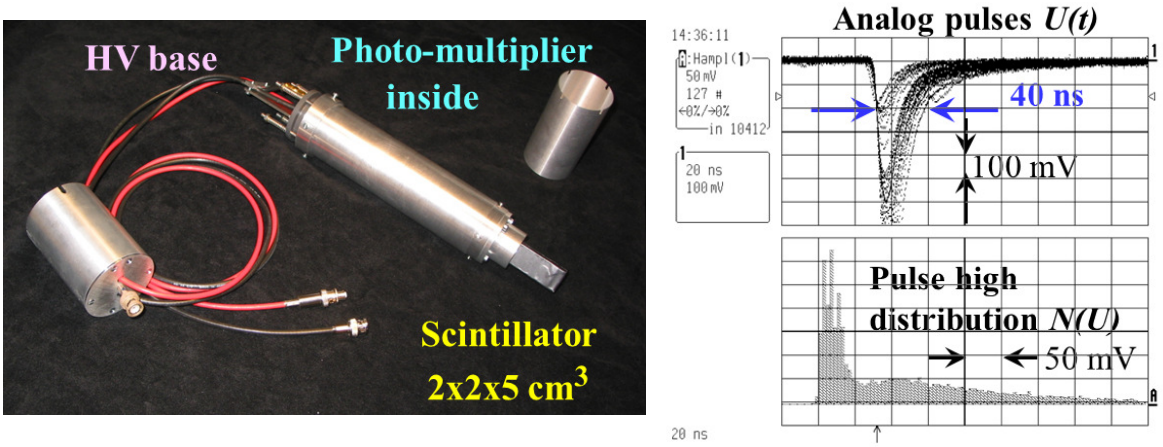


Fig. 33: Left: Typical beam loss monitor based on a plastic scintillator connected to a photomultiplier. Right: Typical pulses from a plastic scintillator (top, 100 mV/division and 20 ns/division) and their pulse height distribution (bottom, 50 mV/division), as measured at the GSI heavy-ion synchrotron.

to the hydrogen atoms of the macromolecules is probable, which leads to a fast proton travelling through the scintillator material and finally generating light emission through electronic stopping. The light is guided to a photomultiplier, converted to electrons, and amplified; see Fig. 33 for a typical realization. The resulting electronic signal can be read out by digitizing the analogue voltage using a relatively small photomultiplier gain, or a particle-counting mode can be used for higher sensitivity and larger dynamic range. A typical output for the counting mode is shown in Fig. 33, which shows analogue pulses of $\sim$10 ns duration. Owing to the wide spectrum of incoming particle species and related energies, the pulse height distribution is quite broad. (This should be compared with the narrow pulse height distribution of single energetic particles shown in Fig. 24.) In the counting mode, the plastic scintillators have a very large dynamic range, from single-particle counting up to a count rate of 10^7 s^{-1}. In most cases, the scintillation monitors are located in crucial areas, such as injection or extraction devices, or close to scrapers. The disadvantage is the low radiation hardness of the plastic materials, owing to the complex chemical composition of the polymers from the plastic matrix. A very approximate number of the destruction threshold is given by the dose $D = 1$ Mrad = 10^4 Gy. where the light transmission is reduced by a factor $1/\mathrm{e} = 0.37$. Typical plastic scintillators are cylindrical rods of 2 cm diameter and 10 cm length. Sometimes, long fibres of plastic or inorganic materials are used.

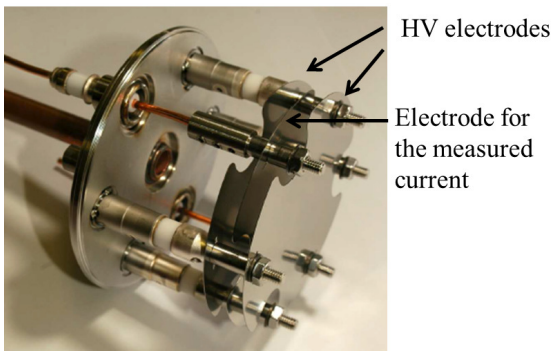


Fig. 34: Three plate SEM beam loss monitor [43, 44]

3.2.2 Liquid scintillator

A scintillator material within a liquid solvent of about one litre is used [17]. The device is also sensitive to charged particles, $e^{\pm}$, γ-rays and neutrons through the same mechanisms as plastic scintillators. A comparable pulse height distribution is observed. The pulse length of the scintillation process is about one order of magnitude longer; therefore, the dynamic range is a factor of ten less for the counting mode. Owing to the microscopic light creation, γ-rays and neutrons can be discriminated using special electronics and so-called pulse shape discrimination [17]. The liquid is about a factor of ten more radiation hard than the plastics, owing to the simpler chemical composition of the solvent.

3.2.3 Secondary electron multiplier and secondary electron multiplier

An aluminium-coated electron multiplier is an electron multiplier in which the photocathode is replaced by an aluminium surface. Secondary electrons are liberated by the interaction with the radiation and amplified by the dynodes. This device offers a fast rise time and high dynamic range, but the sensitivity is much lower than that of scintillators.

If the loss rate is very large, the current of secondary electrons from a surface is large enough to be measured directly. This can be realized using a secondary electron multiplier with three plates installed in a small vacuum vessel [43,44]. The outer plates are biased by some +100 V and the secondary electron current emitted from the central plate is measured by a transimpedance amplifier or a current-to-frequency converter. The inner part of such a monitor is shown in Fig. 34. Relevant simulations and tests of the response to different radiation are reported in Refs. [43, 44].

3.2.4 PIN diode

Solid-state detectors can be used as particle counters. In the typically 100 µm thick depletion layer of the doped Si-crystal, electron–hole pairs are generated. For a minimum ionizing particle, say, a charged proton at the minimum of its electronic stopping power, about 10^4 electron–hole pairs are generated. This signal must be amplified by a charge-sensitive pre-amplifier and can then be counted. As depicted in Fig. 35, two face-to-face mounted PIN diodes are used in coincidence to count only the charged particles and not the low-energy photons, as emitted at electron accelerators by synchrotron radiation. These photons are absorbed in one PIN diode and do not arrive at the second. The detection efficiency is relatively low, owing to the small active area, typically 10×10 mm^2, and the 0.1 mm depletion depth. A device made of two PIN diodes including the amplifier and the counter is commercially available [11]; these are often installed at synchrotron light sources.

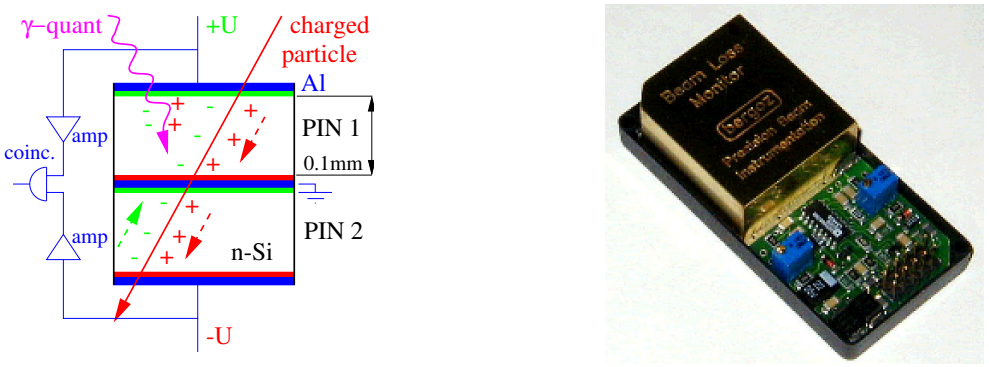


Fig. 35: Scheme of two PIN diodes driven in coincidence and photograph of the device [11]

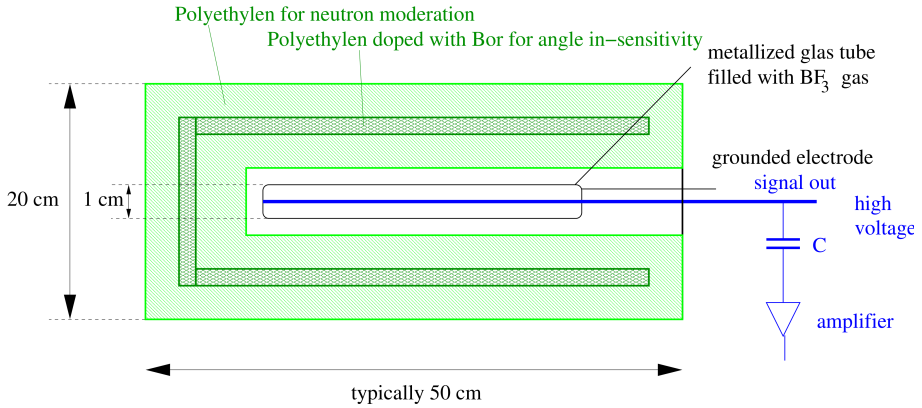


Fig. 36: BF_3 proportional tube as used for beam loss detection. The proportional tube is filled with BF_3 gas surrounded by polyethylene for neutron moderation.

3.2.5 BF_3 proportional tube

To have the possibility of measuring only neutrons, a cylindrical proportional tube (typical diameter, 10 mm and length 500 mm) filled with BF_3 gas can be used [17], as shown in Fig. 36. For thermal neutrons, the reaction $^{10}\mathrm{B} + \mathrm{n} \rightarrow {}^{7}\mathrm{Li} + \alpha$ has a high cross-section ($\sim$ 1 kbarn) and is exothermic by a Q value (total kinetic energies of Li and α) of 2.3 MeV. The gas-filled tube is surrounded by concentric layers of polyethylene, with an outer diameter of $\sim$200 mm. The neutrons from the primary beam interaction are slowed down by elastic collisions with the hydrogen nuclei of the polyethylene. Special precautions are taken to obtain a flat detection efficiency as a function of the angle. Normally, these detectors are used at nuclear power plants for neutrons with energies of up to 10 MeV, but the thermalization yield of the neutrons can be extrapolated at least up to 100 MeV. In the traditional application, the glass tube is operated as a proportional tube; the count rate is limited to several 10^4 s^{-1}, owing to the time needed to recharge the inner wire of the proportional tube. Instead of the BF_3 working gas, comparable types are used, except filled with ^{3}He, having a comparable cross-section of thermal neutron absorption by the reaction ${}^{3}_{2}\mathrm{He} + \mathrm{n} \rightarrow {}^{3}_{1}\mathrm{H} + {}^{1}_{1}\mathrm{H}$ [17].

3.2.6 Ionization chamber

An ionization chamber measures the amount of secondary charges created in a gas volume, see also Section 2.5.2. Figure 37 shows such a round ionization chamber filled with Ar or N_2 gas. Typically, a sealed glass or metal tube contains $\sim$100–1000 cm^3 of gas between an outer high-voltage electrode and an inner grounded and readout electrode [45, 46]. The output current is measured using a sensitive current-to-voltage or current-to-frequency converter [47]. The ionization chamber is not sensitive to neu-

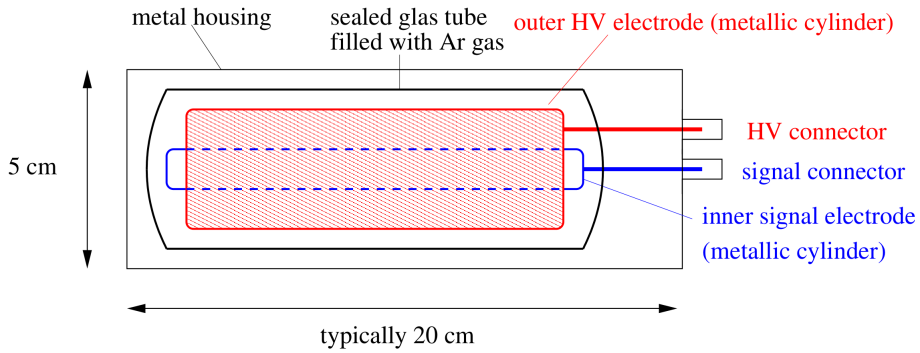


Fig. 37: Co-axial ionization chamber, as used for beam loss detection

Table 8: Basic parameters for the so-called RHIC-type co-axial ionization chamber [45, 46] and the planar LHC-type ionization chamber [48, 49].

Parameter	Co-axial ionization chamber	Planar ionization chamber
Outer length & diameter [cm]	20 & 6	50 & 9
Active gas volume [l]	0.11	1.5
Gas type & pressure [bar]	Ar & 1.1	N_2 & 1.1
Number of electrodes	2	61
Distance of electrodes [mm]	$\varnothing_{\text{inner}} = 6.3$, $\varnothing_{\text{outer}} = 38$	5.7
Voltage [kV]	1	1.5
Reaction time [µs]	$\simeq$3	$\simeq$0.3

trons and has a low detection efficiency for γ-rays; mainly charged hadrons and $e^{\pm}$ are detectable. By definition, the signal strength gives the absorbed dose in Gy directly. Because an inert, rare gas, such as Ar, or diatomic gas, such as N_2, is used in the detection volume, the device is very radiation hard. The signal strength is orders of magnitude less than for detectors in a particle-counting mode. Figure 37 depicts the scheme of a co-axial ionization chamber. For this geometry, the ionization chamber does not react as quickly as the scintillators, because the gas ions, created by the radiation, need $\sim$10 µs to reach the electrode. This time constant leads to a convolution of the primary signal, which is acceptable for most applications. The readout of the ionization chamber current by the digital electronics is therefore not faster than 1 ms in typical applications. The important parameters for such an ionization chamber are summarized in Table 8. A typical installation of a co-axial ionization chamber at an accelerator beam line or synchrotron is depicted in Fig. 38.

The detection threshold of an ionization chamber is proportional to the gas volume. To achieve a fast reaction time, even for a large gas volume, the drift time of the gas ions and electrons should be reduced. This can be realized by a parallel electrode arrangement with alternating biased and readout electrodes. Such an ionization chamber is shown in Fig. 39, as installed at CERN LHC and other facilities. About 4000 beam loss monitors are installed along the 27 km long LHC, corresponding to an average distance of about 6 m. They serve as the main detectors for beam protection and can trigger a fast abort. The design of such ionization chamber for LHC and relevant simulations concerning the response to various radiation species are discussed, e.g., in Refs. [48, 49].

3.2.7 Comparison of different beam loss monitors

Several of the described types of beam loss monitor are tested at the extraction septum magnets at the GSI synchrotron [42]. A typical example is shown in Fig. 40 for a O^{8+} beam of 800 MeV/u extracted; the maximum number of stored particles was 4×10^{10}. The linearity of the different detectors is checked

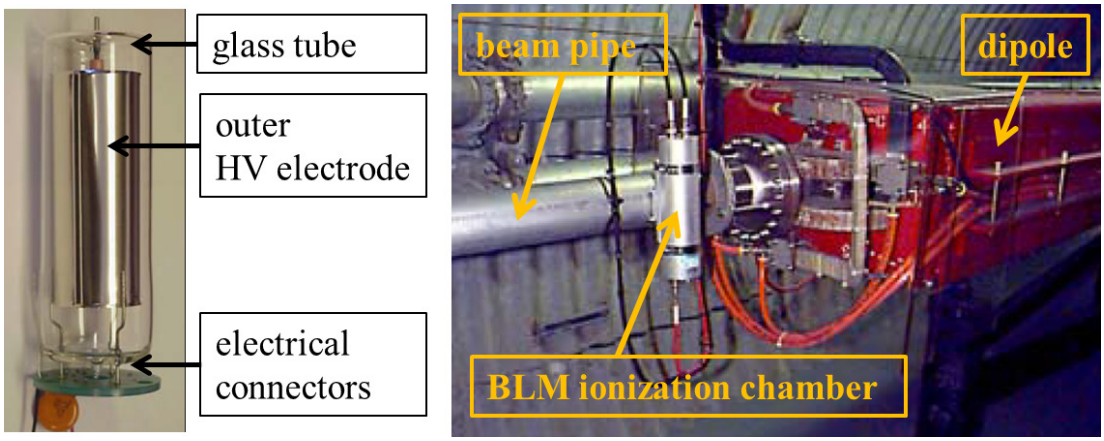


Fig. 38: Left: A 15 cm long co-axial ionization chamber used as a beam loss monitor (BLM). Right: Installation of such a ionization chamber for beam loss detection in a transport line.

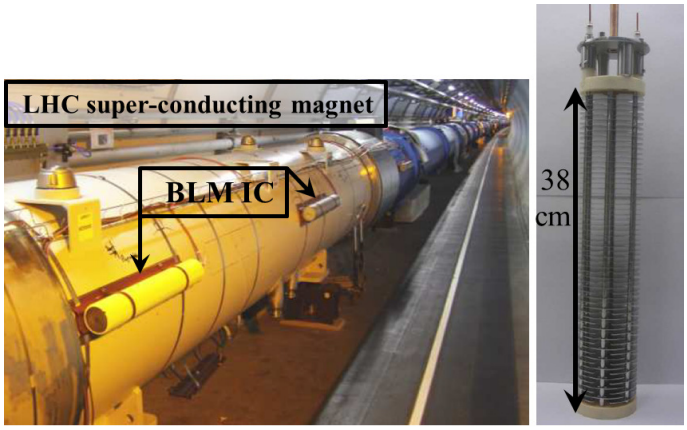


Fig. 39: Installation of a 50 cm long ionization chamber comprising 61 parallel plates at CERN LHC [48, 49]

by plotting the total counts for one spill as a function of the current detected at the experiment. The count rates obtained using the different types of monitor are quite different. The plastic scintillator shows the largest rate, about a factor of 30 greater than the BF_3-tube, owing to the detection of more categories of secondary particles. The liquid scintillator shows a smaller rate; the saturation for the highest rate is due to the slow-integrating pre-amplifier. The signal strength of the ionization chamber is smaller by a factor of 200 than that of the plastic scintillator. The dynamic range is highest for the plastic scintillator. Results for the PIN diode are not shown here, but the count rate would have been about three orders of magnitude less than that of the plastic scintillator. The choice of beam loss monitor type for a given range of beam parameters is therefore mainly driven by the expected count rate for a given beam loss scenario and the expected composition of the secondary particles.

3.3 Machine protection using beam loss detectors

The most frequent application of beam loss monitors is the protection of accelerator components. The high energy stored in the beam can lead to the damage of surrounding materials. In particular, the vacuum pipe can be destroyed by the temperature increase due to the energy release of the particles, leading to a vacuum leakage. Moreover, other materials, like electrical feed-through connections, the insulation of magnet coils or any electronic devices, can be destroyed as a result of the modification of their material by the radiation. As discussed in Section 3.1, nuclear interaction leads to radioactive nuclei, which have lifetimes of hours or longer. This results in activation of the accelerator components, preventing access

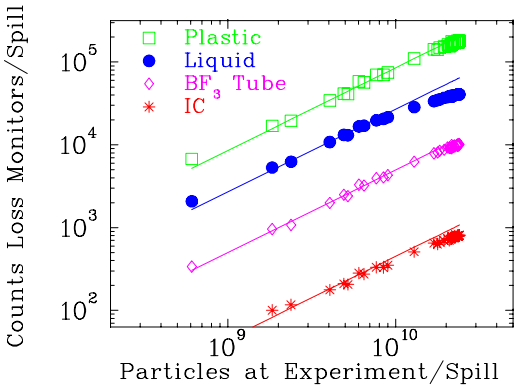


Fig. 40: Comparison of different beam loss monitors as a function of beam delivered to the target current for a O^{8+} beam of 800 MeV/u during slow extraction [42]. The lines are linear fits.

to the accelerator tunnel and strongly hampering maintenance. As a general rule, the power dissipation of the beam losses should be below 1 W/m to enable 'hands-on maintenance'.

Another important application of beam loss monitors is the quench protection of superconducting magnets and cavities. Even a small amount of loss can heat up the superconducting material to above the critical temperature, causing a quench, i.e., a transition to normal conductivity. This dangerous situation is always controlled by loss monitors. They create an interlock: as soon as a critical count rate or dose is reached, the beam is immediately kicked out of the accelerator. The design criteria of an interlock system are discussed in Ref. [47]. Last but not least, the reason for reducing losses is protection of the environment and an efficient transfer of particles to the experiment. In the control room of most high-current accelerator facilities, the loss rate at sensitive locations is monitored online to visualize the full functionality of the machine.

We discuss briefly two categories of loss.

- *Irregular or fast losses* occur because of malfunctions of accelerator devices, e.g., magnets or RF cavities of power supplies. Such a malfunction leads to a beam with wrong parameters and subsequently the loss of part or all of the beam particles. Moreover, a misalignment of the device settings (e.g., the beam is not centred inside a quadrupole leading to steering, or a wrong RF cavity phase or amplitude produces a beam with the wrong kinetic energy) results in a mismatch of beam at the subsequent stages of the accelerator and part of the beam might be lost. The task of the beam loss monitors is to localize these losses, warning the operator or triggering an interlock to stop beam delivery.
- *Regular or slow losses* are known losses, e.g., at aperture limits of a collimator system, or losses due to the finite lifetime of the beam in a synchrotron. Most of these are unavoidable, but an increase in the loss rate indicates a malfunction. In addition, beam loss monitors can be used to find the appropriate parameters, e.g., for an emittance-reducing collimator.

It is obvious, particularly from the first item, that a beam loss monitor system should have a high sensitivity as well as a very large dynamic range. The high sensitivity is necessary to produce a usable signal when only a very small fraction of the beam is lost. Using scintillators or PIN diodes, single-particle counting can be performed with a very low detection threshold. Detecting the lost particles directly gives a much higher sensitivity than looking at the lost fraction of the total signal using a current transformer. The high dynamic range is mainly necessary for the detection of irregular losses, because the losses at one location can easily vary by several orders of magnitude, e.g., due to a power failure of a magnet or cavity. The monitoring has to be operated over several orders of magnitude concerning the signal strength without changing the full-scale range, to prevent saturation of the signal for large losses.

Here, a particle-counting mode delivers this high dynamic for rates ranging from $10\ \mathrm{s}^{-1}$ to $10^7\ \mathrm{s}^{-1}$. Close to a collimator, scintillators offer a high dynamic range, which is needed for sensitive detection of the tails in the transverse particle distribution. For other regular losses, the required dynamic range is not so large, owing to the slower time variation. Ionization chambers are often installed for those applications.

Beam loss monitors are relatively cheap instruments and are installed outside the vacuum pipe, see Fig. 32. At most facilities, they are mounted at many locations along the accelerator to give good spatial coverage. The required time resolution depends very much on the application. To resolve the properties of individual bunches, fast-responding detectors (reaction time less than 100 ns), such as scintillators or PIN diodes, are necessary. Ionization chambers are much slower, having, in combination with the connected analogue and digital electronics, only $\sim$10 µs time response. However, ionization chambers offer direct determination of dose and are therefore often installed for machine projection systems.

The positions of the beam loss monitors should be well chosen. For greatest sensitivity, the distribution of the secondary particles should be a maximum at this location. The secondary particle distribution must be calculated using an adequate software code. For high energies, the well-known code GÉANT [50] is used in most cases. A loss of beam particles at a quadrupole is assumed in many cases because the beam has the largest width there. Because of the alternating focusing, the differences in the horizontal and vertical beam widths offers the possibility of distinguishing between horizontal and vertical losses. Moreover, monitors are often installed at insertions that limit the acceptance, such as injection of extraction septa.

3.4 Beam loss monitoring for beam alignment

Beam loss monitors have two main advantages, namely the high sensitivity and the fact that these monitors can be installed at many locations. For beam alignment at critical positions with small beam pipes, beam loss monitors can be mounted densely. An example for the slow extraction at the GSI synchrotron is shown in Fig. 41. Here, the septum elements, used to kick the beam, have small pipe diameters, comparable to the actual beam size. If the angle at the input of the septum magnets is slightly wrong, part of the beam is lost there. Therefore, the angle kick at the first element (here the electrostatic septum) must be aligned carefully in small steps to find the minimal loss rate. Using beam loss monitors, the alignment procedure can be executed with more sensitive input signals than by monitoring the transmitted current, as demonstrated by the top curve in Fig. 41. Beam loss monitors can serve as a basic tool for daily operation. Beam loss monitors are also useful for the time-dependent monitoring of a possible misalignment at different sections of a synchrotron or a transfer line, e.g., during slow extraction, because they can be installed at several critical locations. At some linac-based facilities, beam loss monitors are used as the operation tool for transmission optimization, owing to their high sensitivity, which might be greater than achievable by measuring the transmitted beam via current transformers, see, e.g., Ref. [51].

4 Measurement of beam profile

The beam profile can be influenced by quadrupole magnets installed in all accelerating structures and in transfer lines. Measurement is important to control the beam width, as well as transverse matching between different parts of an accelerating facility. The large number of bending, focusing, and correction magnets gives rise to the need for many profile measurement stations. Depending on the beam particles, current, and energy, a very large variety of devices exists. We can only briefly discuss the most popular ones. The beam spot can be observed directly by intercepting it with scintillator screens and viewing the emitted fluorescence with a camera. Secondary electron emission grids (SEM grids) are widely used as an electronic alternative to achieve a large dynamic range; a grid of wires with typically 1 mm spacing is inserted. Instead of a fixed grid, a single wire can be scanned through the beam to give high spatial resolution; this is called a wire scanner. A third approach is a so-called ionization profile monitor, in which the ionization products of the interaction of the beam with the residual gas atoms or molecules inside the

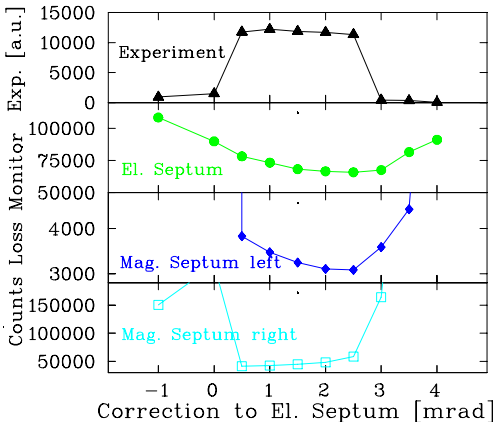


Fig. 41: Counts per spill from three loss monitor locations and the current at the experiment as a function of the angle electrostatic septum at the GSI synchrotron [42].

(imperfect) vacuum of the tube are detected. In these devices, the resulting electrons or ions are accelerated by an external electrical field to a detector with a spatial resolution. The ionization profile monitor is well-suited as a non-destructive method for a synchrotron. When charged particles with relativistic velocities pass through a metallic foil, radiation is emitted, as described by classical electrodynamics; this process is called optical transition radiation. This light is recorded by a camera to determine the beam profile. High-energy electrons emit synchrotron radiation if the trajectory is curved. Monitoring this light by synchrotron radiation monitors yields direct information about the beam spot.

Normally, the diameter of an electron beam is less than 1 mm, while proton or heavy-ion beams have large diameters, up to some centimetres. In particular, in a synchrotron with multiturn injection, diameters of several centimetres can be produced.

The beam width is mainly given by the settings of the focusing magnets and the beam emittance ϵ. Knowing the lattice, i.e., the β function $\beta(s)$ and the dispersion $D(s)$ at the monitor location s, the measured beam width $\sigma(s)$ is given by

$$\sigma_x^2(s) = \epsilon_x \beta_x(s) + \left(D(s) \frac{\Delta p}{p} \right)^2 \qquad \text{and} \qquad \sigma_y^2(s) = \epsilon_y \beta_y(s)\,. \tag{20}$$

In a synchrotron, the lattice functions are well-known, or can be measured separately. If the momentum spread $\Delta p/p$ is also known, the emittance can be calculated. (In the vertical direction, the dispersion is zero in most cases, because only horizontal bending magnets are used.) Nevertheless, the contribution due to the dispersion must be taken into account for the interpretation of the beam width. In a linac, the lattice functions are not so precisely fixed, owing to the variable input emittance orientation, leading to a less stringent relation between profile width and emittance.

4.1 Scintillation screen

The most direct method of beam observation involves recording the light emitted from a scintillation screen intersecting the beam, monitored by a commercial video, CMOS (complementary metal-oxide-semiconductor), or charge-coupled device (CCD) camera, see, e.g., Refs. [52–54] for an overview. These devices are installed in nearly all accelerators from the source up to the target. A scintillation screen is schematically shown in Fig. 42 together with a realization in which the pneumatic drive is mounted on a $\varnothing$200 mm flange.

When a charged particle penetrates a material, the energy loss can be transformed to fluorescence light, as discussed in Section 2.5 for the scintillation counter. The important properties of such a scintillator are as follows.

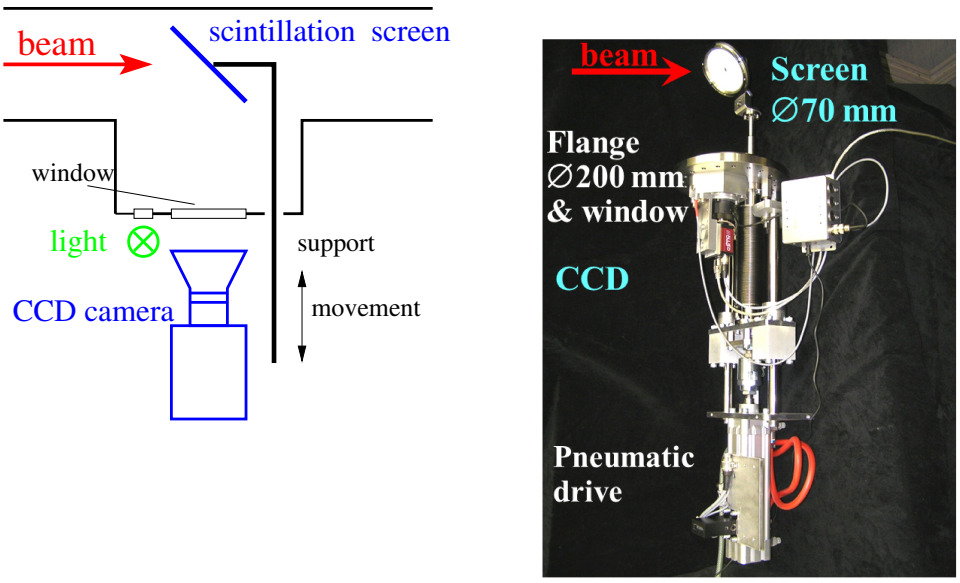


Fig. 42: Left: Intercepting scintillation screen. Right: A P43 phosphor scintillation screen of ⌀70 mm and CCD camera are mounted on a ⌀200 mm flange with pneumatic drive.

Table 9: Chemical composition and some basic optical properties of inorganic scintillation screens. The matrix Al_2O_3 is used in the form of a ceramic. The last four materials are so-called phosphor screens, where powder is deposited on glass or metal plates.

Material	Type	Composition	Activator	Maximum emission [nm]	Decay time
Quartz	Glass	SiO_2	None	470	<10 ns
Alumina	Ceramics	Al_2O_3	None	380	∼10 ns
Chromox	Ceramics	Al_2O_3	Cr	700	∼10 ms
YAG	Crystal	$Y_3Al_5O_{12}$	Ce	550	200 ns
LuAG	Crystal	$Lu_3Al_5O_{12}$	Ce	535	70 ns
Caesium iodide	Crystal	CsI	Tl	550	1 µs
P11	Powder	ZnS	Ag	450	3 ms
P43	Powder	Gd_2O_2S	Tb	545	1 ms
P46	Powder	$Y_3Al_5O_{12}$	Ce	530	300 ns
P47	Powder	$Y_2Si_5O_5$	Ce & Tb	400	100 ns

- High light output matched to the optical system of the camera in the optical wavelength range of $400\ \text{nm} < \lambda < 700\ \text{nm}$.
- High dynamic range, i.e., a good linearity between the incident particle flux and the light output. In particular, a possible saturation of the light gives rise to a deformation of the recorded profile.
- No absorption of the emitted light to prevent artificial broadening by stray light inside the material.
- No significant contributions from reflection at the surfaces of the scintillation light produced inside the material; depending on the geometry, this can lead to an artificial profile broadening. In this case, e.g., anti-reflex coating should be applied.
- Fast decay time, to enable the observation of possible variations of the beam size.
- Good mechanical properties for producing up to ⌀100 mm large screens.
- Radiation hardness to prevent permanent damage.

As stated, plastic scintillators have only low radiation hardness and various kinds of inorganic material are used instead. Some properties are given in Table 9 [52, 53]; for more details see Refs. [17, 29].

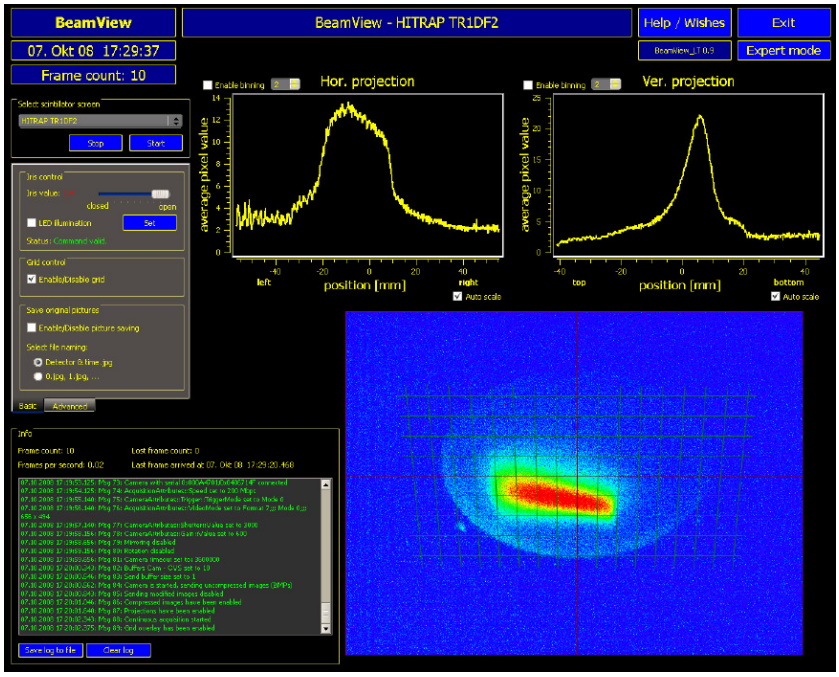


Fig. 43: Beam spot of low-current 4 MeV/u beam at GSI on a YAG:Ce screen recorded with a digital CCD camera. The original greyscale image is converted to false colour for better visibility. The projection on the horizontal and vertical axes are shown in the upper part. From Ref. [65].

The material Chromox is a ceramic, which can be fabricated like pure Al_2O_3 and used widely as an insulating material. The scintillation is due to the Cr activators; chemically, it is comparable to ruby. The disadvantages are its long decay time of $\sim$10 ms and the large absorption inside the material. Nevertheless, owing to its robustness, it is quite often used. The well-known Ce-activated crystal materials, such as YAG ($Y_3Al_5O_{12}$) have better optical properties and are widely used for particle-counting applications [17]. However, it is difficult to produce crystalline discs of about 10 cm diameter from this material. Instead of the expensive single-crystal disc, one can use small grains of $\sim$10 µm size, deposited on glass or metal discs that are several millimetres thick. These phosphor screens are also used very frequently in analogue oscilloscopes, electron microscopes, and image intensifiers. P46 is an example for a phosphor screen powder offering a much lower production cost compared with the chemically equivalent single-crystal YAG:Ce. The properties of doped materials depend strongly on the activator concentration; therefore, Table 9 gives only approximate values. The sensitivities of the different materials span several orders of magnitude and also depend on the particle species, see, e.g., Refs. [54–56]. Scintillation screens are used in all types of accelerators [54]; some examples of investigations are discussed for electron beam [57–60] and proton or ion [55, 56, 61–64] accelerators with quite different applications.

The beam image from a YAG:Ce screen is shown in Fig. 43 for a low-current beam of only $\sim 10^6$ ions at 4 MeV/u, proving the high sensitivity of that material.

For high-intensity beams, it must be ensured that the material is not destroyed by the absorbed beam power. In particular, for slow heavy ions with a range comparable to the screen thickness, this restricts the use; see also the discussion in Section 2.3. A disadvantage of the screen is related to the interception. The used material is so thick (several millimetres) that it causes a large energy loss, so that it cannot be used for the diagnostics of a circulating beam inside a synchrotron.

The screen is observed with a CCD or CMOS camera. A modern approach uses a digital link, with digital data transfer of the camera pixel values, enabling fast data rates for long cable lengths without

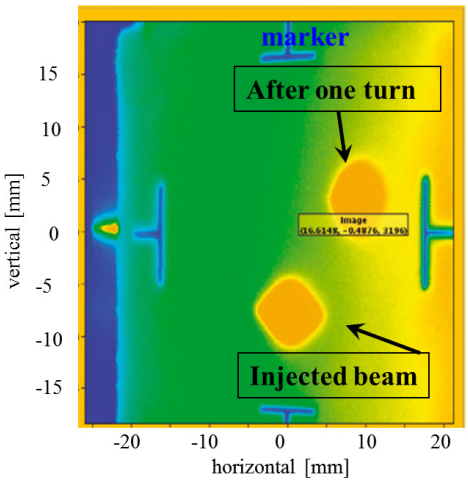


Fig. 44: Beam spot of injected beam and beam after one turn at LHC, recorded using a Chromox screen

signal degeneration, see, e.g., Ref. [65]. One problem is the radiation sensitivity of the CCD sensor and the digital equipment. At high levels of radiation, old-fashioned analogue vidicon cameras are used; in this case, the analogue video output is digitized by a frame grabber.

For alignment of the injected beam into a synchrotron, a so-called first-turn diagnostics unit is installed. After a one-turn passage, the beam hits a profile measurement device, in most cases a scintillation screen, to check for the beam properties after one turn (for this case, the synchrotron is more or less treated as a single-pass transfer line). In Fig. 44, a historically important example is depicted from the injection to the LHC, as recorded at the first day of its operation, in which a scintillation screen is installed just after the injection location to monitor the injected beam as well as the beam after one turn. For larger synchrotrons, such first-turn diagnostics are installed at several sectors. The profile diagnostics are often accompanied by the determination of beam current, using a Faraday cup for low-energy ions, a current transformer for multiturn injection or bream position monitor (BPM) readouts in a single-pass mode for single-turn bunch-to-bucket injection.

4.2 Secondary electron emission grid

When particles hit a surface, secondary electrons are knocked out from the surface, as described in Section 2.5.3. For the profile determination, individual wires or ribbons interact with the beam; this device is called a secondary electron emission grid or a harp, as reviewed in Ref. [66]. Each of the wires has an individual current-to-voltage amplifier. This is an electronic alternative to a scintillation screen with a much higher dynamic range, i.e., the ratio of minimal to maximal detectable current is orders of magnitude greater. W-Re alloys are often used for the metallic wires or flat ribbons, owing to their excellent refractory properties. This is particularly important for low-energy linacs because no cooling can be applied, owing to the geometry. A photograph of such a device is shown in Fig. 45; the specifications are given in Table 10.

For low energies at proton or heavy-ion linacs, the particles are stopped in the material or undergo significant energy loss. The diameter-to-spacing ratio of the wires determines the attenuation of the beam current (and, of course, also the signal strength on the individual wires). Typically, only 10% of the beam area is covered by the wires; in this sense, the profile measurement is nearly non-destructive. For energies above 1 GeV/u, the relative energy loss is negligible and large ribbons are used. A typical example of the SEM grid measurement used for beam optimization is shown in Fig. 46 for a transfer line.

Each wire has an individual pre-amplifier, in most cases an I/U converter, followed by a sample-and-hold circuit (Fig. 47). These electronics must be installed close to the accelerator hardware. With

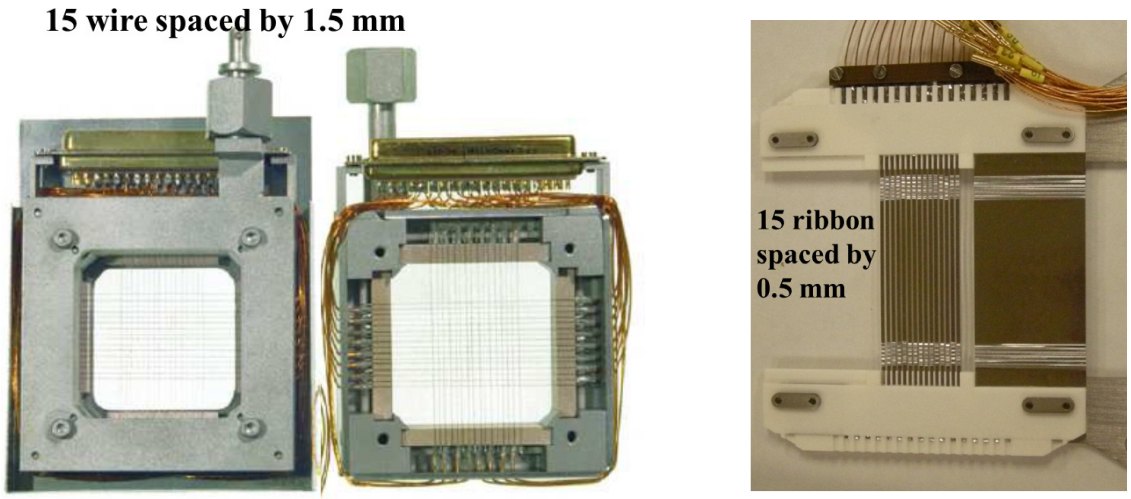


Fig. 45: Left: Secondary electron emission grid for both planes with 15 tungsten wires spaced by 1.5 mm as typically used at a proton linac. The individual wires are insulated with glass-ceramics. Middle: Same grid with tin cover removed. Right: Grid based on ribbons of 25 µm thick aluminium foil by laser cutting; the spacing is 0.5 mm. This type is used for high-energy protons, typically >1 GeV. The devices are mounted on a pneumatic drive to move them in and out of the beam (not shown).

Table 10: Typical specification for SEM grid used at proton and heavy-ion linacs

Parameter	Value
Diameter of the wires	0.05–0.5 mm
Spacing	0.5–2 mm
Length	50–100 mm
Material	W or W-Re alloy
Insulation of the frame	Glass or Al_2O_3
Number of wires	10–100
Maximum power rating in vacuum	1 W/mm
Mininimum sensitivity of I/U conversion	1 nA/V
Dynamic range of electronics	$1{:}10^6$
Integration time	1 µs to 1 s

a multiplexer, the analogue values are transported to an analogue-to-digital converter (ADC) located outside the accelerator tunnel. Readout of a full SEM grid in less than a millisecond is typical for the use of pulsed or d.c. beams.

4.3 Wire scanner

Instead of using several wires with individual, expensive, electronics, a single wire can be swept through the beam [66]. The advantage is that the resolution is not limited by the wire spacing; therefore, this technique is often used in electron accelerators with beam sizes in the sub-millimetre range. It can also be applied in proton synchrotrons, owing to the small amount of intercepting matter.

An arrangement with a straight wire on a fast pendulum mechanics in shown in Fig. 48 [67, 68]. Scanning velocities up to 10 m/s can be achieved with a special pneumatic mechanism. Sometimes this set-up is also called a flying wire. Carbon or SiC is used as the wire material, owing to its low weight and low nuclear charge, Z, resulting in a low-energy deposition in the wire (the Bethe–Bloch equation (Eq. 13) gives $\mathrm{d}E/\mathrm{d}x \propto \rho \cdot Z/A$). In addition, these materials can withstand high temperatures without melting. The wire can be as thin as 10 µm. However, owing to the scanning of the single wire, the

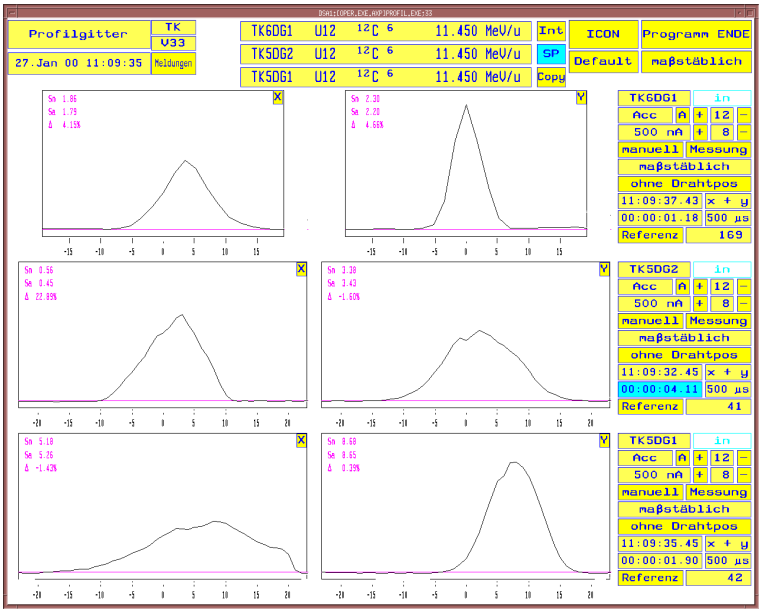


Fig. 46: Measured profiles [mm] of SEM grids in one transfer line at the GSI heavy-ion linac. Left: horizontal profile. Right: vertical profile.

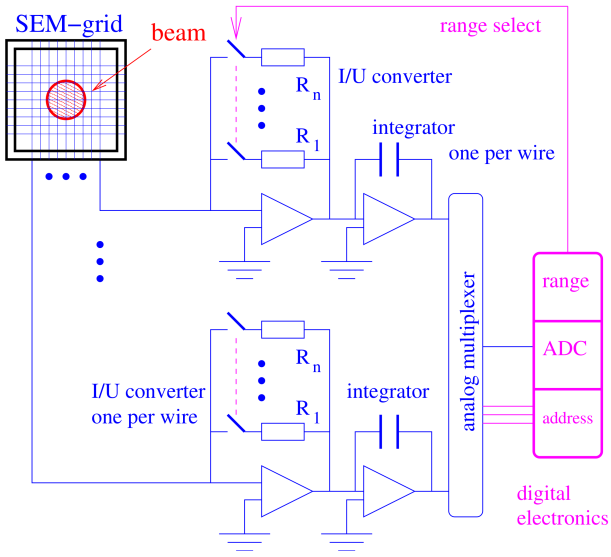


Fig. 47: Typical analogue signal processing for SEM grid

profile is not taken at a single instant, even at high scanning velocities. Therefore, only the steady-state distribution can be probed.

For profile display, the position of the wire, determined by the position encoder, is plotted along the horizontal axis. The beam signal for the vertical axis can be deduced from the current given by the emitted secondary electrons, as for a SEM grid. This is done, in particular, for low-energy protons and heavy ions. In most cases, for beam energies larger than 150 MeV/u for ions (the threshold for π-meson production) or 10 MeV for electrons, the signal is deduced by monitoring the secondary particles outside the beam pipe, see Fig. 49. These secondary particles might be hadrons created by the nuclear interaction of the proton or ion projectiles and the wire, having enough kinetic energy to leave the vacuum chamber. For the case of electron accelerators, the secondary particles are mainly Bremsstrahlung photons. The detector is simply a type of well-suited beam loss monitor, e.g., a scintillator installed several metres

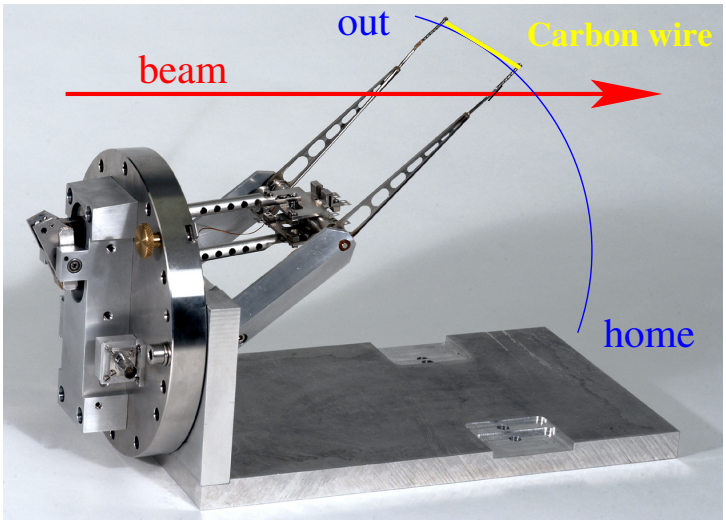


Fig. 48: Pendulum scanner or 'flying wire' used at CERN synchrotron [67]

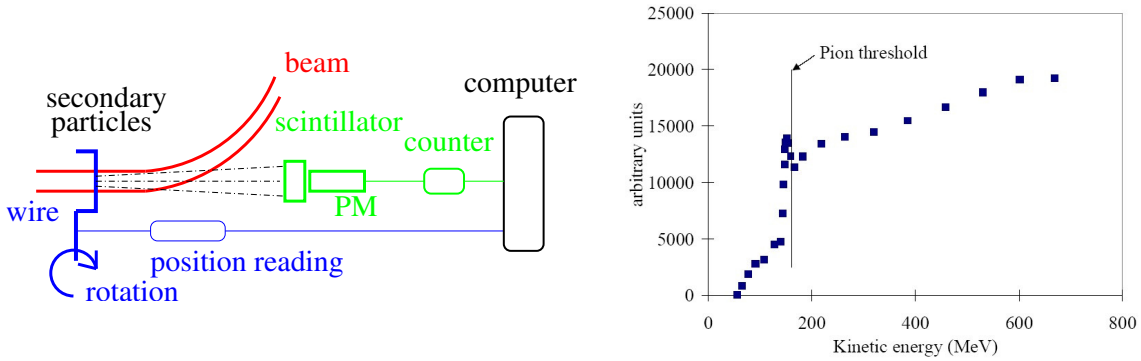


Fig. 49: Left: Scheme of wire scanner using the production of secondary particles as the signal source. Right: Total rate from wire scanner under proton impact at CERN PS-Booster, showing the rate increase above the π-threshold at $\sim$150 MeV [69].

away (see also Section 3). The count rate is plotted as a function of the wire position, as a precise representation of the beam profile.

For fast movements, vibration of the wire restricts the spatial resolution. A much greater precision can be achieved using a much slower movement of, for example, only 0.1 m/s, with a linear drive, as shown in Fig. 50 [70]. With these types, a resolution of 1 µm is reached in electron accelerators [71]. To achieve this low beam width reading, σ_{beam}, for a Gaussian beam, the contribution of the wire (4 µm thick carbon wire in this case) must be deconvoluted from the measured width σ_{meas} according to

$$\sigma^2_{\mathrm{beam}} = \sigma^2_{\mathrm{meas}} - 4 \cdot r^2_{\mathrm{wire}} \quad . \tag{21}$$

In most cases, the wire is mounted on a fork, which is inserted into the beam pass by a drive mounted at 45°. Then only one drive is sufficient for measuring both transverse planes by mounting the wires in a crossed orientation, as shown in Fig. 50. If the signal generation is performed by reading the secondary emission current, two isolated wires can be mounted in a compact manner. If the signal is generated by beam loss monitors outside the beam pipe, the two wires must be separated (as shown in the left-hand panel of Fig. 50), so that the wires cross the beam one after the other.

A comparison of the wire scanner and the SEM grid shows the advantages and disadvantages of both types [66].

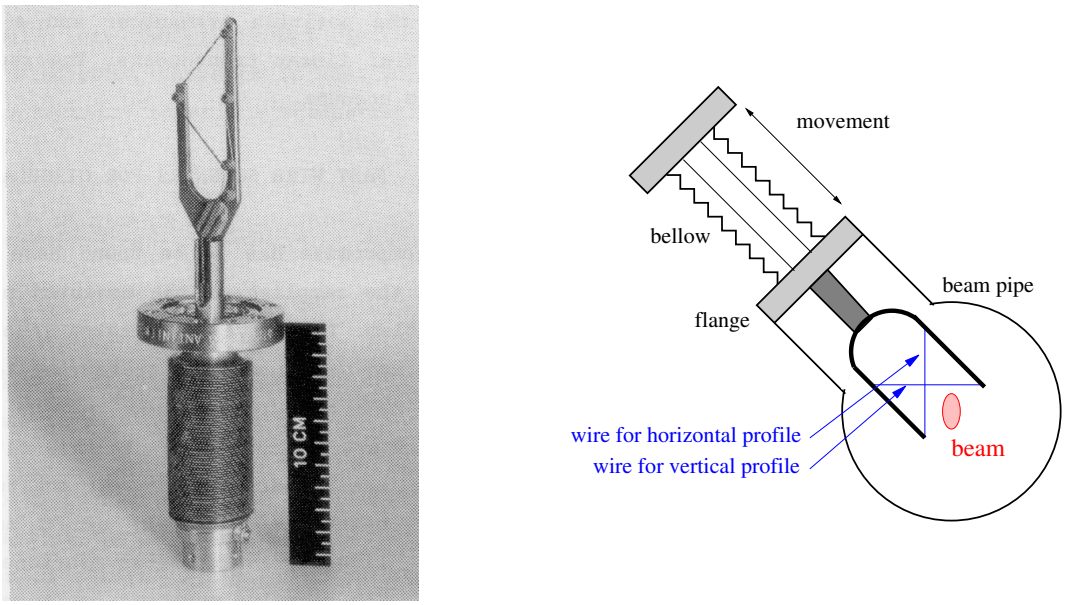


Fig. 50: Left: Linear scanner used at CERN [70]. Right: Schematics of the movement

– With a SEM grid, the beam intensity is sampled concurrently, whereas a moving wire samples the parts of the profile at different locations at different times. Therefore, variations of the beam intensity in time will be mixed with transverse intensity variations using a scanning device.
– In the case of pulsed beams, a synchronization between the readout of the electronics and the beam pulse is required, as well as a movement within the beam pulse. The synchronization between readout and beam pulse is more easy to achieve for a SEM grid.
– The resolution of a SEM grid is fixed by the wire spacing (typically 1 mm), while a wire scanner can have much higher resolution, down to 10 µm, owing to its constant movement. (For high resolution, mechanical vibration must be avoided.)
– The electronics for data acquisition are cheaper for a scanning system. A SEM grid requires one channel per wire.
– For the cost of the mechanics, it is the other way around: the precise vacuum actuator for the scanner is more expensive than the pneumatic drive needed for a SEM grid.

4.4 Multiwire proportional chamber

For slowly extracted beams from a synchrotron, the current is much too low to be measured using a SEM grid. One can use the amplification of electrons in a gas, as is done in a multiwire proportional chamber; for the principle of operation see, e.g., Ref. [17]. As discussed for the ionization chamber for current measurements in Section 2.5.2, the primary particles traverse a gas (such as 90% Ar mixed with 10% CH_4 or CO_2), creating secondary electrons. A multiwire proportional chamber consists of a grid held at a high voltage, typically several kilovolts, and a grounded grid, which is read by a charge-sensitive pre-amplifier, as for SEM grids. The distance between the anode and the cathode plane is typically 1 cm and the spacing of the wires about 1 mm. The principle is shown in Fig. 51. Close to the anode wires, typically of 25 µm diameter, the electrical field increases in proportion to $1/r$, see Fig. 52. After reaching a threshold, the energy of the electrons accelerated towards the wires is high enough to knock out additional electrons from the gas atoms or molecules. This gives rise to an avalanche, which results in a $\sim 10^4$ amplification in the number of electrons. This amplification inside the detector volume is nearly noise-free, owing to the high source electrical impedance of the free charge carriers. The resulting low noise could not be achieved by an electric amplifier, owing to its thermal noise. The following electronics

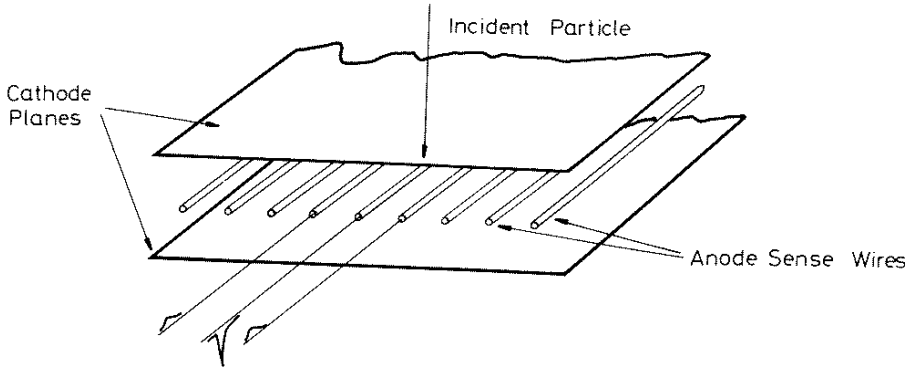


Fig. 51: Scheme of multiwire proportional chamber for one plane, showing signal generated by particle interaction

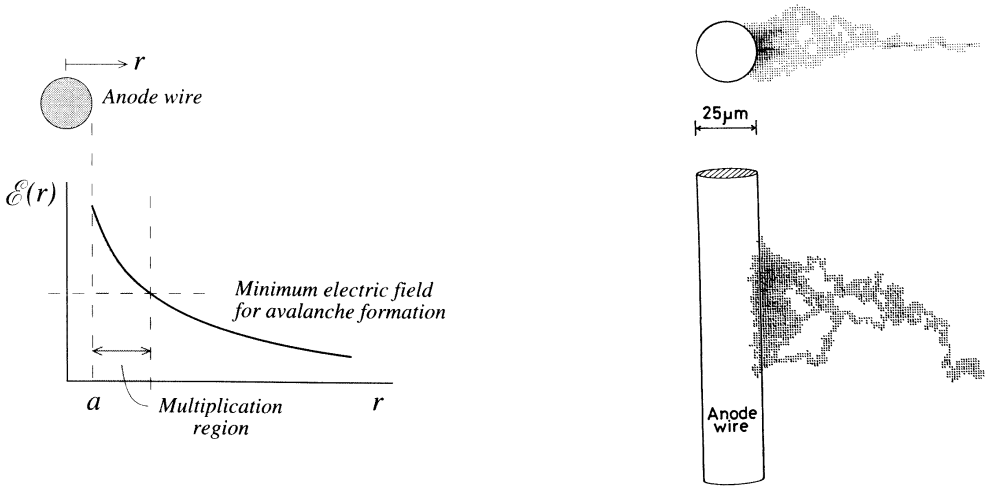


Fig. 52: Left: Electric field close to anode wires, showing amplification region. Right: Monte Carlo simulation showing electron avalanche as grey shading [72].

(further amplifier and ADC) and the display method are comparable to the procedure for SEM grids

An example of a multiwire proportional chamber used at GSI is shown in Fig. 53.

4.5 Ionization profile monitor

A frequently used non-destructive method for profile determination is the ionization profile monitor, sometimes also called the residual gas monitor. These monitors are installed in nearly every proton or heavy-ion synchrotron for the detection of beam sizes between some millimetres and several centimetres; a review is given in Ref. [73, 74]. They are seldom used for electron synchrotrons, owing to the smaller electron beam. The principle is based on the detection of the ionized products from a collision of the beam particles with residual gas atoms or molecules in the vacuum pipe. A scheme for such a monitor is shown in Fig. 54. Typical pressures for linacs and transfer lines are in the range 10^{-8}–10^{-6} mbar, containing mainly N_2 and O_2, and for synchrotrons are in the range 10^{-11}–10^{-9} mbar, containing mainly H_2; the different compositions are due to the different vacuum pumps used. Owing to the electronic stopping process, electrons are liberated and electron–ion pairs are generated. An estimate of the signal strength can be obtained using the Bethe–Bloch formula. However, owing to the single collision behaviour, the W values (see Section 2.5.2) are only a first-order approximation for the conversion of the energy loss to the numbers of secondary particles; the process is better approximated by considering some 100 eV for the creation of an electron–ion pair. To separate the electron–ion pairs, an external electric field of the

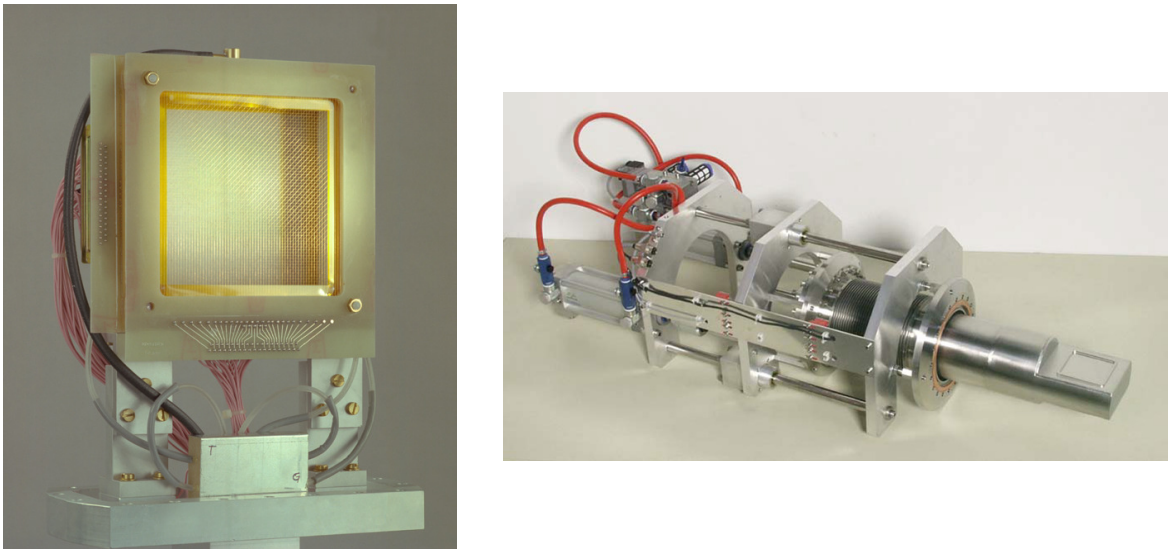


Fig. 53: Left: Multiwire proportional chamber used for profile measurement of slowly extracted ion beams at GSI. Right: Pneumatic drive with multiwire proportional chamber mounted inside pocket having normal pressure and separated from vacuum by a 100 µm thick foil.

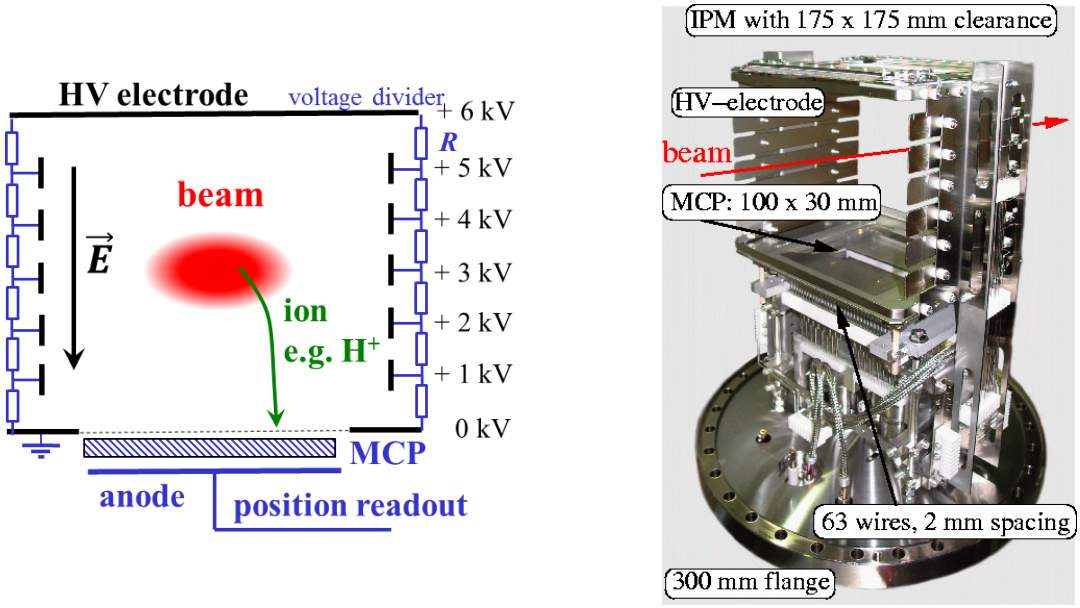


Fig. 54: Left: Scheme of ionization profile monitor for horizontal profile determination. Right: The large aperture ionization profile monitor installed at the GSI synchrotron for the horizontal direction; the clearance is 175 mm × 175 mm. The monitor is mounted on a ⌀300 mm flange. The readout behind the multichannel plate (bottom) is achieved with an array of 63 wires with 2 mm spacing.

order of 1 kV/cm is applied by metallic plates installed in the vacuum pipe, but outside the beam path. To obtain a more homogeneous field, guiding strips are installed at the side to get a smooth transition from the high-voltage side to the ground plane. For most applications, the residual gas ions are detected in this set-up. A realization from the GSI synchrotron, with a large opening, of 17 cm, is shown in Fig. 54 [75, 76].

For the linac case, the vacuum pressure is high and the energy loss is larger, owing to the lower beam energy. Enough electron–ion pairs are generated to give a measurable current (down to 0.1 nA

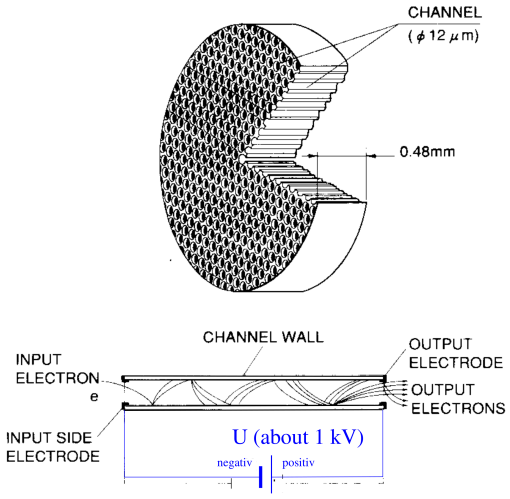


Fig. 55: Scheme of a single multichannel plate made of a glass plate with small channels, where the amplification of the secondary electrons takes place.

per strip) of secondary ions for direct detection by a sensitive SEM grid-like wire array [77, 78]. In a synchrotron, the pressure and the stopping power are smaller. A multichannel plate particle detector is used as a 'pre-amp'. A multichannel plate is made of a 1 mm thick glass plate with round channels of $\sim$10 µm, see Fig. 55. These channels are coated with high-resistivity materials and a voltage of $\sim$1 kV is applied across them. When a residual gas ion (or electron) hits the surface, secondary electrons are liberated, i.e., the ions are converted to electrons. These electrons are multiplied, as in a photomultiplier, by a factor of typically 10^3 per multichannel plate. Owing to the diameter of the channels, the spatial resolution of a single step multichannel plate is of the order of 0.03 mm. For higher amplifications, a two-step device (Chevron geometry) with a gain of $\sim 10^6$ is used, resulting in a resolution of about 0.1 mm.

To detect the secondary electrons at the multichannel plate output, two technologies concerning the anode are available.

- A SEM grid made of wires with $\sim$0.5 mm diameter and 1 mm spacing. The advantage is the fast readout made possible using appropriate electronics [79]. The disadvantage is the limited spatial resolution.
- A phosphor screen installed close to the back of the multichannel plate. The secondary electrons create a light pulse, which is imaged by a CCD camera. The advantage is the high resolution, as there are several hundreds of pixels for one plane, and the cheap and direct data acquisition system supplied by the commercial CCD camera. However, with this technique, the time resolution and the image repetition rate are limited by the CCD camera frame-rate, of about 100 frames per second. For faster readout, the CCD is replaced with a photomultiplier or avalanche diode, e.g., Ref. [80].

In the following, we discuss some examples for the profile measurement inside a synchrotron.

Adiabatic damping. During the acceleration of a beam inside a synchrotron, the beam width should shrink. This is related to the fact that the emittance is defined in the laboratory frame by the angle between the transverse and longitudinal velocities. During acceleration, the longitudinal velocity increases (for the non-relativistic case), while the transverse component remains the same, leading to a decrease of the angle between both vectors, as shown schematically in Fig. 56. This is called transverse adiabatic damping by the conservation of the normalized emittance; see also the discussion in Section 5 and Eq. 40. The normalized emittance is defined as $\epsilon_{\mathrm{norm}} = \beta\gamma\epsilon$ for a beam of velocity β, Lorentz factor γ and actual emittance ϵ. The beam width σ scales as $\sigma \propto \sqrt{\epsilon}$. In Fig. 56, the decrease of the beam width

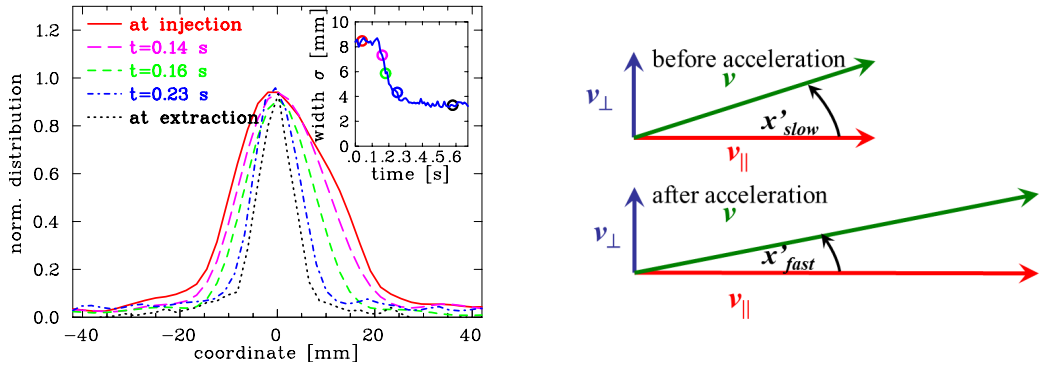


Fig. 56: Left: Evolution of the horizontal beam profile during the acceleration of C^{6+} from 6.7 MeV/u to 600 MeV/u within 0.4 s at the GSI synchrotron. The inset depicts the evolution of the beam width of one standard deviation during the acceleration started at 0.15 s; the depicted beam profiles are indicated by coloured circles. Right: Schematic visualization of transverse, $v_\perp$, and longitudinal, $v_\parallel$, velocity evolution during acceleration, leading to a decrease of the angle x'.

is demonstrated for the acceleration of a C^{6+} ion beam from a kinetic energy of $E_{\mathrm{ini}} = 6.7$ MeV/u, corresponding to a velocity of $\beta_{\mathrm{ini}} = 12\%$ and a Lorentz factor of $\gamma_{\mathrm{ini}} = 1.007$, to $E_{\mathrm{final}} = 600$ MeV/u, corresponding to $\beta_{\mathrm{final}} = 79\%$ and $\gamma_{\mathrm{final}} = 1.64$. The assumption of normalized emittance preservation would lead to the scaling of the expected beam width as

$$\sigma_{\mathrm{final}}^{\mathrm{theo}} = \sqrt{\frac{\beta_{\mathrm{ini}} \cdot \gamma_{\mathrm{ini}}}{\beta_{\mathrm{final}} \cdot \gamma_{\mathrm{final}}}} \cdot \sigma_{\mathrm{ini}}^{\mathrm{theo}} = 0.33 \cdot \sigma_{\mathrm{ini}}^{\mathrm{theo}}$$

for the case depicted in Fig. 56. The experimental value $\sigma_{\mathrm{final}}^{\mathrm{exp}} = 0.37 \cdot \sigma_{\mathrm{ini}}^{\mathrm{exp}}$ is close to that theoretical value; the larger value for the experimental case is related to a change of the focusing and further imperfections during the acceleration. An ionization profile monitor, as a non-intercepting device, is well-suited for a long-time observation without any influence on the beam.

Properties during multiturn injection. During a typical multiturn injection, the horizontal phase space is filled like a spiral. This is controlled by bumper magnets inside the synchrotron, which shifts the orbit of the stored particles for each turn and the duration of the beam delivery, as varied by a chopper in the transfer line in front of the synchrotron. The process is performed such that the central orbit, i.e., the central part of the horizontal phase space, is filled first. If now the length of the injected beam pulse is varied, the process can be demonstrated, as displayed in Fig. 57. For a short beam pulse, the beam width and, therefore, the beam emittance is small and increases with the duration of the beam delivery. The vertical emittance is not influenced in the displayed case, showing the absence of horizontal–vertical coupling in the synchrotron. However, the stored beam current is significantly smaller if only one or a few turns are injected. During operation, the amount of accelerated particles and their emittance can be balanced; for a proton synchrotron, the normalized emittance is preserved (as shown already) and therefore the beam quality is dominated by the injection process.

Steering of the injected beam. If the beam is injected at an improper angle with respect to the synchrotron reference orbit, the particles might still be stored. However, particles at a large angle compared with the (imaginary) reference particle fill the phase space at the outer periphery. The beam profile is the projection of the phase space to the spatial axis and this peripheral distribution might correspond to a non-Gaussian beam with large shoulders or even a 'hollow beam'. The is demonstrated in the left-hand panel of Fig. 58. Here, the beam's steering angle with respect to the reference orbit of the synchrotron beam axis is changed by a transfer line dipole in front of the synchrotron. Depending on the chosen angle, the entire synchrotron acceptance can be filled, as demonstrated by an increase in the width and, correspondingly, the emittance of the stored beam. To prevent such an emittance blow-up, the steering angle

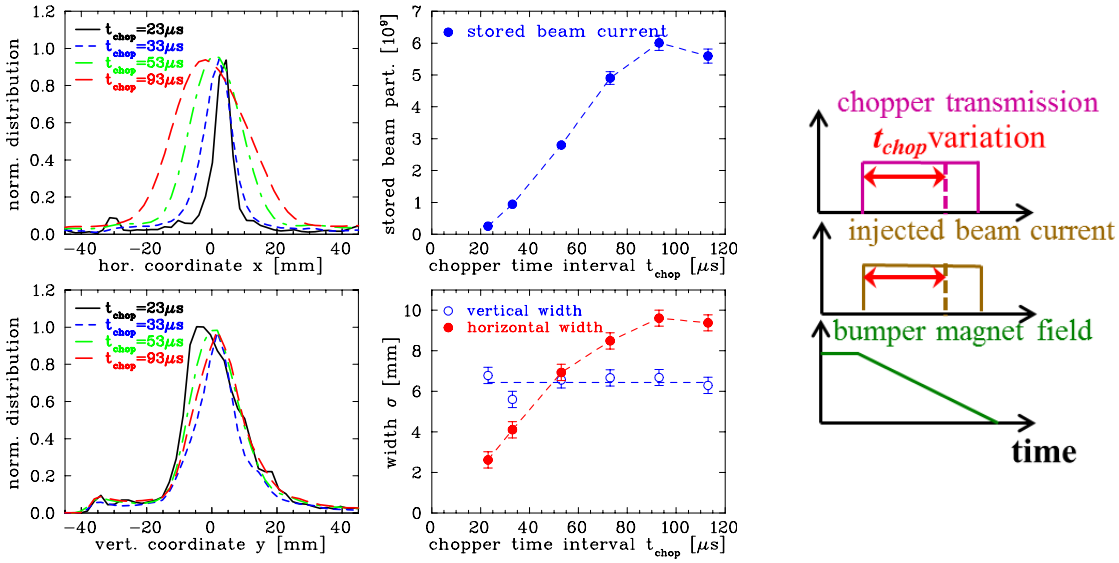


Fig. 57: Multiturn injection properties at an ion synchrotron for C^{6+} beam at 6.7 MeV/u with up to 6×10^9 ions per fill using a multiturn injection at the GSI synchrotron, with a revolution time of 5 µs per turn; each profile is an average over about 100 turns. Top left: Horizontal beam profile for different beam delivery times. Bottom left: Vertical profile. Top centre: Beam current measured by a d.c. transformer. Bottom centre: Calculated standard deviation of the profiles. Right: Schematic diagram of transfer line chopper and synchrotron bumpers.

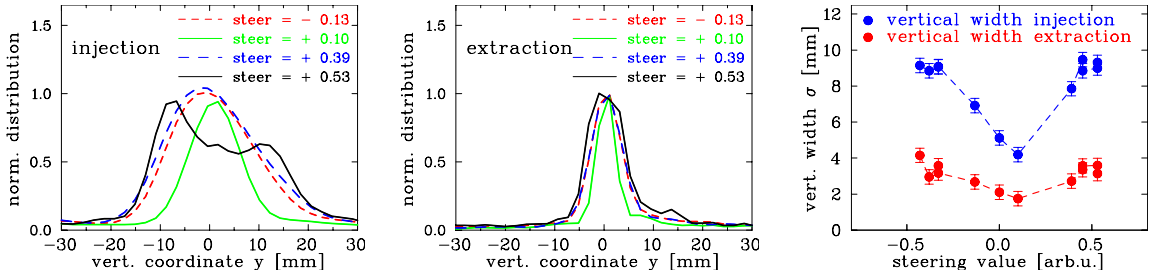


Fig. 58: Demonstration of the effect of vertical injection steering for the stored beam. The beam parameters are the same as of Fig. 57, followed by an acceleration to 600 MeV/u. Left: Vertical profiles at injection energy of 6.7 MeV/u. Right: Vertical profiles at extraction energy of 600 MeV/u. Right: Beam width as a function of steering angle.

must be optimized to achieve a minimal beam width. For hadron synchrotrons, the enlarged emittance is conserved even after acceleration, leading to a decreased beam quality for the user. This is demonstrated in the middle panel of Fig. 58, which shows the profile measurement of the circulating beam after acceleration for the same values of the steerer in the injection line. The right-hand panel of this figure depicts the beam width before and after acceleration as a function of the steering angle and demonstrates the emittance blow-up for these cases.

Betatron mismatch for synchrotron injection. To prevent transverse emittance enlargement, it is important to control the injection into a synchrotron for a single-turn injection. If the orientation of the injected beam emittance is wrong, owing to misaligned focusing in the transfer line, beam storage might still be possible but then the emittance ellipse would start to rotate with an angle per revolution given by the fractional part q of the tune Q of the synchrotron, as shown in Fig. 59. A rotating ellipse is equivalent to a variation in the beam profile from turn to turn. After several hundred turns, the full synchrotron acceptance is filled, owing to the decoherence of the single particle's betatron-phases (e.g., caused by the momentum spread of the particles Δp and the coupling to tune via chromaticity ξ as $\Delta Q/Q_0 = \xi \cdot$

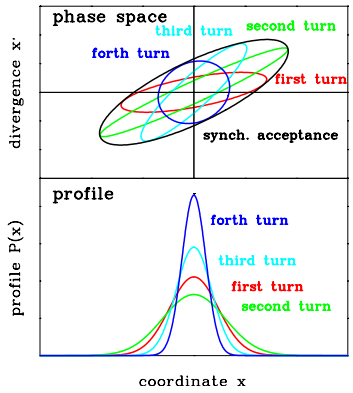


Fig. 59: Schematics for betatron mismatched injection into a synchrotron. If the orientation of the injected beam emittance does not fit the acceptance, variations of the beam profile occur.

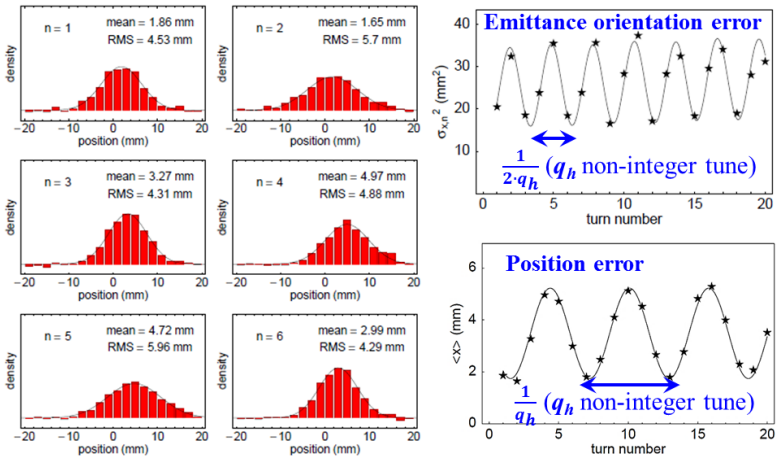


Fig. 60: Left: Profile measurement for a mismatched injection into the CERN PS synchrotron for the first $n = 1, \ldots, 6$ turns. Right: Fitted square of the beam width and centre as a function of time [81].

$\Delta p/p_0$), leading to a larger circulating beam and an emittance blow-up. The mismatch can be determined by recording the profile width on a turn-by-turn basis.

An example is shown in Fig. 60 for a proton injection into the CERN PS synchrotron with 1.4 GeV energy and a bunch length of 80 ns [81, 82]. A thin SEM grid is used and the data are digitized for each revolution of the short bunch. The energy loss of the protons in the wires is too small to influence this measurement during only a few turns. The oscillation of the profile width can be seen; the fitted Gaussian width is plotted as a function of the turn number. The mismatch of the emittance orientation leads to as oscillation of the beam width with a periodicity given by the fractional part of the tune as $1/2q_\mathrm{h}$; because of the symmetric phase space distribution, depicted in Fig. 59, one half rotation in phase space leads to reproduction of the profile. For the depicted case, the mismatch results in a blow-up of the stored emittance by $\sim$10%. The variation of the centre of the transverse profile is caused by a mismatch in terms of dispersion; the centre oscillates with a period of $1/q_\mathrm{h}$. A better injection is created by changing the setting of the quadrupole magnets in the transfer line. Such measurements can also be performed by a turn-by-turn readout of an ionization profile monitor; an example is shown in [83, 84].

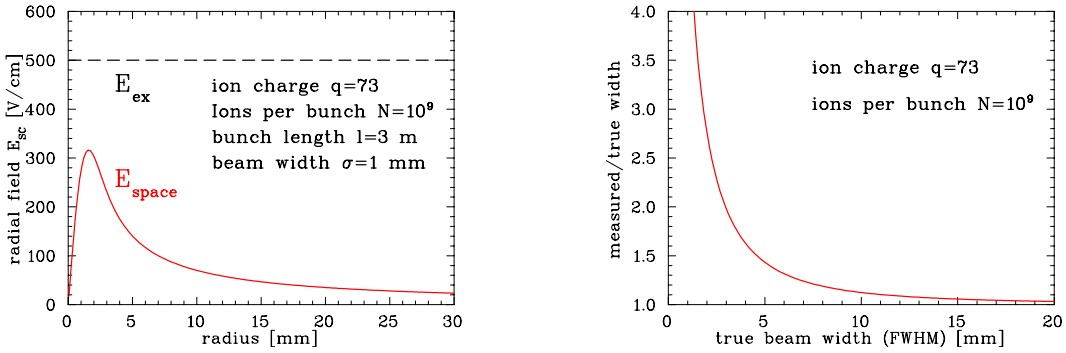


Fig. 61: Left: Space charge field of 7×10^{10} charges per 3 m bunch length. Right: Resulting relative broadening using an ionization profile monitor with ion detection, plotted using Eq. 24.

4.5.1 Ion detection: resolution limit by the beam's space charge

The residual gas ions or electrons are accelerated towards the detector by applying an external electrical field of typically 0.3–3 kV/cm. However, the beam, made of charged particles, also carries an electrical field, the space charge field, which influences the trajectory as well. The radial electrical field E_{space} of a round beam is given non-relativistically for a Gaussian distribution with standard deviation σ (full width at half-maximum = $2.35 \cdot \sigma$) by

$$E_{\mathrm{space}}(r) = \frac{1}{2\pi\epsilon_0} \cdot \frac{qeN}{l} \cdot \frac{1}{r} \cdot \left(1 - \mathrm{e}^{-\frac{r^2}{2\sigma^2}}\right) , \tag{22}$$

assuming equal beam sizes in the horizontal and vertical directions, which leads to a radially symmetrical field, where qeN/l is the number of particles in charge state q per unit length. For high-current operation, this field strength is comparable to the external field, see Fig. 61. The residual gas ions or electrons are accelerated by both fields, leading to a broadening of the measured profile, see Fig. 62. To estimate the real beam size, the correction can be calculated via

$$\sigma^2_{\mathrm{beam}} = \sigma^2_{\mathrm{meas}} - \sigma^2_{\mathrm{corr}} , \tag{23}$$

where σ_{corr} is given to a first-order approximation by the beam space charge and the external field according to [85–88]

$$\sigma^2_{\mathrm{corr}} = \frac{\mathrm{e}^2 \ln 2}{4\pi\epsilon_0 \sqrt{m_{\mathrm{p}} c^2}} \cdot d_{\mathrm{gap}} \cdot qN \cdot \sqrt{\frac{1}{eU_{\mathrm{ex}}}} . \tag{24}$$

The correction is proportional to the number of beam particles N. It is inversely proportional to the square root of the external field (created by the static voltage U_{ex}); therefore, an increase of the external field has only a relatively weak influence on the broadening. A large distance d_{gap} between the two high-voltage electrodes results in a longer drift time and therefore in a prolonged influence of the space charge field. To calculate this correction for a real case with different horizontal and vertical sizes and a longitudinal extension within the bunch, numerical simulations are required, see, e.g., Refs. [87, 88].

The initial velocity of the residual gas ions immediately after the ionizing atomic collision between the residual gas molecule and the beam particle (Bethe–Bloch regime) is too small to affect the resolution (in contrast with the electron detection scheme, see next section).

4.5.2 Electron detection in the presence of a magnetic field

To measure intense beams, electron detection is more suited for overcoming space charge induced broadening, see Fig. 62. For this, the high voltage is reversed, to accelerate the negative electrons created by

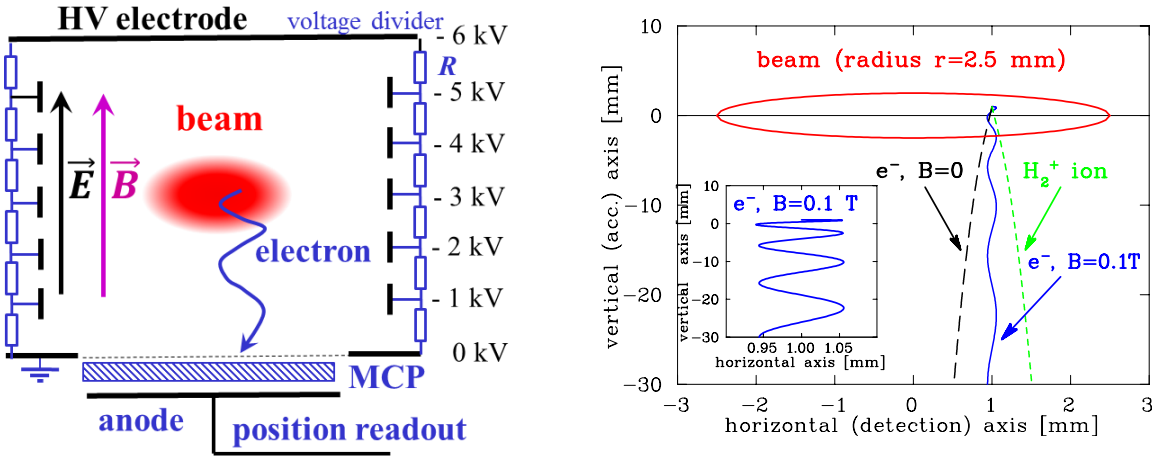


Fig. 62: Left: Trajectory of electron in ionization profile monitor in the presence of a homogeneous electric E and magnetic field B. Right: Comparison of trajectories of H_2^+ ion and electron with and without magnetic field, showing the broadening due to the beam space charge field. By applying a magnetic field, the electron is fixed to the field line, as shown in the inset. Beam parameters: 10^9 U^{73+} within a 10 m long bunch with radius 2.5 mm.

the beam interaction towards the detector. If an additional magnetic field B is applied parallel to the external electrical field, the electrons spiral around the magnetic field lines with the cyclotron radius $r_c = \sqrt{2m_e E_{kin,\perp}}/eB$ with $E_{kin,\perp}$ being the kinetic energy fraction perpendicular to the magnetic field. This transverse kinetic energy is determined by the initial transverse velocity given by the kinematics of the atomic collision. The cross-section depends on the beam and residual gas molecule properties [89], resulting in kinetic energies of up to kiloelectronvolts for the emitted electron, for the typical case. The cyclotron radius, and therefore the detector resolution, is mainly determined by this initial energy and is almost independent of the beam space charge field. The projection of the trajectory in the plane perpendicular to the field is a circle with less than 0.1 mm radius (corresponding to the spatial resolution of the multichannel plate) if a magnetic field of $B \simeq 0.1$ T is applied. The movement along the field is a linear acceleration with a typical time of flight of ~ 2 ns. In most cases, the necessary magnetic field of about 0.1 T is generated using a dipole magnet of the same type as used for the beam bending, see, e.g., Ref. [80].

4.6 Optical transition radiation screens

At electron accelerators with relativistic particles, the profile is determined from the so-called optical transition radiation at an intercepting thin metallic foil. The optical transition radiation is produced in a purely classical electromagnetic process, including special relativity, when a charged particle passes from one medium into another. While passing to the vacuum in front of the foil, the particle has a certain electromagnetic field configuration, which is different from the field inside the medium, because the foil has a (complex) dielectric constant $\epsilon_r \neq 1$, i.e., different from vacuum. On approaching the foil, the particle's electromagnetic field leads to a time-dependent polarization at the foil boundary. When the charged particle traverses the foil, the field configuration is suddenly changed. The change of this polarization at the foil surface generates radiation with a characteristic intensity and angular distribution.

A typical set-up of an optical transition radiation measurement is shown in Fig. 63. The foil is inserted at <45° with respect to the beam path in most cases. The foil is made of aluminium or aluminium coated on Mylar with a thickness of 1 µm or less. The light is emitted in the forward direction as well as at 90°, because the metallic surface acts as a mirror. Depending on the particle energy, the angular distribution peaks at an angle $\theta = 1/\gamma$, where γ is the relativistic Lorentz factor. For typical values, 100 to 1000 beam particles yield one photon in the optical wavelength range. With appropriate optics, an

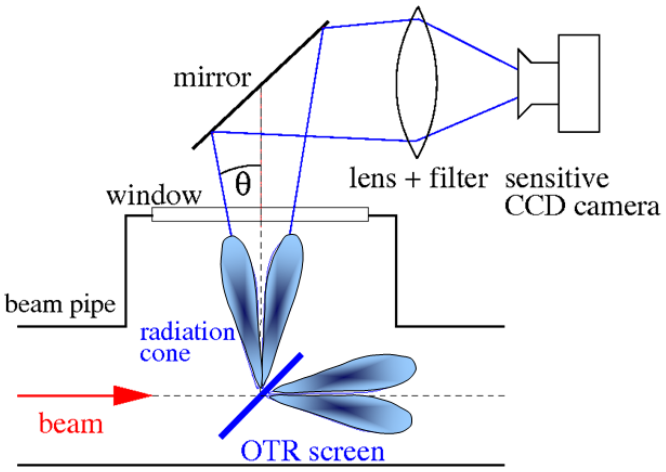


Fig. 63: Optical transition radiation (OTR) screen measurement

image of the foil is recorded with a CCD or CMOS camera. In some cases, an image amplifier is used, owing to the relatively small number of photons.

The general process is treated in, e.g., Refs. [18, 90, 91], leading to closed but extensive formulae. For relativistic beam particles and some other valid assumptions, the radiated energy $\mathrm{d}W$ into a solid angle $\mathrm{d}\Omega$ per frequency interval $\mathrm{d}\omega$ can be approximated as

$$\frac{\mathrm{d}^2 W}{\mathrm{d}\Omega \mathrm{d}\omega} = \frac{2e^2\beta^2}{\pi c} \cdot \frac{\theta^2}{\left(\gamma^{-2} + \theta^2\right)^2} \tag{25}$$

where c, e, and γ are the velocity of light, the elementary charge, and the relativistic Lorentz factor, respectively. The observation is oriented at an angle θ perpendicular to the beam path, the so-called specular angle, see Fig. 63. There is no difference in the radiation for electrons or protons moving with the same Lorentz factor, reflecting the fact that only the electromagnetic fields of the beam particles are concerned. Note that the radiated energy does not depend on the frequency, ω, i.e., the whole spectrum is covered. This is valid up to the plasma frequency of the metal, which, for most metals, is in the deep ultraviolet, at about 10 eV. The radiated energy of Eq. 25 is converted to the number of photons by $W = N_{\mathrm{photon}} \cdot \hbar\omega$, observed within a wavelength interval from λ_{begin} to λ_{end} in the optical region by the camera and integration over the interval λ_{begin}–λ_{end}. This yields the number of photons per solid angle

$$\frac{\mathrm{d}N_{\mathrm{photon}}}{\mathrm{d}\Omega} = N_{\mathrm{beam}} \cdot \frac{2e^2\beta^2}{\pi\hbar c} \cdot \log\left(\frac{\lambda_{\mathrm{begin}}}{\lambda_{\mathrm{end}}}\right) \cdot \frac{\theta^2}{\left(\gamma^{-2} + \theta^2\right)^2}\,, \tag{26}$$

where N_{beam} is the number of beam particles. This function is plotted in Fig. 64 for three different values of γ. It is clearly seen that the radiation is more tightly focused for higher energies (more precisely, higher values of the Lorentz factor γ), having the advantage that a larger proportion of the photons reach the camera. At electron accelerators, optical transition radiation is used even at moderate energies above $\simeq$100 MeV (corresponding to a Lorentz factor $\gamma \simeq 200$). For proton acceleration with a Lorentz factor $\gamma \simeq 200$ or equivalent, a beam energy for protons above 200 GeV is only achieved at some high-energy facilities, such as CERN SPS and LHC, where optical transition radiation screens are installed [92, 93].

The optical transition radiation profile determination has some advantages and disadvantages compared with a scintillating screen measurement [52, 53, 58].

– It is based on a classical electromagnetic process, leading to a linearity between the number of photons and the beam intensity.

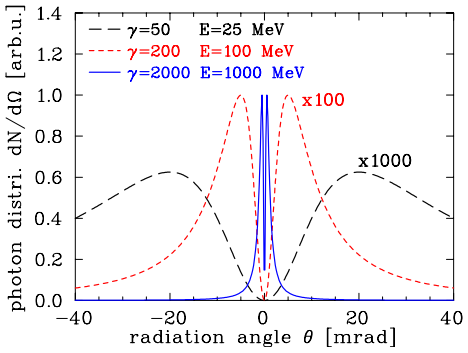


Fig. 64: Intensity distribution of optical transition radiation as a function of observation angle for three different values of the Lorentz factor γ. Corresponding beam energies for electrons are also given. The photon intensity is enhanced by a factor of 1000 for $\gamma = 50$ and 100 for $\gamma = 200$ in the plot.

- The number of photons and their distribution are independent of the thickness of the foil. Therefore, very thin foils of pure aluminium foil or aluminium coated on Mylar down to 0.25 µm can be used.
- It is still an intercepting method, but a thin foil minimizes the influence on the beam, owing to the low scattering probability. It can also be applied at high beam power, because of the low energy loss in the thin, low-Z material.
- It is mainly used for electron beams with energies above 100 MeV. For proton beams, the necessary γ is only reached at very-high-energy accelerators.
- The intensity of usable light using an optical transition radiation screen is orders of magnitude less than for a scintillation screen. Therefore, more advanced and expensive cameras must be used.

4.7 Synchrotron radiation monitor

For electron accelerators, the effect of synchrotron radiation, emitted by accelerated electrons, can be used for profile determination. As known from classical electrodynamics [18], the radiation power P_{synch} is given for a momentum change $\mathrm{d}p/\mathrm{d}t$ of a particle with mass m_0 by

$$P_{\mathrm{synch}} = \frac{e^2 c}{6\pi\varepsilon_0 (m_0 c^2)^2} \left[\frac{\mathrm{d}p}{\mathrm{d}t}\right]^2 . \tag{27}$$

For a circular accelerator, the direction of the momentum is changed in the bending dipole magnets, leading to the emission of synchrotron radiation, which is significant only for relativistic energies with a Lorentz factor above at least $\gamma > 100$. Therefore, the process is of importance for most circular electron accelerators but only for proton accelerators with an energy above some 100 GeV.

In the centre-of-mass system, the radiation is emitted perpendicular to the momentum change. The Lorentz transformation to the laboratory frame gives a factor of γ, yielding a forward peak distribution with opening cone of half angle γ, as demonstrated in Fig. 65. The physical basis, the achievable resolution, and some applications are reviewed in Ref. [94].

The light emitted by the electron's bend in a dipole magnet can be used to monitor its horizontal and vertical beam profile, as shown schematically in Fig. 66. For diagnostic purposes, the optical part of the emitted spectrum is observed, in most cases, by using optical band-pass filters. For this wavelength, high-quality optics are available and standard CCD or CMOS cameras can be used. A typical resolution limit for the related optics is of the order of ∼100 µm, and is related to the diffraction pattern created at the finite-size lenses and filters [94, 95]. A higher resolution can be achieved by observing the light from wigglers or undulators, if they are installed in the synchrotron.

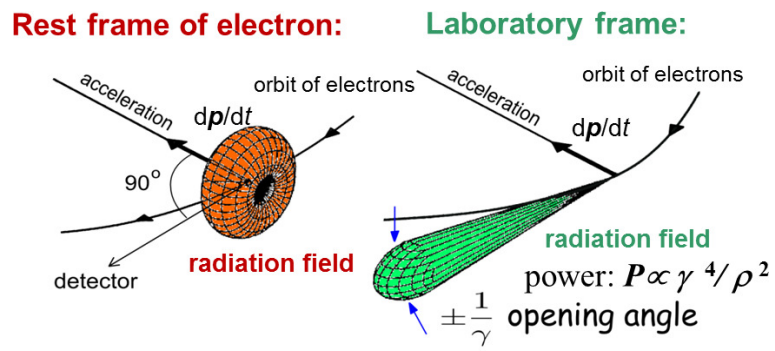


Fig. 65: Forward peaked synchrotron radiation for relativistic particles from a bending magnet

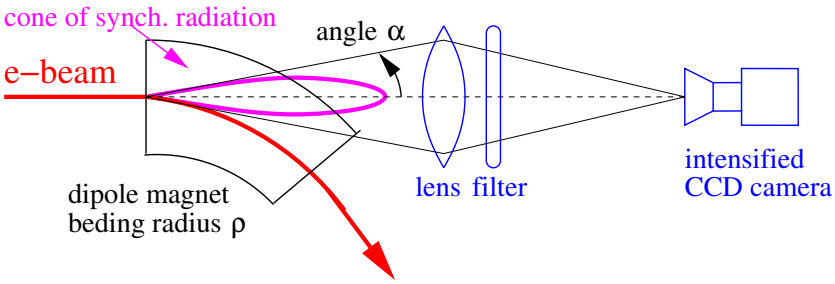


Fig. 66: Scheme for a synchrotron radiation profile monitor to observe the radiation from a dipole

A realization of a synchrotron light profile monitor used at CERN LEP is shown in Fig. 67 [96]. The bending radius here is 3.1 km; diffraction gives the largest contribution to the resolution, by about $\sigma \sim 300$ µm, which is comparable to the real electron beam size close to the final energy. The set-up consists of a metallic mirror to deflect the light out of the plane of the beam. Owing to the high power of the synchrotron radiation, the mirror must be cooled; possible deformations can spoil the optical quality. With the help of some curved mirrors, the light is focused; an optical filter selects the wavelength range of blue or ultraviolet light; the detector is an image-intensified CCD camera. Even though the resolution is not too high, this relatively straightforward system can be quite helpful, owing to its non-destructiveness, as demonstrated in Fig. 68 from the synchrotron light source APS [97]. Such synchrotron light monitors can also be installed at dipoles in transfer lines, as realized at some facilities, e.g., Ref. [98]. However, there might be some technical problems, as the amount of light emitted by the single pass of the electrons might be insufficient; instead optical transition radiation screens, owing to their simpler technical realization, serve as frequently used profile diagnostics in transfer lines at third-generation light sources, while scintillation screens are used for low beam current applications at other electron accelerators.

An example of the application of a synchrotron radiation monitor during acceleration at the ALBA booster synchrotron is shown in Fig. 69 [99, 100]. For the two-dimensional beam image, the widths in the horizontal and vertical directions can be calculated and compared with simulations. At first glance, the adiabatic damping is less pronounced than in the example of the hadron beam shown in Fig. 56: the reason is the contribution to the beam width of the relative momentum spread $\Delta p/p_0$ of the beam particles and the non-zero dispersion D at the measurement location,

$$x_{\mathrm{tot}} = \sqrt{\varepsilon\beta + \left(D \cdot \Delta p/p_0\right)^2}\,,$$

and the contribution of the quantum fluctuations related to the emission of photons [99, 100].

The diffraction limit of the standard monitor can be compensated for by using an interference technique with a double slit, known from astronomy. The beam width can be be calculated from the

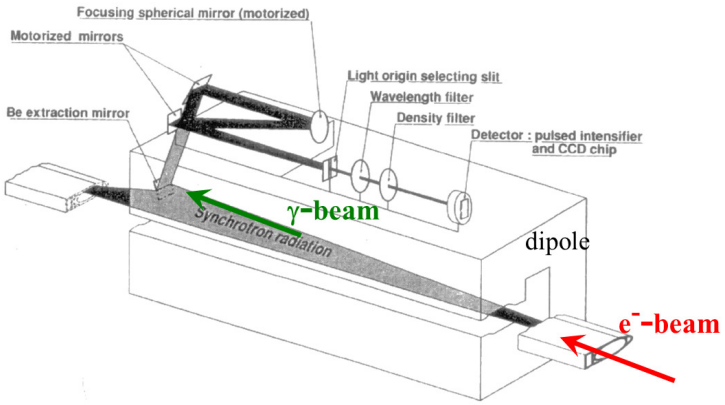


Fig. 67: Synchrotron radiation profile monitor at LEP. The optical system is installed close to the dipole magnet with bending radius of 3100 m [96].

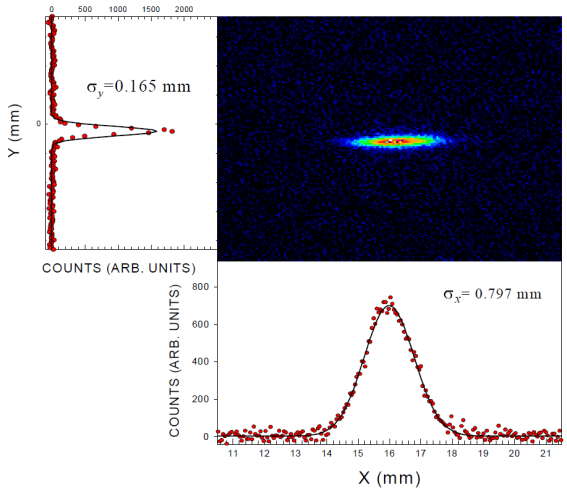


Fig. 68: Image of electron beam in the APS accumulator ring, obtained using synchrotron radiation monitor [97]

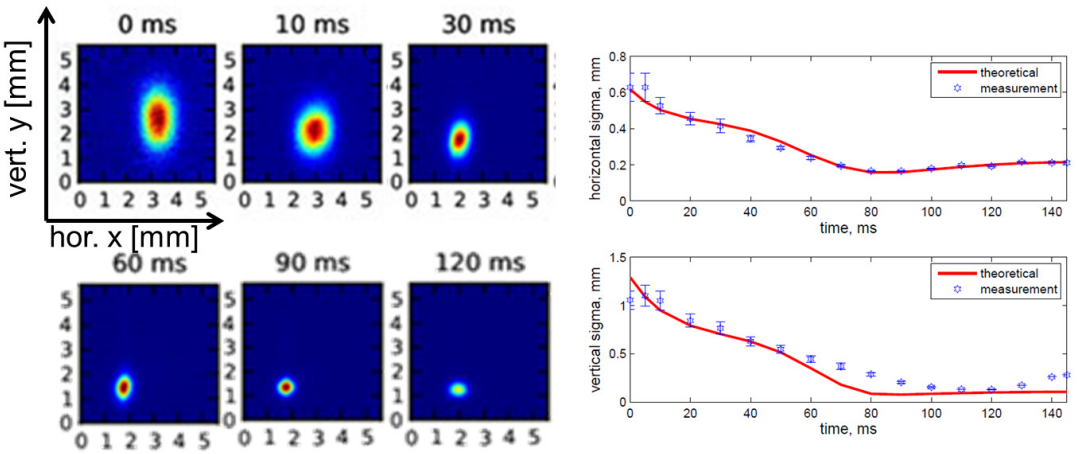


Fig. 69: Left: Examples of two-dimensional beam images obtained using a synchrotron light monitor at the ALBA booster synchrotron. Right: Fitted horizontal and vertical beam widths during acceleration from 0.1 to 3 GeV within 130 ms and comparison with simulations [99, 100].

distance of the minima of the interference pattern. Owing to non-coherent emission by an ensemble of electrons, the fringes fade out for large beam sizes. A resolution down to the micrometre range has been realized with this method in combination with elaborate analysis tools, see, e.g., Refs. [94, 101, 102].

A higher resolution can also be achieved by monitoring the profile at much shorter wavelengths; X-ray pinhole cameras are used for this purpose. Only the emitted X-rays, scraped by a typically 20 µm aperture of high-Z material are recorded by a detector. A typical resolution, concerning the beam width, of 10 µm can be obtained using such an X-ray observation scheme [94] but this requires a more complex technical realization than for the observation in the optical range.

5 Measurement of transverse emittance

The emittance describes the quality of a beam. Its determination is based on profile measurements; it is unimportant what method of profile measurement is used, as long as it has an adequate resolution. In the following, we start with a slit-grid device, in which the spatial co-ordinate is fixed by an aperture and the angle distribution is measured. This is suited for particles with a range in matter below $\sim$1 cm, i.e., protons or ions with $E_{\mathrm{kin}} < 100$ MeV/u. The emittance can also be determined by fitting the beam envelope measured at one location with different focusing conditions or at different locations. This can be applied to transfer lines for all particle conditions. However, here it is problematic to include emittance growth due to space charge forces, i.e., blow-up of the transverse size due to the forces between the charged particles. A detailed review of the involved physics, technologies, and data acquisition is given in Ref. [103].

In a synchrotron, it is sufficient to measure the beam profile at one location only. For the stationary state of stable storage, the orientation of the ellipse, which is equivalent to knowledge of the lattice function dispersion $D(s)$ and β function $\beta(s)$, is fixed (or can be measured separately). The emittance is calculated from the beam width σ as

$$\epsilon = \frac{1}{\beta(s)} \left[\sigma^2 - \left(D(s) \frac{\Delta p}{p} \right)^2 \right] , \tag{28}$$

which is valid for the horizontal and vertical planes, respectively. If only horizontal bends occur in a facility, the dispersion in the vertical direction is, in most cases, $D_y = 0$, and Eq. (28) simplifies to $\epsilon_y = \sigma_y^2 / \beta_y(s)$.

5.1 Definition of emittance

The motion of the beam particle can be described by a linear second-order differential equation. This assumes the absence of any non-linear coupling, such as space charge forces or beam–beam effects, as well as coupling between the two transverse planes and the longitudinal plane. The beam quality is given by the phase space volume, which, in this case, is a constant of the motion. The emittance for one plane is defined by

$$\epsilon_x = \frac{1}{\pi} \int_A \mathrm{d}x \mathrm{d}x', \tag{29}$$

where $A = \pi\epsilon$ is the area of the phase space occupied by the beam, see Fig. 70. Determination of the emittance is equivalent to determination of the distribution of the spatial co-ordinate x (i.e., the beam profile), the distribution in angle x' and the correlation between x and x'.

The interpretation of an area assumes a hard-edge, homogeneous distribution. A more realistic case is the Gaussian density distribution $\rho(x, x')$, defined at the position of the vector $\vec{x} = (x, x')$ for each location s along the beam path. The two-dimensional density is then (the notation is defined later)

$$\rho(x, x') = \frac{1}{2\pi\epsilon_x} \exp\left[\frac{-1}{2} \vec{x}^{\mathrm{T}} \boldsymbol{\sigma}^{-1} \vec{x} \right] \equiv \frac{1}{2\pi\epsilon_x} \exp\left[\frac{-1}{2 \det\sigma} \left(\sigma_{22} x^2 - 2\sigma_{12} x x' + \sigma_{11} x'^2 \right) \right] . \tag{30}$$

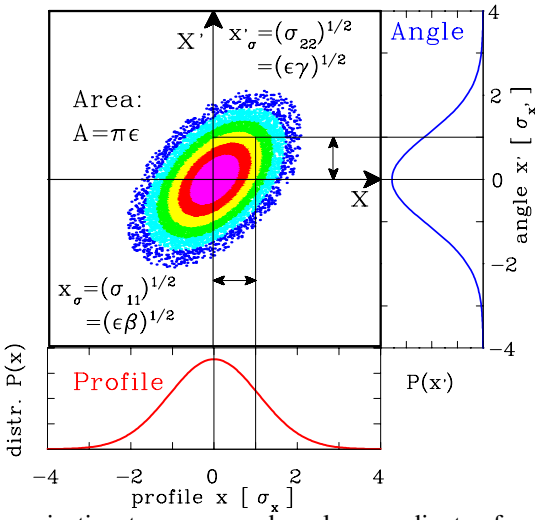


Fig. 70: Emittance ellipse and projection to space- and angle co-ordinates for a Gaussian density distribution. Values of the independent variables are given in units of the standard deviation.

The profile distribution $P(x)$ is obtained by integrating the density ρ over x' as

$$P(x) = \int \rho(x, x') \mathrm{d}x' \,.$$

The parameter $\sqrt{\sigma_{11}}$ is the standard deviation of this distribution; $\sqrt{\sigma_{22}}$ is the corresponding value for the angular distribution $P(x')$ obtained by integrating $\rho(x, x')$ over x as $P(x') = \int \rho(x, x') \mathrm{d}x$. σ_{12} is called the covariance; it describes the correlation between x and x' and is related to the orientation of the ellipse in the phase space. With the help of the three-parameter σ_{ij}, the so-called beam matrix $\boldsymbol{\sigma}$ can be defined as

$$\boldsymbol{\sigma}(s) = \begin{pmatrix} \sigma_{11}(s) & \sigma_{12}(s) \\ \sigma_{12}(s) & \sigma_{22}(s) \end{pmatrix} \tag{31}$$

at one location s in the beam line. The beam matrix $\boldsymbol{\sigma}$ is a representation of the beam ellipse at this location s and varies along the beam path. It is used in connection with the transfer matrix $\mathbf{R}$, which describes the action of the optical elements, such as drifts, dipoles, quadrupole magnets, or solenoids, along the beam path in a linear approximation, as

$$\boldsymbol{\sigma}(s_1) = \mathbf{R} \cdot \boldsymbol{\sigma}(s_0) \cdot \mathbf{R}^{\mathrm{T}} \tag{32}$$

from a location s_0 to s_1. See Section 5.3 for a more detailed discussion.

The absolute value of the emittance at each location can be defined using this notation as

$$\epsilon_x = \sqrt{\det \boldsymbol{\sigma}} = \sqrt{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \tag{33}$$

and corresponds to an occupation of 15% of the full phase space area. The unit of this value is the metre radian or, more frequently, the millimetre milliradian. To be consistent with the geometrical interpretation of an area surrounded by an ellipse, the number is multiplied by π; in this case, 39% of the beam is inside the area

$$\pi \cdot \sqrt{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \,.$$

In other words, the Gaussian beam quality at a given location s is fully described by the beam matrix $\boldsymbol{\sigma}(s)$.

Frequently, the Twiss parameters are used; these are the beam matrix elements normalized by the emittance as

$$\alpha = -\sigma_{12}/\epsilon \,, \qquad \beta = \sigma_{11}/\epsilon \,, \qquad \gamma = \sigma_{22}/\epsilon \,. \tag{34}$$

Table 11: Fraction of beam inside different values of emittance for Gaussian distribution

Factor of ϵ_{rms}	$1 \cdot \epsilon_{\mathrm{rms}}$	$\pi \cdot \epsilon_{\mathrm{rms}}$	$2\pi \cdot \epsilon_{\mathrm{rms}}$	$4\pi \cdot \epsilon_{\mathrm{rms}}$	$6\pi \cdot \epsilon_{\mathrm{rms}}$	$8\pi \cdot \epsilon_{\mathrm{rms}}$
Fraction of beam f [%]	15	39	63	86	95	98

The beam matrix is then

$$\boldsymbol{\sigma} = \epsilon \cdot \begin{pmatrix} \beta & -\alpha \\ -\alpha & \gamma \end{pmatrix} \tag{35}$$

and the equation of the beam ellipse can be written as

$$\gamma x^2 + 2\alpha x x' + \beta x'^2 = \epsilon\,, \quad \text{with the normalization} \quad \beta\gamma - \alpha^2 = 1. \tag{36}$$

The widths of the profile and angular distribution are given by

$$x_\sigma = \sqrt{\sigma_{11}} = \sqrt{\epsilon\beta} \quad \text{and} \quad x'_\sigma = \sqrt{\sigma_{22}} = \sqrt{\epsilon\gamma}\,. \tag{37}$$

Their geometric meaning is one standard deviation of the transverse profile and angular distribution of the beam. The geometrical size of the phase space ellipse changes along the beam pass s; therefore, the parameters $\alpha(s)$, $\beta(s)$ and $\gamma(s)$ are functions of the position s. In particular, for a synchrotron, $\beta(s)$ is called the beta function, describing the beam size via $x_\sigma(s) = \sqrt{\epsilon \cdot \beta(s)}$.

For theoretical calculations, one prefers analytical descriptions of the density distribution. An example is the Gaussian distribution of Eq. 30; other functions are discussed, e.g., in Ref. [104]. However, the beam does not always have a Gaussian shape. For any arbitrary phase space distribution, the beam emittance can be calculated via the statistical moments of a two-dimensional distribution $\rho(x, x')$.

To describe the beam quality via the emittance, the r.m.s. value can be calculated as

$$\epsilon_{\mathrm{rms}} = \sqrt{\det \begin{pmatrix} \langle x^2 \rangle & \langle x x' \rangle \\ \langle x x' \rangle & \langle x'^2 \rangle \end{pmatrix}} = \sqrt{\langle x^2 \rangle \langle x'^2 \rangle - \langle x x' \rangle^2} \tag{38}$$

and follows the same general mathematical role as Eq. 33.

For a Gaussian distribution, the fraction f of the beam inside a certain value of the emittance can be expressed via the r.m.s. value as

$$\epsilon(f) = -2\pi\epsilon_{\mathrm{rms}} \cdot \ln(1 - f)\,, \tag{39}$$

leading to Table 11.

When a certain value of emittance is given for a beam, one must check its meaning carefully, because the different accelerator facilities do not share a common definition for the emittance around or of what beam fraction for a given emittance value is meant.

The emittance is a quantity defined in the laboratory frame, owing to the definition of the divergence. When a beam is accelerated, the divergence shrinks ('adiabatic damping') as a result of the changing ratio of longitudinal velocity v_s to transverse velocity v_x: $x' = v_x/v_s$. To compare the emittance for different longitudinal momenta, $p_s = m_0 \cdot \gamma_{\mathrm{rel}} \cdot v_s$, the normalized emittance ϵ_{norm}

$$\epsilon_{\mathrm{norm}} = \frac{v_s}{c}\gamma_{\mathrm{rel}} \cdot \epsilon \tag{40}$$

is referred to a value $v_s/c \cdot \gamma_{\mathrm{rel}} = 1$, where c is the velocity of light and $\gamma_{\mathrm{rel}} = 1/\sqrt{1 - (v_s/c)^2}$ is the relativistic Lorentz factor. The normalized emittance is constant under ideal accelerating conditions.

A measurement of emittance means a determination of the numerical value of ϵ, as well as of the orientation and shape of the phase space distribution.

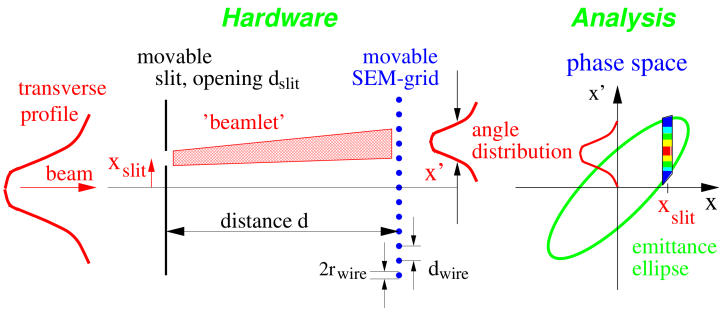


Fig. 71: Scheme of slit-grid emittance measurement device

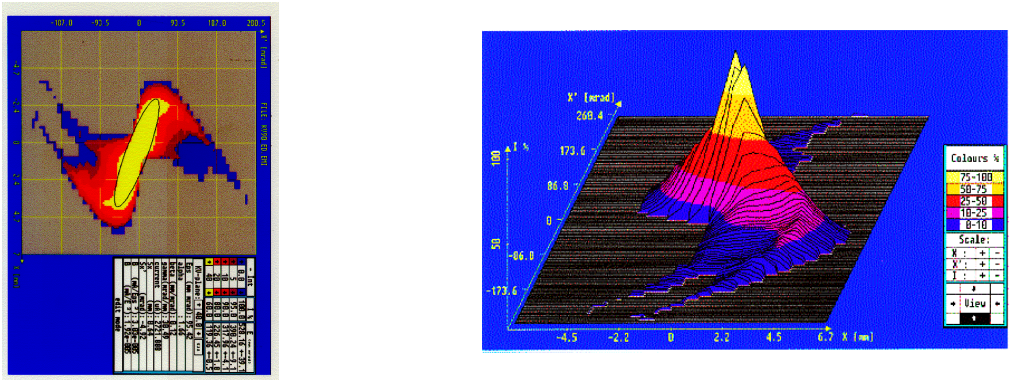


Fig. 72: Emittance measurement using a slit-grid device with a low-energy ion beam after the source at GSI display. Left: contour plot. Right: mountain plot.

5.2 Slit-grid method

A popular method at proton or ion linacs is the slit-grid device depicted in Fig. 71, where the beam particles have a penetration depth less than 1 cm. Here, the position x is fixed for one direction, with a thin slit having an opening of typically 0.1–0.5 mm to filter out only a small fraction of the beam at a known location. In the perpendicular direction, the full beam is transmitted to get a large signal. The angle x' is determined using a SEM grid having a distance from the slit of 10 cm to 1 m, depending on the ion velocity. In the field-free drift space, the trajectories of the particles, combined in a 'beamlet', are straight lines [105, 106]. The contribution to the emittance plot in the phase space is given by the angle distribution at the slit location. The slit is then scanned through the beam to determine the emittance at all positions. After making the full scan, the emittance is plotted and the r.m.s. value of the emittance ϵ_{rms} is calculated using the statistical moments, as given in Eq. 38 from the measured data. An elliptical shape is fitted to the data and the Twiss parameters can be calculated. An example is shown in Fig. 72 for a low-energy ion beam as a contour or three-dimensional plot. This method can also be used to determine more pathological phase space distributions, not only Gaussian distributions. This happens quite often close to an ion source, owing to the large space charge forces or the large profile width, where aberrations of the magnets could be significant.

The resolution for the space co-ordinate Δx is limited by the slit width $\Delta x = d_{\mathrm{slit}}$. The angle resolution $\Delta x'$, measured at the distance d, is given by the radius of the wire r_{wire} and the width of the slit, resulting in $\Delta x' = (d_{\mathrm{wire}} + 2r_{\mathrm{wire}})/d$. The size of discrete elements in the phase space is given by $\Delta x \cdot \Delta x'$. This leads to a discretization error, in particular, in the case of small beam sizes (focused beam) or small angle distributions (parallel beam). The resolution is improved by scanning the SEM grid in steps that are smaller than the distance of the wires d_{wire}, increasing the density of the discrete elements in the phase space analysis. This can lead to overlapping elements, because their size $\Delta x \cdot \Delta x'$ remains constant. The same holds for a movement of the slit with step sizes smaller than the slit width.

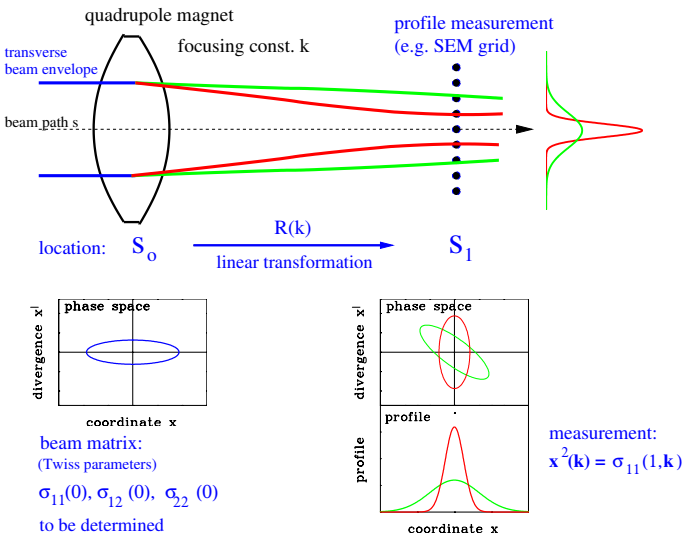


Fig. 73: Variation of quadrupole strength for determination of emittance at location s_0

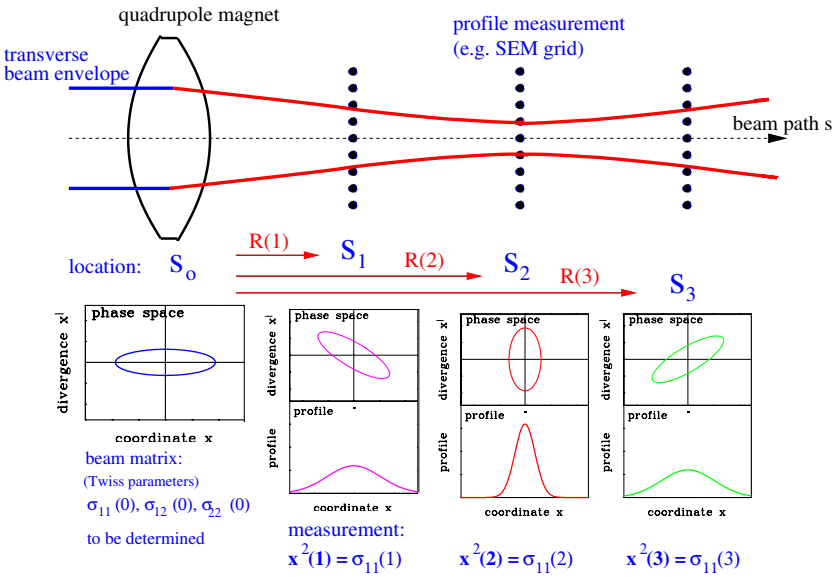


Fig. 74: Scheme for the emittance determination at a transfer line using profile measurement at three locations

5.3 Quadrupole variation

The emittance can be determined from a series of profile measurements, either by changing the focusing strength of a quadrupole, or by measuring at three or more positions along a transfer line, as shown schematically in Figs. 73 and 74. To derive the emittance from such a measurement, linear transformations are used. The described measurement is simply the reverse of normal transport calculations using such codes as MADX or WinAgile: Starting for a given emittance (size and orientation) at the input of a transport line, the beam envelope is calculated along the line. If the profile is measured, one calculates the emittance at the input of the line by using the same transformations in reverse (see later).

For a straight, non-dispersive transfer line, the transformation from a location s_0 to s_1 is given by the 2×2 transfer matrix $\mathbf{R}$. A determination of the emittance at the position s_0 is equivalent to the evaluation of the beam matrix $\boldsymbol{\sigma}$ at this position. The beam matrix is transformed to a second location s_1 with the help of the product of transfer matrices for the individual elements $\mathbf{R} = \prod \mathbf{R}_{\text{elements}}$ from the

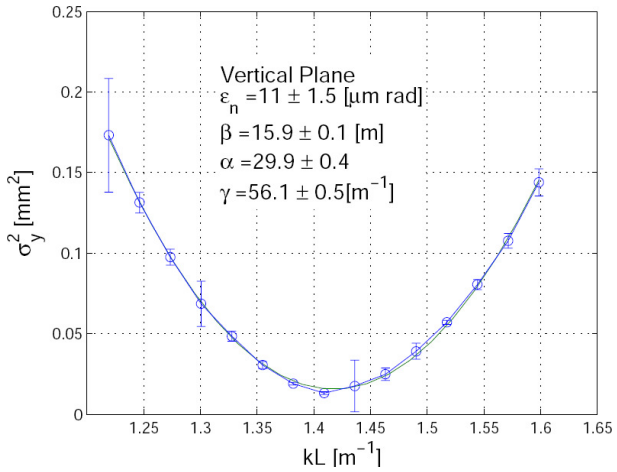


Fig. 75: Profile width determined using a YAG:Ce screen at Elettra electron linac for a quadrupole variation and parabolic fit [107].

quadrupole to the profile measurement location via

$$\boldsymbol{\sigma}(1) = \mathbf{R} \cdot \boldsymbol{\sigma}(0) \cdot \mathbf{R}^{\mathrm{T}} \tag{41}$$

The beam width $x_{\mathrm{rms}}(1)$ is measured at s_1 and the equation for the element $\sigma_{11}(1)$ is given by

$$x^2_{\mathrm{rms}}(1) \equiv \sigma_{11}(1) = R^2_{11}\sigma_{11}(0) + 2R_{11}R_{12}\sigma_{12}(0) + R^2_{12}\sigma_{22}(0)\,. \tag{42}$$

This is a linear equation for the unknown three beam matrix elements $\sigma_{ij}(0)$ at location s_0, as shown in Fig. 73, in front of the focusing quadrupole magnet.

To get a solution, we need at least three different settings of the quadrupole strength k_i, and therefore different transfer matrices $\mathbf{R}(k_i)$ leading to three different readings of the profile width, as depicted in Fig. 73. Assuming $i = 1, 2, \ldots, n$ different settings of the quadrupole strength $k_1, k_2, \ldots, k_n$ and n measurements of the beam width $x^2_{\mathrm{rms}}(1, k_i) \equiv \sigma_{11}(1, k_i)$, a redundant system of linear equations is obtained in the form

$$\begin{aligned}
\sigma_{11}(1,k_1) &= R^2_{11}(k_1)\cdot\sigma_{11}(0) + 2R_{11}(k_1)R_{12}(k_1)\cdot\sigma_{12}(0) + R^2_{12}(k_1)\cdot\sigma_{22}(0)\,, \quad \text{focusing } k_1\,,\\
\sigma_{11}(1,k_2) &= R^2_{11}(k_2)\cdot\sigma_{11}(0) + 2R_{11}(k_2)R_{12}(k_2)\cdot\sigma_{12}(0) + R^2_{12}(k_2)\cdot\sigma_{22}(0)\,, \quad \text{focusing } k_2\,,\\
&\vdots\\
\sigma_{11}(1,k_n) &= R^2_{11}(k_n)\cdot\sigma_{11}(0) + 2R_{11}(k_n)R_{12}(k_n)\cdot\sigma_{12}(0) + R^2_{12}(k_n)\cdot\sigma_{22}(0)\,, \quad \text{focusing } k_n\,.
\end{aligned} \tag{43}$$

The solutions of this system are the values of the beam matrix $\sigma_{ij}(0)$ (or, equivalently, the Twiss parameters) at the location s_0, the entrance of the quadrupole magnet. With these values, the size and orientation of the phase space ellipse is fixed. For three measurements (k_1, k_2, k_3), we can have a unique solution, but nothing is learnt about the errors. Therefore, more than three measurements must be made, leading to a redundant system of linear equations. The solution is reached by a least square fit to the best parameters of $\sigma_{ij}(0)$ or by solving the linear regression problem via so-called normal equations. Both algorithms are described in textbooks on linear algebra or numerical mathematics.

An example is shown in Fig. 75, obtained at the electron linac at Elettra for a beam with 107 MeV using a YAG:Ce scintillation screen for profile determination. To get a small error for the emittance determination, it is recommended to pass a beam waist using quadrupole variation. Only in this case can a parabola be fitted to the profile data with sufficient accuracy.

Instead of solving the redundant system of linear equations, one can start from the parabola fit of the beam size squared as a function of the quadrupole strength, as shown in Fig. 75. Most frequently, the transfer line comprises a quadrupole followed by a drift towards the profile measurement location. A quadratic dependence is observed for the following reason. Assuming a thin-lens approximation with a focal length f of the quadrupole action, one can write the transfer matrix as:

$$\mathbf{R}_{\text{focus}} = \begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ K & 1 \end{pmatrix} . \tag{44}$$

After a drift of length L, one gets the transfer matrix of the transfer line

$$\mathbf{R} = \mathbf{R}_{\text{drift}} \cdot \mathbf{R}_{\text{focus}} = \begin{pmatrix} 1+LK & L \\ K & 1 \end{pmatrix} . \tag{45}$$

Inserting this matrix into Eq. 41 for the transformation of the beam matrix $\boldsymbol{\sigma}(1) = \mathbf{R} \cdot \boldsymbol{\sigma}(0) \cdot \mathbf{R}^{\mathrm{T}}$, one gets, for the measured beam matrix element, $\sigma_{11}(1) = x_{\text{rms}}^2(1)$:

$$\sigma_{11}(1) = L^2\sigma_{11}(0) \cdot K^2 + 2\left(L\sigma_{11}(0) + L^2\sigma_{12}(0)\right) \cdot K + \sigma_{11}(0) + 2L\sigma_{12}(0) + L^2\sigma_{22}(0) \quad . \tag{46}$$

This is the expected quadratic function in the quadrupole gradient K. From the beam width measurement for various quadrupole settings, a parabola fit is performed, as for Fig. 75, with the fit parameters a, b and c (this parametrization is chosen to yield a simple expression for the absolute value of the emittance, as shown later for Eq. 50):

$$\sigma_{11}(K) = a\,(K-b)^2 + c = aK^2 - 2abK + ab^2 + c. \tag{47}$$

A comparison of the coefficients of Eqs. 46 and 47 yields

$$\begin{aligned} a &= L^2\sigma_{11}(0) \\ -ab &= L\sigma_{11}(0) + L^2\sigma_{12}(0) \\ ab^2 + c &= \sigma_{11}(0) + 2L\sigma_{12}(0) + L^2\sigma_{22}(0) \quad . \end{aligned} \tag{48}$$

These linear equations can finally be solved, resulting in the requested matrix elements at the entrance of the optical system:

$$\begin{aligned} \sigma_{11}(0) &= \frac{a}{L^2} \\ \sigma_{12}(0) &= -\frac{a}{L^2}\left(\frac{1}{L} + b\right) \\ \sigma_{22}(0) &= \frac{1}{L^2}\left(ab^2 + c + \frac{2ab}{L} + \frac{a}{L^2}\right). \end{aligned} \tag{49}$$

The absolute value of the emittance is calculated from these fit parameters via

$$\epsilon = \sqrt{\det \boldsymbol{\sigma}(0)} = \sqrt{\sigma_{11}(0)\sigma_{22}(0) - \sigma_{12}^2(0)} = \frac{\sqrt{ac}}{L^2} \quad . \tag{50}$$

Be aware that the fit parameters have some units, as given in Eq. 47.

In this discussion, it is assumed that the method is performed at a location where no dispersion D is present. The lattice function dispersion D gives rise to an additional contribution to the beam width Δx through the momentum deviation Δp of the particles with respect to the mean momentum p_0 as $\Delta x = D \cdot \Delta p/p_0$. The dispersion effect is added in a quadratic manner to the measured beam width,

as $x_{\mathrm{rms}}^2 = \sigma_{11}(0) + (D \cdot \Delta p/p_0)^2$. To avoid measurement of the lattice function D and the beam's longitudinal momentum spread, the location for emittance determination should be dispersion-free.

The described method is based on linear transformations and conservation of emittance during manipulation by the quadrupole. Moreover, an elliptical shape of the emittance is assumed. Depending on the general beam properties, this is a good assumption, as long as non-linear effects, such as aberration of magnets by higher-order field contributions or space charge forces are small. If the investigations are conducted with intense beams, an emittance blow-up may occur, depending on the particle density, i.e., on the beam size, but the size is to be changed and even a waist must be created. For high-intensity beams, self-consistent, iterative algorithms must be applied to obtain an estimate of the emittance at the quadrupole entrance, see, e.g., Refs. [108, 109], and references therein.

5.4 'Three-grid' method

Instead of varying a quadrupole, the beam profiles at different locations along a transfer line can be measured with fixed magnet settings. This is easier to achieve at long transfer lines, where the profile monitors are installed, in any case. There is no general restriction to the involved optics, as long as they can be described by linear optics, e.g., dipoles can be included with their known transfer matrix $\mathbf{R}_{\mathrm{dipole}}$.

The profile must be measured at three or more locations, see Fig. 74. For good accuracy, one profile monitor should be close to a beam waist. To determine the beam matrix elements $\sigma_{ij}(0)$ at the entrance of the line s_0, the profile widths and, therefore, the beam matrix elements $\sigma_{11}(n)$ are determined at n different locations s_n. A system of redundant linear equations for the entrance emittance $\sigma_{ij}(0)$ is given by

$$\begin{aligned}
\sigma_{11}(1) &= R_{11}^2(1) \cdot \sigma_{11}(0) + 2R_{11}(1)R_{12}(1) \cdot \sigma_{12}(0) + R_{12}^2(1) \cdot \sigma_{22}(0)\,, \qquad \mathbf{R}(1) : s_0 \rightarrow s_1 \\
\sigma_{11}(2) &= R_{11}^2(2) \cdot \sigma_{11}(0) + 2R_{11}(2)R_{12}(2) \cdot \sigma_{12}(0) + R_{12}^2(2) \cdot \sigma_{22}(0)\,, \qquad \mathbf{R}(2) : s_0 \rightarrow s_2 \\
&\vdots \\
\sigma_{11}(n) &= R_{11}^2(n) \cdot \sigma_{11}(0) + 2R_{11}(n)R_{12}(n) \cdot \sigma_{12}(0) + R_{12}^2(n) \cdot \sigma_{22}(0)\,, \qquad \mathbf{R}(n) : s_0 \rightarrow s_n. \qquad (51)
\end{aligned}$$

This formula is comparable to the quadrupole variation (Eq. 43), but the transfer matrix $\mathbf{R}(i)$ now describes the beam transport from the starting point of the transfer line s_0 to the location of the individual measurements s_i. The algorithm of the emittance calculation (least square fit or normal equations) is the same as for the quadrupole scan. Such a measurement is shown in Fig. 76, compared with the fitted beam envelope, to enable visualization of possible errors. In practice, the parameters of the entrance ellipse are fitted to match the measured profiles using a linear optics code (e.g., MADX, WinAgile) for this transfer line.

To obtain a reliable measurement, the transfer matrix $\mathbf{R}(n)$ should be an adequate description of the action of the optical elements on the beam. In particular, one must ensure that the beam centre passes the quadrupoles along their magnetic axes, i.e., no 'steering' of the quadrupoles is present. This might be a source of errors in practical situations. As discussed before, no emittance growth, e.g., owing to space charge forces, is assumed here.

The algorithm is not limited to straight transport lines. If a dipole magnet is also present, one must take its action on the beam into account. This results in a dispersion D, i.e., a coupling between the momentum spread $\delta = \Delta p/p$ and the transverse beam size, with $\Delta x = D \cdot \Delta p/p$. For the transport matrix, we get an additional matrix element R_{13}. Also, a dipole has an influence on the angular distribution, called the angular dispersion, resulting in a additional matrix element R_{23}. The single-particle trajectory is now made of a three-dimensional vector $\vec{x} = (x, x', \Delta p/p)$ and the transfer matrix $\mathbf{R}$ is now a 3×3 matrix. The effect of dispersion must be included in the symmetric beam matrix σ, with three new elements, σ_{13}, σ_{23} and σ_{33}. To determine the values of these beam matrix elements, at least

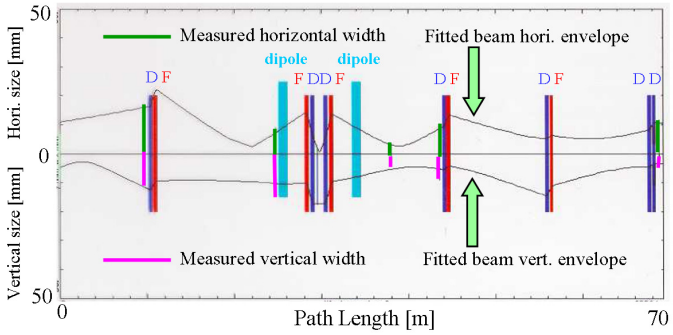


Fig. 76: Determination of beam envelope by linear transformations using different profile measurements. The width of the profile is shown as green lines, measured at a transfer line at GSI. Top: Horizontal direction. Bottom: Vertical direction.

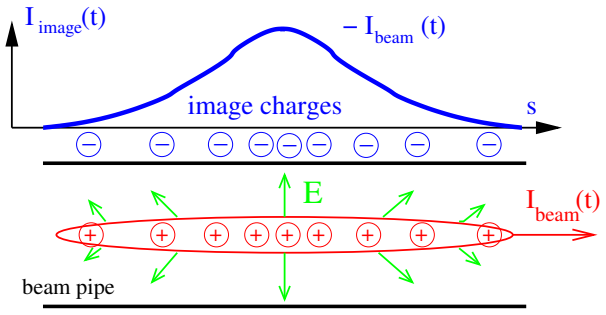


Fig. 77: The beam current induces a wall current of the same magnitude and time structure but reversed polarity

six profile measurements must be made, resulting in at least six equations in the (redundant) system of linear equations for $\sigma_{ij}(0)$ in Eqs. 43 and 51.

6 Beam position monitors

The longitudinal bunch shape, as well as the position of a beam, is very frequently determined using pick-up plates. The idea is to measure, on an insulated metal plate, the image current induced by the electromagnetic field of the beam particles, see Fig. 77. Because the electric field of a bunched beam is time-dependent, an a.c. signal is seen on the plate and the coupling is achieved using RF technologies. Only time-varying signals can be detected by this principle, as generated by a bunched beam. By contrast with bunched beams, the position for a d.c. beams, such as, e.g., behind an ion source in the low-energy transfer lines or beams slowly extracted from a synchrotron, cannot be detected by this method; in this case, the position is determined from the profile measurement. The principle signal shape, as well as the most often used types of pick-up are described. The application is the determination of the beam position, i.e., the transverse centre of mass of the bunches. To this end, four pick-up plates are installed and the difference of the signals yields the centre of mass in both transverse axes. The device with the plates is called the pick-up; the corresponding installation for the centre-of-mass determination is referred to as a beam position monitor (BPM). Using this instrument, the beam position is determined in transfer lines and in synchrotrons. Various measurements based on this position information are possible but are only discussed briefly at the end of this section.

6.1 Signal treatment of capacitive pick-ups

As shown in Fig. 78, a capacitive pick-up consists of a plate or ring inserted in the beam pipe. Here the induced image charge of the beam is coupled via an amplifier for further processing. The plate, at a

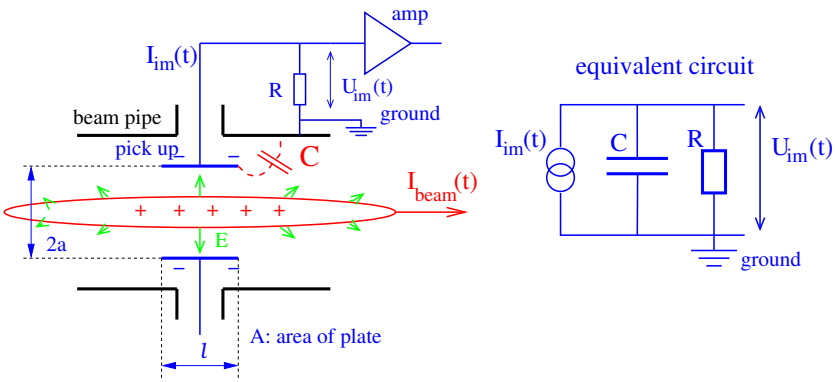


Fig. 78: Scheme of a pick-up electrode and its equivalent circuit

distance a from the beam centre, has an area A and a length in the longitudinal direction l. The current I_{im} driven by the image charge Q_{im} as a function of time t is

$$I_{\mathrm{im}}(t) \equiv \frac{\mathrm{d}Q_{\mathrm{im}}}{\mathrm{d}t} = -\frac{A}{2\pi a l} \cdot \frac{\mathrm{d}Q_{\mathrm{beam}}(t)}{\mathrm{d}t}. \tag{52}$$

Having a beam with velocity β, we can write, for the derivative of the beam charge $\mathrm{d}Q_{\mathrm{beam}}(t)/\mathrm{d}t$,

$$\frac{\mathrm{d}Q_{\mathrm{beam}}(t)}{\mathrm{d}t} = \frac{l}{\beta c}\frac{\mathrm{d}I_{\mathrm{beam}}}{\mathrm{d}t} = -\frac{l}{\beta c} \cdot \mathrm{i}\omega I_{\mathrm{beam}}(\omega)\,, \tag{53}$$

where the beam current is expressed in the frequency domain using the angular frequency ω as $I_{\mathrm{beam}} = I_0 \mathrm{e}^{-\mathrm{i}\omega t}$. (More precisely, the derivative of a function $\mathrm{d}f/\mathrm{d}t$ can be expressed as a multiplication of its Fourier transformation $\tilde{f}(\omega)$ by $-\mathrm{i}\omega$.) Combining Eqs. 52 and 53, the image current can be written as

$$I_{\mathrm{im}}(t) = \frac{1}{\beta c} \cdot \frac{A}{2\pi a} \cdot \mathrm{i}\omega I_{\mathrm{beam}} \quad . \tag{54}$$

As the signal, we use the voltage drop across a resistor R:

$$U_{\mathrm{im}}(\omega) = R \cdot I_{\mathrm{im}}(\omega) = Z_{\mathrm{t}}(\omega, \beta) \cdot I_{\mathrm{beam}}(\omega) \quad . \tag{55}$$

For all types of pick-up plate, the general quantity of longitudinal transfer impedance $Z_{\mathrm{t}}(\omega, \beta)$ is defined by Eq. 55 in the frequency domain according to Ohm's law; for a more detailed discussion, see Refs. [110–113]. The transfer impedance describes the effect of the beam on the pick-up voltage and depends on frequency, the velocity of the beam particles β and geometrical factors. It is very helpful to make the description in the frequency domain, where the independent variable is the angular frequency ω, related to the time domain by Fourier transformation.

The frequency-domain function $\tilde{f}(\omega)$ is calculated from the time-domain function $f(t)$ by Fourier transformation as

$$\tilde{f}(\omega) = \int_{-\infty}^{\infty} f(t) \cdot \mathrm{e}^{-\mathrm{i}\omega t} \mathrm{d}t \quad . \tag{56}$$

By inverse transformation, the time-domain function can be recovered as

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{f}(\omega) \cdot \mathrm{e}^{\mathrm{i}\omega t} \mathrm{d}\omega\,, \tag{57}$$

if the given integral for $\tilde{f}(\omega)$ exists. Normally, the absolute value, $|\tilde{f}(\omega)|$, also called the amplitude or magnitude, and the phase, $\tan\varphi = \mathrm{Re}(\tilde{f})/\mathrm{Im}(\tilde{f})$, of this complex (in a mathematical sense) function are displayed.

The pick-up of Fig. 78 has a certain capacitance C, which is determined by the distance of the plate from the beam pipe and a capacitance contributed by the cable between the plate and the amplifier input. This amplifier has an input resistance R. Using a current source to model the beam and the parallel connection of the equivalent circuit, we can write its impedance Z as

$$\frac{1}{Z} = \frac{1}{R} + \mathrm{i}\omega C \quad \Longleftrightarrow \quad Z = \frac{R}{1+\mathrm{i}\omega RC}\,. \tag{58}$$

Inserting this impedance into Ohm's law, as given in Eq. 55, and using the expression for the image current of Eq. 54, the measurable voltage of the pick-up is

$$U_{\mathrm{im}} = \frac{R}{1+\mathrm{i}\omega RC}\cdot I_{\mathrm{im}} = \frac{1}{\beta c}\cdot\frac{1}{C}\cdot\frac{A}{2\pi a}\cdot\frac{\mathrm{i}\omega RC}{1+\mathrm{i}\omega RC}\cdot I_{\mathrm{beam}} \equiv Z_{\mathrm{t}}(\omega,\beta)\cdot I_{\mathrm{beam}}\,, \tag{59}$$

which results in a general transfer impedance

$$Z_{\mathrm{t}}(\omega,\beta) = \frac{1}{\beta c}\cdot\frac{1}{C}\cdot\frac{A}{2\pi a}\cdot\frac{\mathrm{i}\omega RC}{1+\mathrm{i}\omega RC}\quad . \tag{60}$$

This is a description of a first-order high-pass filter for the transfer impedance $Z_{\mathrm{t}}(\omega,\beta)$ with a cut-off frequency $f_{\mathrm{cut}} = \omega_{\mathrm{cut}}/2\pi = (2\pi RC)^{-1}$. For the so-called linear-cut pick-ups used at proton synchrotrons (see later), a typical value of the capacitance is $C = 100$ pF, with a length $l = 10$ cm. The high-pass characteristic is shown in Fig. 79 for a 50 Ω and a high-impedance 1 MΩ amplifier input resistor. In the figure, the absolute value

$$|Z_{\mathrm{t}}| = \frac{1}{\beta c}\cdot\frac{1}{C}\cdot\frac{A}{2\pi a}\cdot\frac{\omega/\omega_{\mathrm{cut}}}{\sqrt{1+\omega^2/\omega_{\mathrm{cut}}^2}} \quad \text{and the phase relation} \quad \varphi = \arctan\left(\omega_{\mathrm{cut}}/\omega\right) \tag{61}$$

are shown. A pick-up has to match the interesting frequency range, which is given by the accelerating frequency and the bunch length. In a proton synchrotron, typical values of the accelerating frequency are in the range 1–10 MHz, while for linacs and electron synchrotrons, frequencies in the range 100 MHz to 3 GHz are typically used.

We can distinguish two extreme cases for the transfer impedance of Eq. 60.

– *High-frequency range* $\omega \gg \omega_{\mathrm{cut}}$. Here, the transfer impedance converges to

$$Z_{\mathrm{t}} \propto \frac{\mathrm{i}\omega/\omega_{\mathrm{cut}}}{1+\mathrm{i}\omega/\omega_{\mathrm{cut}}} \longrightarrow 1. \tag{62}$$

The resulting voltage drop at R is, for this case,

$$U_{\mathrm{im}}(t) = \frac{1}{\beta cC}\cdot\frac{A}{2\pi a}\cdot I_{\mathrm{beam}}(t). \tag{63}$$

Therefore, the pick-up signal is a direct image of the bunch time structure with no phase shift, i.e., $\varphi = 0$. To get a low cut-off frequency, $\omega_{\mathrm{cut}} = 1/RC$, high-impedance input resistors are used to monitor long bunches, e.g., in a proton synchrotron. The calculated signal shape is shown in the right-hand panel of Fig. 80. Note that a 50 Ω termination is considered in Fig. 80, leading to a large value of the cut-off frequency, $f_{\mathrm{cut}} = 32$ MHz. In the application of a proton synchrotron, high-impedance (∼1 MΩ) termination yields a much smaller value of the cut-off frequency, $f_{\mathrm{cut}} =$ 10 kHz in this case, as shown in Fig. 79, where the condition corresponds to the enlarged flat part of the depicted transfer impedance. A typical signal from this circuit is given in Fig. 81.

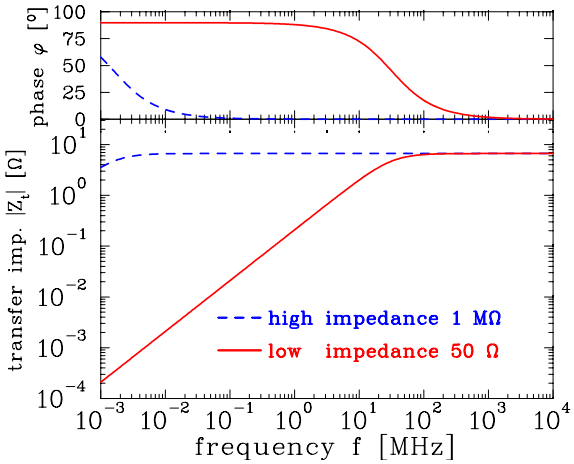


Fig. 79: Absolute value and phase of the transfer impedance for an $l = 10$ cm long cylindrical pick-up with a capacitance of $C = 100$ pF and an ion velocity of $\beta = 50\%$ for high (1 MΩ) and low (50 Ω) input impedances of the amplifier.

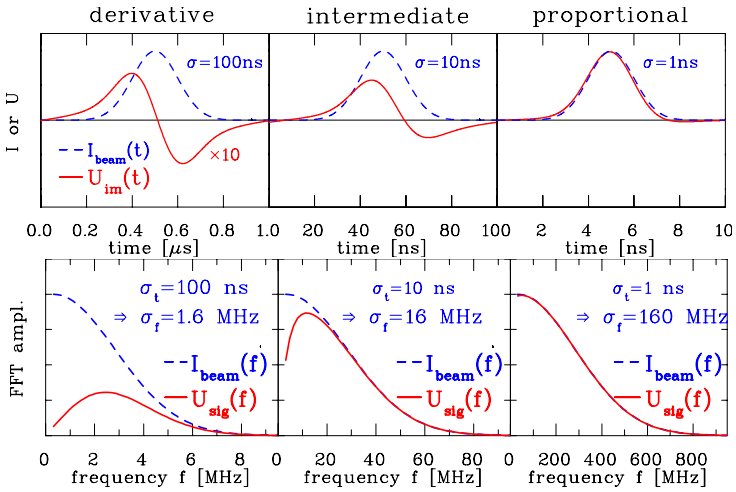


Fig. 80: Simulation of image voltage $U_{\mathrm{im}}(t)$ for the values of the pick-up used in Fig. 79 terminated with $R = 50\,\Omega$ for three different bunch lengths of Gaussian distribution, with $\sigma = 100$ ns, 10 ns, and 1 ns, respectively. The cut-off frequency f_{cut} is 32 MHz. Note the different time scales. (The bunch length in the last case is artificially short for a proton synchrotron.) A Gaussian function in the time domain of width σ_t has a Fourier transformation described by a Gaussian function of width $\sigma_f = 1/(2\pi\sigma_t)$, centred at $f = 0$.

– *Low-frequency range* $\omega \ll \omega_{\mathrm{cut}}$. The transfer impedance here is

$$Z_{\mathrm{t}} \propto \frac{\mathrm{i}\omega/\omega_{\mathrm{cut}}}{1 + \mathrm{i}\omega/\omega_{\mathrm{cut}}} \longrightarrow \mathrm{i}\frac{\omega}{\omega_{\mathrm{cut}}}. \tag{64}$$

Therefore, the voltage across R is, in this case,

$$U_{\mathrm{im}}(t) = \frac{R}{\beta c} \cdot \frac{A}{2\pi a} \cdot \mathrm{i}\omega I_{\mathrm{beam}} = -\frac{R}{\beta c} \cdot \frac{A}{2\pi a} \cdot \frac{\mathrm{d}I_{\mathrm{beam}}}{\mathrm{d}t}\,, \tag{65}$$

again using the frequency-domain relation $I_{\mathrm{beam}} = I_0 \mathrm{e}^{-\mathrm{i}\omega t}$. We see that the measured voltage is proportional to the derivative of the beam current. This can also be seen from the phase relation of the high-pass filter in Fig. 79, where a phase shift of 90° corresponds to a derivative. The signal

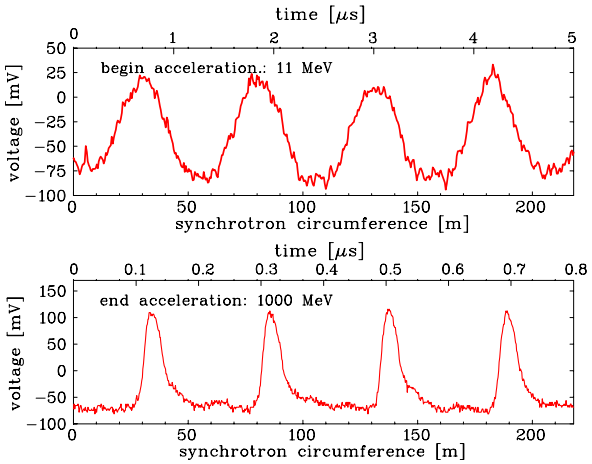


Fig. 81: Bunch signals from a 'shoebox' pick-up for 1 MΩ termination, as installed at the GSI ion synchrotron. Top: Bunches of the proton beam along the synchrotron circumference at the start of the acceleration. Bottom: Bunches of the proton beam along the synchrotron circumference after attaining the final energy. Note the different time scales.

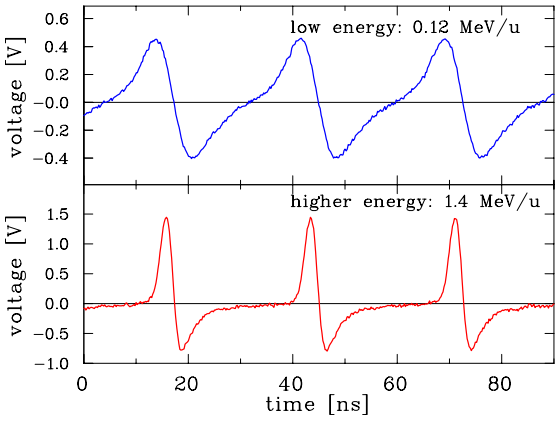


Fig. 82: Bunch signal from a capacitive pick-up at the GSI ion linac. Top: 0.12 MeV/u. Bottom: 1.4 MeV/u. The derivative behaviour is caused by 50 Ω termination to achieve a large bandwidth of the amplifier. The accelerating frequency is 36 MHz.

is bipolar, as shown in the left-hand panel of Fig. 80. A measurement from the GSI linac, which fulfils this condition, is shown in Fig. 82.

The signal at the amplifier output depends on the frequency range, as compared with the cut-off frequency. Of course, the bunches are not pure sine waves; therefore, it might be that a mixture of proportional and derivative components are present, see the middle panel of Fig. 80. In general, the signal shape, as depicted in Fig. 80, can be calculated for the different parameters using the transfer impedance concept by the following steps.

- The beam current $I_{\mathrm{beam}}(t)$ is Fourier transformed, yielding $\tilde{I}_{\mathrm{beam}}(\omega)$.
- This is then multiplied by $Z_{\mathrm{t}}(\omega)$, yielding the frequency-dependent voltage $\tilde{U}_{\mathrm{im}}(\omega) = Z_{\mathrm{t}}(\omega) \cdot \tilde{I}_{\mathrm{beam}}(\omega)$.
- To retrieve the time-dependent signal, the inverse Fourier transformation is applied, yielding $U_{\mathrm{im}}(t)$.

With respect to the different limits of the transfer impedance, two important applications are discussed for illustration.

– *Range $\omega \gg \omega_{\rm cut}$ realized by low $\omega_{\rm cut}$ due to high impedance.* In a proton or ion synchrotron, a low frequency is used for acceleration in the range of $f_{\rm acc} = 1$–10 MHz, resulting in bunches of several metres in length. In these machines, large apertures are, necessarily, typically $a = 10$ cm, reducing the transfer impedance due to $Z_{\rm t} \propto 1/a$. To increase sensitivity, the length of the pick-up in the beam direction is increased, typically $l \sim 10$ cm as $Z_{\rm t} \propto l$. Note that the pick-up length is still much shorter than the bunch length. To have a flat curve of the sensitivity, i.e., a large bandwidth, a low value of $\omega_{\rm cut} = 1/RC$ is used by feeding the signal into a high-impedance field-effect transistor as the first step of the amplifier. A bandwidth of the amplifier circuit of 100 MHz can be achieved, but it is difficult to build amplifier circuits with larger bandwidths. To prevent signal degeneration due to the limited amplifier bandwidth, the application of high-impedance amplifiers is restricted to proton or ion synchrotrons with acceleration frequencies less than ∼10 MHz. Observation of the contribution in the bunch spectrum from the tenth harmonic of the acceleration frequency is sufficient for most applications. Figure 81 shows that a direct image of the bunch is seen in this case.
– *Range $\omega \ll \omega_{\rm cut}$ realized by high $\omega_{\rm cut}$ due to $50\ \Omega$ impedance.* At proton linacs and all electron accelerators, the bunches are short, up to the millimetre range, owing to the higher accelerating, $f_{\rm acc} = 100$ MHz to 3 GHz. Firstly, one does not gain signal strength by increasing the length l. Secondly, a $50\ \Omega$ termination is used to prevent reflections and achieve smooth signal processing with a large bandwidth, up to several gigahertz. The short l, and therefore smaller capacitance C, along with $R = 50\ \Omega$, leads to a high value of $\omega_{\rm cut}$; the derivative of the bunches is seen, as displayed in Fig. 82.

The foregoing discussion does not take into account the effect of the frequency response of the associated cables or other band-limiting electronics; see Refs. [111, 114] for a more detailed discussion. Further on, the fact that a pick-up has a high-pass characteristic leads to the fact that a d.c. beam cannot be detected (d.c. beam detectors are available, e.g., behind the ion source or for a transfer line during slow extraction).

6.2 Characteristics for position measurement by BPMs

The deviation of the beam centre with respect to the centre of the vacuum chamber is frequently monitored using four isolated plates or buttons by determining the voltage difference $\Delta U_x = U_{\rm right} - U_{\rm left}$ or $\Delta U_y = U_{\rm up} - U_{\rm down}$ of opposite plates. A shorter distance from one of the plates leads to a higher induced voltage. This is called the 'proximity effect'; it is shown schematically in Fig. 83. Normalizing to the total signal $\Sigma U_x = U_{\rm right} + U_{\rm left}$, the horizontal displacement x can be obtained via

$$x = \frac{1}{S_x} \cdot \frac{U_{\rm right} - U_{\rm left}}{U_{\rm right} + U_{\rm left}} = \frac{1}{S_x} \cdot \frac{\Delta U_x}{\Sigma U_x} \quad \text{(horizontal)}\,, \tag{66}$$

which is independent of the beam intensity. For the vertical plane, the position y is given by

$$y = \frac{1}{S_y} \cdot \frac{U_{\rm up} - U_{\rm down}}{U_{\rm up} + U_{\rm down}} = \frac{1}{S_y} \cdot \frac{\Delta U_y}{\Sigma U_y} \quad \text{(vertical)}. \tag{67}$$

This position measurement is the most frequent application of pick-ups; hence, they are called beam position monitors (BPMs). The proportional constant S_x (or S_y) between the measured normalized voltage difference and the beam displacement is called the position sensitivity and its unit is %/mm. Sometimes, the inverse is used, $k = 1/S$; this is given in units of millimetres and is called the BPM position constant or position sensitivity. It is possible that the position sensitivity depends on the beam position itself, corresponding to a non-linear voltage response for a large beam displacement and, additionally, on the evaluation frequency; hence, it is a function $S = S(x, y, \omega)$.

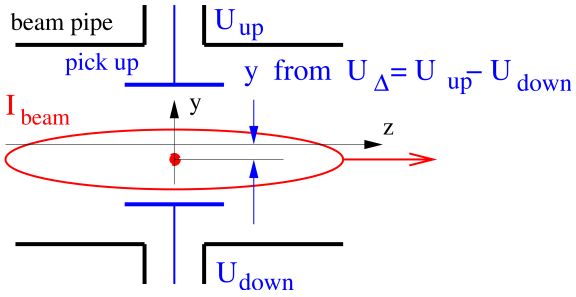


Fig. 83: Schematics of BPM for vertical position reading, based on the proximity effect

For typical beam displacements of less than one-tenth of the beam pipe aperture, the difference ΔU is less by about this factor than the sum voltage ΣU, i.e.,

$$\text{typically} \quad \Delta U < \frac{\Sigma U}{10} \quad . \tag{68}$$

Sensitive signal processing is required for the difference voltage, in order to achieve a sufficient signal-to-noise ratio. This concerns the usage of low-noise amplifiers, which must be matched to the signal level. Sometimes, difference and sum voltages are generated by analogue means using so-called Δ–Σ hybrids. For these cases, the difference voltage can be more greatly amplified than the sum signal (typically amplification by 10 dB corresponds a factor of three for the voltage amplification) to match the optimal signal level for the successive electronics or analogue-to-digital conversion. The most effective noise reduction is achieved by limitation of the signal processing bandwidth, because the thermal noise voltage U_{eff} at a resistor R scales with the square root of the bandwidth Δf as $U_{\text{eff}} = \sqrt{4k_{\text{B}}TR\Delta f}$, where k_{B} is the Boltzmann constant and T is the temperature. The bandwidth limitation is performed by band-pass filtering at a harmonics of this frequency or mixing with the accelerating frequency in narrowband processing, as discussed in Section 6.5. A review of general BPM properties and BPM types is given in Refs. [110–113].

6.3 Position measurement using button pick-ups

For a round arrangement of buttons, a simple two-dimensional electrostatic model can be used to calculate the voltage difference as a function of beam displacement. As shown in Fig. 84, we assume a thin, 'pencil' beam having a transverse extension much smaller than the beam pipe radius of current I_{beam}, which is located off-centre by distance r at an angle θ. The wall current density j_{im} at the beam pipe, of radius a, is given as a function of the azimuthal angle ϕ as

$$j_{\text{im}}(\phi) = \frac{I_{\text{beam}}}{2\pi a} \cdot \left(\frac{a^2 - r^2}{a^2 + r^2 - 2ar \cdot \cos(\phi - \theta)} \right) \tag{69}$$

and is depicted in Fig. 84; see Ref. [115] for a derivation of this formula.

As discussed, this represents the proximity effect, where the current density depends on the distance from the beam centre. The button-type BPM electrodes cover an angle α; the image current I_{im} is recorded as

$$I_{\text{im}} = a \int_{-\alpha/2}^{+\alpha/2} j_{\text{im}}(\phi) \mathrm{d}\phi \quad . \tag{70}$$

The resulting signal difference for opposite plates as a function of horizontal beam displacement, corresponding to $\theta = 0^\circ$, shows a significant non-linear behaviour, as displayed in Fig. 85. The signal voltage is calculated according to Eq. 55 as $U = R \cdot I_{\text{im}}$. It can be seen that, for the normalized difference

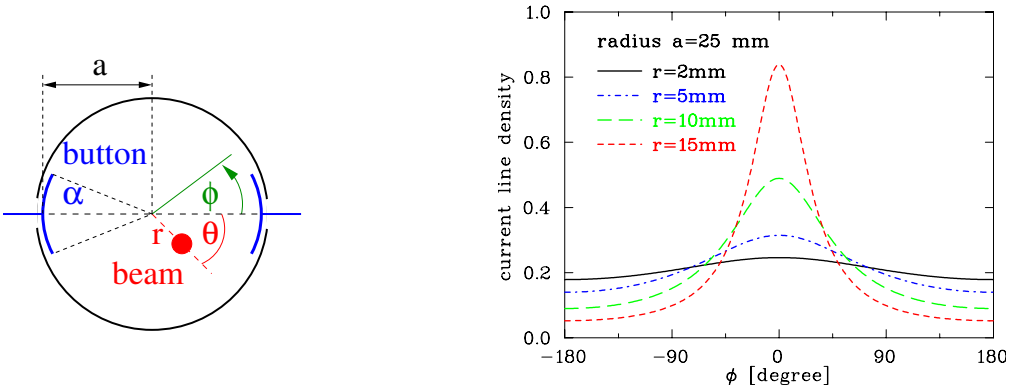


Fig. 84: Left: Schematics for a button BPM. Right: Image current density generated by a 'pencil' beam at different displacements r for an azimuth $\theta = 0$.

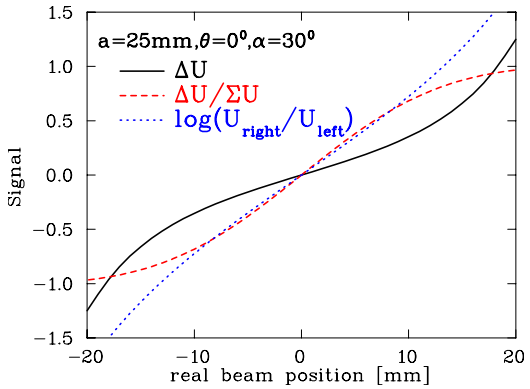


Fig. 85: Difference voltage, normalized difference and logarithmic ratio for a button BPM arrangement of angular coverage $\alpha = 30^\circ$ as a function of horizontal beam displacement, i.e., $\theta = 0$ for a beam pipe radius $a = 25$ mm.

$\Delta U/\Sigma U$, the linear range continues to larger beam offsets. The non-linearity increases if the beam centre moves outside the horizontal axis, as depicted in Fig. 86 for different values of the azimuthal orientation θ as a function of horizontal displacement $x = r \cdot \cos\theta$, according to Eqs. 69 and 70. For an orientation along the diagonal line $\theta = 45^\circ$, a significant deviation from linearity starts for this case, even at about one-quarter of the beam pipe radius. This non-linearity can be influenced by the button size, as represented in this two-dimensional approach by the angle α. However, in the central part, the reading is almost independent of orientation, leading to a universal position sensitivity S. The dependence between the horizontal and vertical planes is better depicted in the right-hand panel of Fig. 86. Here, real beam positions are plotted at equidistant steps, as well as the results obtained using $1/S \cdot \Delta U/\Sigma U$ calculations with S fitted at the central part. The preceding two-dimensional electrostatic model delivers satisfying result for typical electron beams of relativistic beam velocities, i.e., TEM-like field pattern; only minor corrections are necessary in the case of a circular pick-up.

If the beam is accelerated to a non-relativistic velocity, $\beta \ll 1$, and has a bunch length less than 1 cm, as is often the case for proton linacs and cyclotrons, the electric field can no longer be described by a TEM wave. Instead, the field pattern, as originated by the charged particles, has a significant longitudinal extension because the electric field propagation is faster than the beam velocity (see the discussion in Section 7.1). Moreover, this field pattern, and therefore the wall current, depends strongly on the beam displacement for opposite BPM plates, resulting in velocity dependence of the position sensitivity $S(v)$. This is discussed analytically in Ref. [116]; a numerical example is given in Ref. [117, 118].

A BPM arrangement with circular beam pipe geometry, installed at CERN LHC, is depicted in Fig. 87. Owing to the size and the round electrodes, this arrangement is commonly called a button BPM.

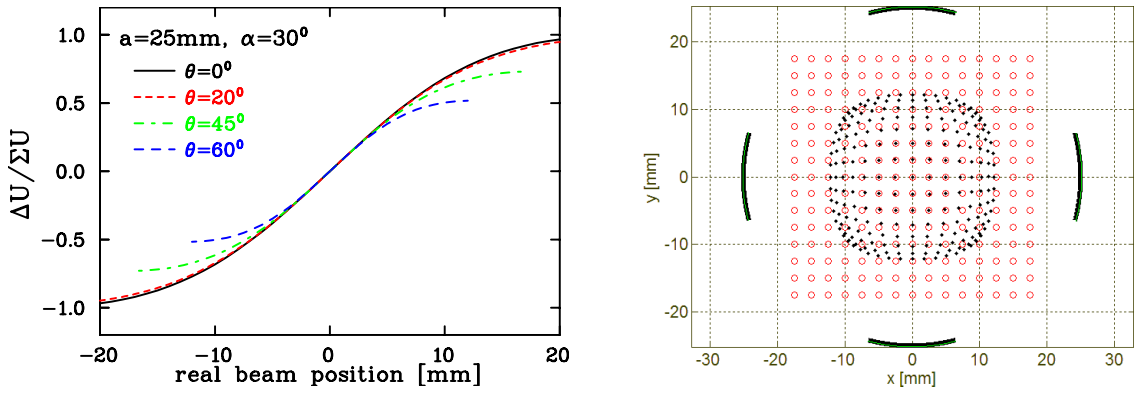


Fig. 86: Left: Calculated horizontal position for different azimuthal beam orientation θ for the parameters of Fig. 85. Right: In the position map, the open circles represent the real beam position and the dots are the results of the $1/S \cdot \Delta U/\Sigma U$, with $S = 7.4$ %/mm for the central part.

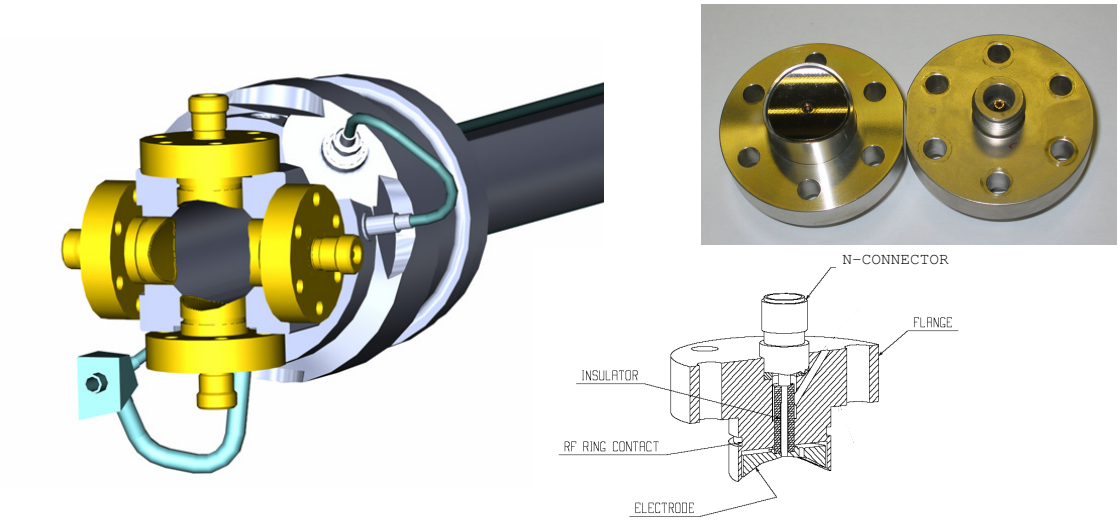


Fig. 87: Left: Installation of curved ⌀24 mm button BPMs at the LHC beam pipe of ⌀50 mm [119, 120]. Right: Photograph and technical drawing of BPM used at LHC: the air side is equipped with a N-connector.

The reason for such an installation is to ensure a smooth transition from the regular beam pipe to the BPM region and to prevent excitation of an electromagnetic field, the so-called wakefield, by the beam. In most other cases, planar buttons are used, owing to their simpler mechanical realization.

Button pick-ups are the most popular devices for electron accelerators. A button pick-up consists of a circular plate, typically 10 mm diameter, mounted flush with the vacuum chamber. The cross-section of the chamber is not changed by this insertion, to avoid excitation of wakefields by the beam. The button itself should have a short vacuum feed-through with a smooth transition to the 50 Ω cable, to avoid excitation of standing waves and to reach a bandwidth up to 10 GHz.

6.4 'Shoebox' pick-ups using the so-called linear cut

Owing to the long bunches at proton or ion synchrotrons, typically several metres, long electrodes, typically 20 cm, are installed to enhance the signal strength. A box-like device is usually used, to achieve a precise linear dependence with respect to the beam displacement, see Fig. 88. For the previously discussed button pick-up geometries, the signal of the plate closer to the beam's centre of mass is larger than that of the more distant plate; this is called the proximity effect. By contrast, shoebox pick-ups are based on another principle. The influenced signal is proportional to the actual plate length at the beam

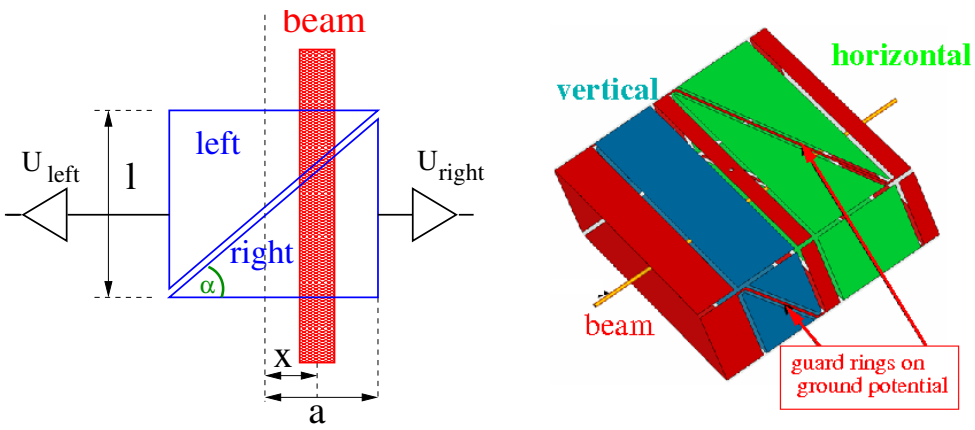


Fig. 88: Left: Position measurement using the so-called linear cut and an example of an electrode arrangement for the horizontal plane. Right: Linear-cut BPM for both planes.

centre position. For a given beam displacement x, the electrode's image voltage U_{im} is proportional to the lengths l_{left} and l_{right} of the beam projected on the electrode surface, as shown for the horizontal direction in the left-hand panel of Fig. 88. For triangle electrodes with half-aperture a, one can write:

$$l_{\mathrm{right}} = (a + x) \cdot \tan\alpha \quad \text{and} \quad l_{\mathrm{left}} = (a - x) \cdot \tan\alpha \quad \Longrightarrow \quad x = a \cdot \frac{l_{\mathrm{right}} - l_{\mathrm{left}}}{l_{\mathrm{right}} + l_{\mathrm{left}}} \quad . \tag{71}$$

The position reading is linear and can be expressed by the image voltages as

$$x = a \cdot \frac{U_{\mathrm{right}} - U_{\mathrm{left}}}{U_{\mathrm{right}} + U_{\mathrm{left}}} \equiv \frac{1}{S_x} \cdot \frac{\Delta U_x}{\Sigma U_x} \quad \Longrightarrow \quad S_x = \frac{1}{a}\,, \tag{72}$$

which shows that the position sensitivity for this ideal case is simply given by the inverse of the half-aperture. Compared with other types of pick-up, the position sensitivity is constant for almost the full range of displacements, i.e., nearly no corrections for non-linearity are required [121]. This is demonstrated in Fig. 89; numerical simulations show a linear scaling of the difference voltage ΔU_x of opposite plates normalized to the sum voltage ΣU_x as a function of beam offset up to about two-thirds of the half-aperture. Moreover, the position reading in the horizontal plane is independent of the beam position in the vertical plane. Owing to the linearity, the position reading is independent of the beam size, which is of importance for the relatively large beam size compared with the chamber aperture for low-energy proton or ion synchrotrons. This position linearity is the main advantage of the linear-cut type over the button BPM type. An analogue signal from such a linear-cut BPM used at an ion synchrotron is shown in Fig. 81.

As a summary, the basic parameters for linear-cut and button BPMs are compared in Table 12; this table contains significant simplifications and serves as an overview only.

For position measurement, other techniques are also used, including the measurement of the magnetic field with so-called inductive pick-ups or the excitation of cavity modes within a so-called cavity BPM. Further descriptions and citations can be found in Refs. [110, 112, 113, 122].

6.5 Electronic treatment for position determination

To determine the position of the beam, the signals from the electrodes must be compared. For this comparison, the signal shape (derivative or proportional behaviour) is of minor importance. The electronics used for this purpose are described only briefly in this contribution; a detailed review is given in Refs. [123–125]. For position resolution, the signal-to-noise ratio is important. Beside the stray fields from the RF cavities, the broadband amplifier noise and the electronic noise of the subsequent devices

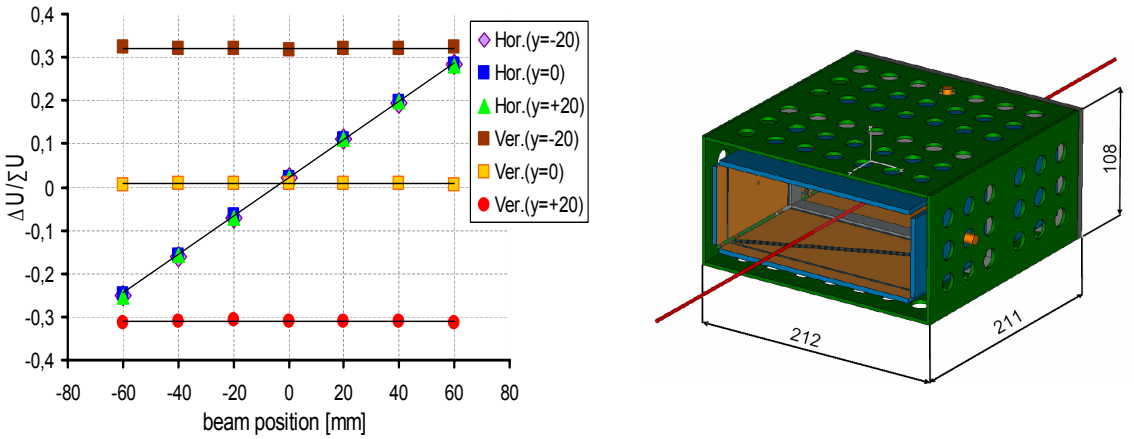


Fig. 89: Left: Numerical calculation of position reading, $\Delta U/\Sigma U$, as a function of beam displacement for the horizontal plane. The simulated beam is swept in the horizontal direction (curves with constant slope) and vertical direction (constant functions) and proves the decoupling of both planes [121]. Right: BPM geometry for these calculations, with horizontal half-aperture $a = 100$ mm, vertical half-aperture $b = 35$ mm and length 211 mm, comparable to the BPM shown in Fig 90.

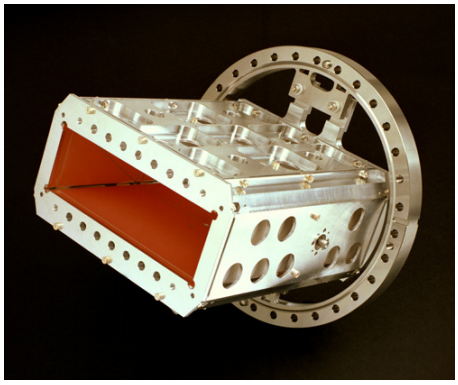


Fig. 90: Linear-cut position pick-up with an aperture limitation of 200×70 mm^2, from the GSI synchrotron ring

Table 12: Simplified comparison of linear-cut and button BPMs

	Linear-cut BPM	Button BPM
Precaution	Bunches longer than BPM	Bunches comparable to BPM
Length (typical)	10–20 cm per plane	$\varnothing$0.5–5 cm
Shape	Rectangular or cut cylinder	Orthogonal or planar orientation
Mechanical realization	Complex	Simple
Coupling	Often 1 MΩ, sometimes 50 Ω	50 Ω
Capacitance (typical)	30–100 pF	3–10 pF
Cut-off frequency (typical)	1 kHz for $R = 1$ MΩ	0.3–3 GHz for $R = 50$ Ω
Usable bandwidth (typical)	0.1–100 MHz	0.3–5 GHz
Linearity	Very linear, no x–y coupling	Non-linear, x–y coupling
x-position sensitivity	Good	Good
Usage	Proton synchrotron, $f_{\mathrm{acc}} < 10$ MHz	Proton linac, all electron accelerators $f_{\mathrm{acc}} > 100$ MHz

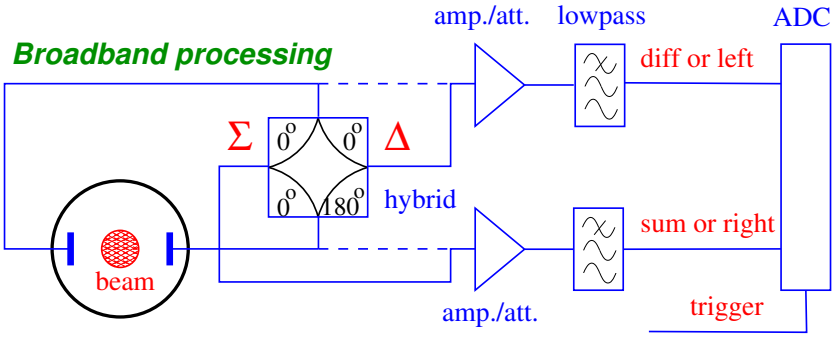


Fig. 91: Broadband signal processing. Plate signals are either fed directly to the amplifier (dashed line) or via a hybrid. The sum and difference signals are generated in an analogue manner (solid line).

have an effect. Therefore, a minimum bunch current is necessary for a reliable position measurement. Two different principles are commonly used, based on analogue electronics; so-called broadband and narrowband processing.

Broadband processing. In the broadband case, as shown in Fig. 91, the signals from the individual plates are amplified (or even attenuated) to adapt the signal strength to the ADC input level. The sum and difference signals are then calculated from the digitized values. For noise reduction and alias-product suppression, a low-pass filter is used, matched to the sample rate of the ADC. In older installations, fast sampling was not possible; an external trigger, generated from the bunch passage in an analogue manner, forced the digitalization. In many applications, the sum and difference voltages are analogously generated by a 180° hybrid or a differential transformer. Because these are purely passive devices, they can be mounted quite close to the BPM plates, even when there is high radiation. The resulting sum and difference signals are then stored in the ADC. The difference signal, which is normally smaller by at least a factor of ten, can be amplified by a larger amount than the sum signal to exploit the full ADC range. (An overview of standard RF components can be found in Ref. [126].) The analogue electronics are required to match the signal shape to the properties of the ADCs; an appropriate trigger is used for the digitalization. Modern installations do not use these types of analogue electronics; instead, the signal is digitized directly by fast ADCs or digital receivers, which are even commercially available [127]. With the help of high-speed digital signal processing, the bunch signal is than reduced to one value per bunch and the beam position is calculated from the sum and difference values, see, e.g., Ref. [128]. Broadband processing is less precise than narrowband processing. For electron machines, with small beam size, a resolution of 100 µm can be achieved using broadband processing. The advantage is the possibility of performing bunch-by-bunch analysis (i.e., always measuring the position of the same bunch rotating in a synchrotron) using an appropriate trigger setting or precise digital signal processing. The broadband scheme is installed at transfer lines between synchrotrons, where only one or few bunches are transferred. A further modification is the use of logarithmic amplifiers, which are commercially available, e.g., Ref. [11], and which are often installed at transfer lines; in these, an order-of-magnitude increase of input signal is converted to a linear increase of the output signal. For the typical button BPM arrangement, a good linearity for the position sensitivity is attained, as depicted in Fig. 85. The advantage is that a large dynamic range can be covered by a single amplifier stage without range switching. The disadvantage is the smaller achievable accuracy, which is often acceptable at transfer lines.

Narrowband processing. Narrowband processing is used to obtain a higher precision of the position reading, attaining 1 µm in electron machines using the electronics scheme shown in Fig. 92. The better signal-to-noise ratio is achieved by reducing the bandwidth of the signal processing by several orders of magnitude. As an illustrative example, the Fourier transform of the signal from the GSI linac pick-up is shown in Fig. 93. The spectral power of the signal is mainly available at the bunch repetition harmonics nf_0. The position evaluation uses only this large signal power within the frequency band

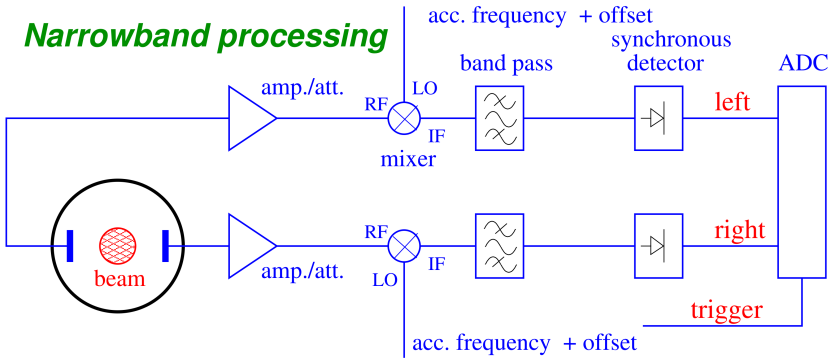


Fig. 92: Signal analysis for narrowband treatment

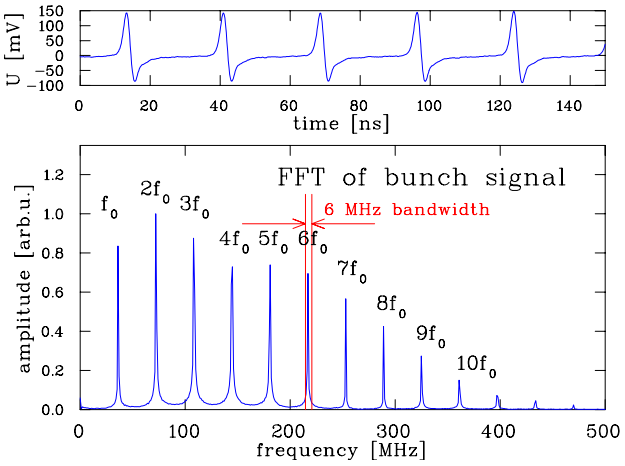


Fig. 93: Bunch signal (top) and its Fourier transform (bottom) at the GSI linac of a 1.4 MeV/u beam. In the Fourier spectrum, each line corresponds to the indicated harmonic of the 36 MHz accelerating frequency. The position is evaluated at $6f_0 = 216.8$ MHz using 6 MHz narrowband processing, as indicated by the red lines.

Δf, while the thermal noise is reduced as $U_{\text{eff}} \propto \sqrt{\Delta f}$. Technically, the pick-up signal is mixed with the accelerating frequency, e.g., using RF components described in Ref. [126]. The resulting quasi-d.c. signal is then digitized by an ADC and the position calculated via software. In general, this technique is called heterodyne mixing; the same principle is used in a spectrum analyser. The mixing is equivalent to an average over many bunches, leading to much greater precision in the position reading, but does not allow turn-by-turn observation. This method can also be applied for a synchrotron with varying acceleration frequency, e.g., a proton synchrotron for non-relativistic velocities, because the varying accelerating frequency changes in phase with the bunch signal, resulting in a constant intermediate frequency. Such a system is commercially available [11]. Moreover, for a synchrotron, such narrowband processing is applied at linacs or behind cyclotrons with a pulse length greater than 10 µs as, e.g., used for multiturn injection into a synchrotron. If only one or few bunches are transferred, e.g., between two synchrotrons, narrowband processing cannot be applied.

These methods are used to perform BPM signal conditioning in an analogue manner; for most cases, this leads to a small-bandwidth output, which is then digitized with a slow sampling ADC. Motivated by their intensive use in telecommunication, the sampling rate and voltage resolution of ADCs have significantly improved over the last decade; an overview of ADC technology can be found in Ref. [129]. Moreover, digital signal processing capabilities have been enhanced. The tendency of modern technologies is to digitize the signal at an early stage in the signal path and replace the analogue conditioning with digital signal processing, e.g., performing the narrowband treatment in the digital domain.

In most cases, the required parallel processing is realized using a field programmable gate array to ensure fast, real-time, response. On this digital basis, versatile treatments can be implemented without hardware exchange [127, 130], leading to higher flexibility for the application. An example is the filtering of one data set with different bandwidths to match different requirements for position and time resolution; such digitalization and processing units are commercially available [127]. Related methods, compared with analogue treatment, are reviewed in Ref. [125].

6.6 Trajectory measurement

The trajectory indicates the position of a single bunch during its transport through the accelerator; this gives basic information of the initial alignment of the accelerator setting. Because the single-bunch position is monitored, broadband processing is used for the BPM electronics; the reduced position resolution is of minor concern for this measurement.

Figure 94 shows, schematically, typical instrumentation for a transfer line between synchrotrons. It is assumed that this transfer line comprises an entry section, a long central section, with a more or less periodic arrangement of several focusing-defocusing (FODO) cells equipped with a quadrupole and a steerer for each direction, and an exit section with more flexible matching control. The main goal during operation is steering correction of the beam using the BPM readings as inputs. In the optimal case, the correction should be performed independently for the horizontal and vertical directions. Each section has a different task and related instrumentation.

- *Entry section.* The position and angle of the beam are gained from two BPMs spaced a certain distance to achieve good resolution for both parameters. Steering control is most often achieved by changing the values for the septa within the synchrotron. In addition, a profile measurement is foreseen to estimate the validity of the assumed transverse matching and the expectation with respect to the transverse emittance.
- *Central section.* Steering control in the central part can be performed incrementally by modifying the settings of the dipole steerers, which are often installed close to the corresponding focusing quadrupole and separated for the horizontal and vertical directions. The kick produced by an individual steerer is monitored by a BPM, which should be mounted so as to have a betatron phase advance of about $\mu_\beta \simeq 90^\circ$ to transfer the steerer's kick to an offset at the BPM.
- *Exit section.* Beside beam control via a dipole steerer, transverse matching to the closed orbit requirements of the synchrotron is performed by a doublet or even triplet quadrupole, arranged for flexibility, and is checked with a BPM and, additionally, profile measurement. For a fixed target following the transfer line, the same arrangement is chosen in many cases.

In addition to position determination, the sum signal of the BPM can be used for relative estimation of the beam current. Nevertheless, for the absolute current determination, transformers must be installed, at least at the entrance and exit sections, as well as behind possible aperture limits (not depicted in Fig. 94). Beam loss monitors can assist in transmission optimization and are therefore installed at the entrance and exit sections, as well as at possible acceptance limits. In the central section, the possibility of emittance measurement, using the method of quadrupole variation and an additional profile monitor, is normally considered, see Section 5.3.

As an example of a measurement, Fig. 95 depicts the measurement of the trajectory of a single bunch injected into the LHC in the first days of its commissioning [131]. The position reading at each of the 530 BPMs on the 27 km long synchrotron is displayed showing some oscillations caused by mismatched injection parameters, which could be aligned later on with the help of the diagnostics. In this respect, the LHC is treated as a transfer line.

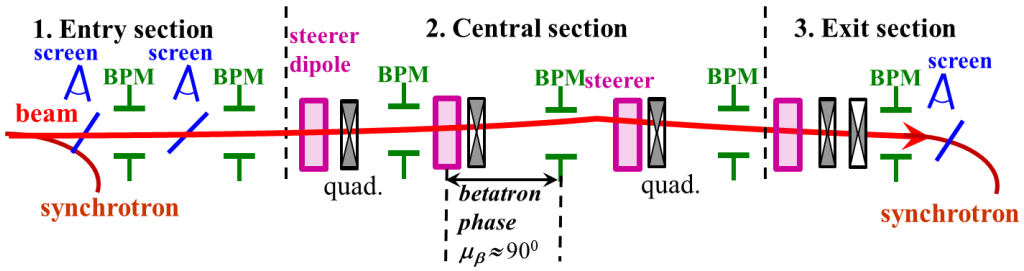


Fig. 94: Typical arrangement of diagnostics in a transfer line for bunched beam transfer between synchrotrons

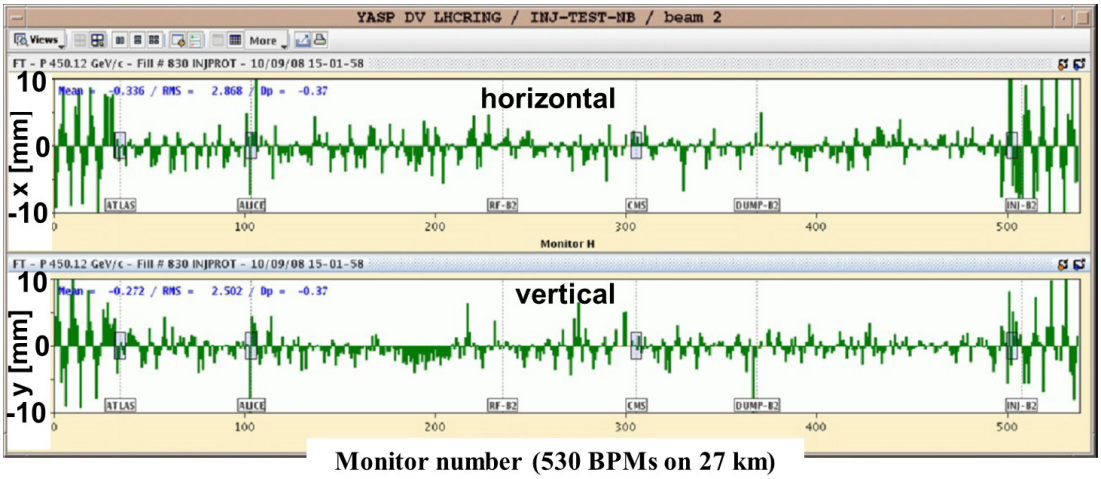


Fig. 95: Measurement of the trajectory of a single bunch injected in the LHC with a position display at the 530 BPMs around the synchrotron [131]. Image from R. Jones (private communication).

6.7 Closed orbit measurement at a synchrotron

It is important, for operation, to match the beam centre from the transfer line to the reference orbit inside the synchrotron. For proton synchrotrons in particular, the reference orbit must be determined to ensure stable storage of the beam without emittance enlargement. The closed orbit is defined by averaging the position over many turns; the required accuracy is of the order of 10 µm for proton accelerators and less than 1 µm for synchrotron light sources. In a synchrotron, several BPMs are installed to determine the closed orbit. As a rule of thumb, a good choice is to have four BPMs per tune value separated by about $\mu_\beta \simeq 90^\circ$ betatron phase advance. The closed orbit is the central beam path along the machine, taking existing imperfections into account. The closed orbit might differ from the ideal path, defined by the centre of the quadrupole magnets. Only for a good alignment does the real beam behave as expected from beam optics calculations, e.g., no beam steering by the quadrupole magnets occurs. An example of the use of position measurement during the acceleration in the GSI synchrotron is given in Fig. 96.

The position reading of the BPMs around the ring can be used as the input of a feedback loop for active beam correction to compensate for systematic and temporal uncertainties, see, e.g., Ref. [132]. Such a feedback system is installed in most synchrotrons. Using the results of the position measurement, various beam parameters for a beam stored in a synchrotron can be determined, such as the tune, chromaticity, and beta function at the BPM location; see Refs. [1–7, 133, 134] for an introduction to those measurements. Comparable methods for beta functions and dispersion determination can be applied to transfer lines as, e.g., described in Ref. [135, 136]. A detailed description is beyond the scope of this contribution.

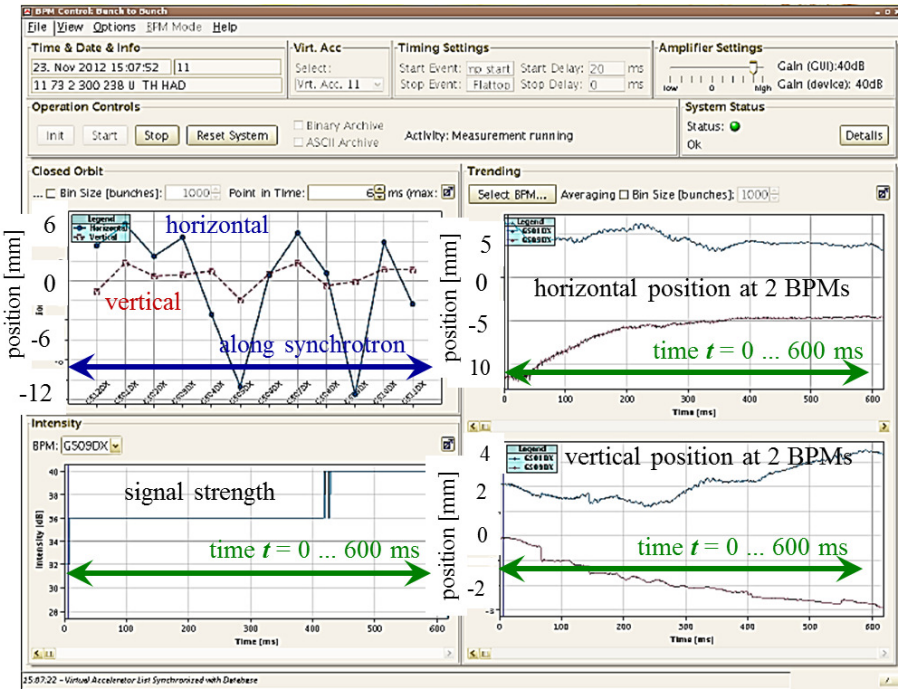


Fig. 96: Position measurement at the GSI synchrotron of a partly misaligned beam during a 0.6 s acceleration. Top left plot: Position around the ring at 12 locations at the beginning of the ramp. Top right plot: Horizontal position during acceleration ramp for two BPM locations. Bottom right plot: Vertical position for these BPMs. Bottom left plot: Signal power $\propto U_\Sigma^2$ during ramp for a single BPM. The panel at the top is used to control pre-amplifiers and timing.

7 Measurement of longitudinal parameters

Longitudinal parameters are as important as transverse ones. The longitudinal phase space is spanned by the longitudinal spread and the momentum spread.

- The longitudinal spread of the bunch l is given in units of length (millimetres), time (nanoseconds) or phase (degrees with respect to accelerating frequency), see Fig. 97. The mean value is the centre of the bunch relative to the RF or relative to the ideal reference particle. The corresponding transverse value is the transverse beam profile. For an unbunched d.c. beam, e.g., in a proton storage ring, the quantity has no meaning.
- The momentum spread $\delta = \Delta p/p$ is the deviation relative to the momentum p of the ideal reference particle. Instead of momentum, it is sometimes common in proton or ion linacs and cyclotrons to record the quantity beam energy $\Delta E_{\mathrm{kin}}/E_{\mathrm{kin}}$ or even only ΔE_{kin} . The corresponding transverse value is the beam divergence.

The value of emittance ϵ_{long} is given by the product of the two quantities,

$$\epsilon_{\mathrm{long}} = \frac{1}{\pi}\int_A \mathrm{d}l\mathrm{d}\delta\,, \tag{73}$$

where A is the area of the phase space occupied by the beam particles (see Fig. 97). Linear transformations can be applied in the same way as for the transverse case; see Section 5.1. The normalized longitudinal emittance

$$\epsilon_{\mathrm{long}}^{\mathrm{norm}} = \frac{v_s}{c}\gamma_{\mathrm{rel}} \cdot \epsilon_{\mathrm{long}} \tag{74}$$

is preserved under ideal conditions, where v_s is the longitudinal velocity and γ_{rel} is the relativistic Lorentz factor.

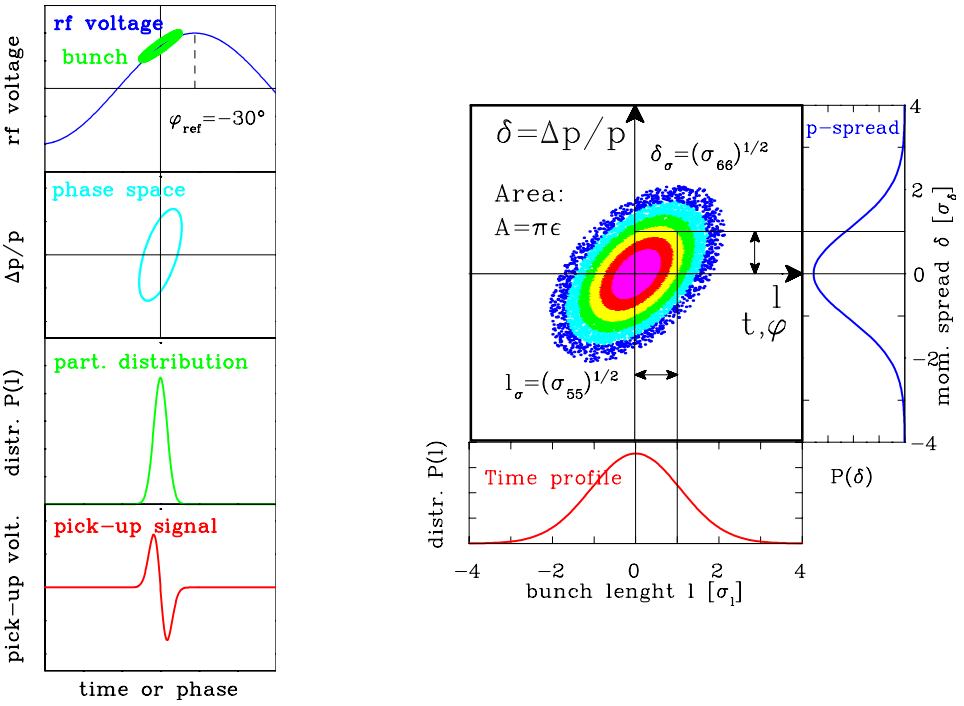


Fig. 97: Left: Relation between accelerating voltage and longitudinal emittance, as well as time distribution of the bunches, as measured with the differentiated pick-up signal. Right: Longitudinal phase space plot of Gaussian distribution.

Using a pick-up or a fast current transformer, the projection of the phase space on the time axis could be determined for some beam parameters, giving the bunch position and width. Determination of the emittance by linear transformation is possible. Some conditions must be fulfilled for a pick-up or transformer to guarantee an adequate bunch shape measurement. Firstly, the signal from the pick-up or transformer must reflect the bunch shape, i.e., the frequency spectrum of the bunch must fit in the bandwidth of the recording electronics; for short bunches, e.g., at high-energy electron accelerators, this might not be the case, see Section 7.7. Secondly, the bunch must be much longer in the longitudinal direction than the pick-up. For bunches with a length comparable to the pick-up, a further condition is that the beam must be sufficiently relativistic that the beam's electric field is essentially transverse. For the transverse electric field component $\vec{E}_\perp$, Lorentz transformation results in an enhancement in the laboratory frame, compared with the rest-frame of the moving charge, by

$$E_{\perp,\mathrm{lab}}(t_{\mathrm{lab}}) \simeq \gamma \cdot E_{\perp,\mathrm{rest}}(t_{\mathrm{rest}})\,; \tag{75}$$

see the more detailed discussion in Section 7.1 concerning the transformation of t_{rest} to t_{lab}, which is important for low-energy, non-relativistic particles, e.g., at proton linacs. In the case of a slow beam from a proton linac, the last condition is not fulfilled and other methods must be applied. For bunches shorter than the pick-up, the signal does not reflect the longitudinal bunch shape. For electron beams, the bunch length is so short that the bunch structure is smeared out by integration over the pick-up capacitance, i.e., is expressed in terms of electrical parameters by the limited bandwidth. Here, synchrotron radiation is monitored in connection with the fast optical method using streak cameras, see Section 7.7.

Measurement of the energy or momentum spread $\delta = \Delta p/p$ is not discussed in detail here. A magnetic spectrometer can be used for this purpose, giving point-to-point focus in connection with small slits located at appropriate locations. This is discussed in textbooks on beam optics, e.g., Ref. [137].

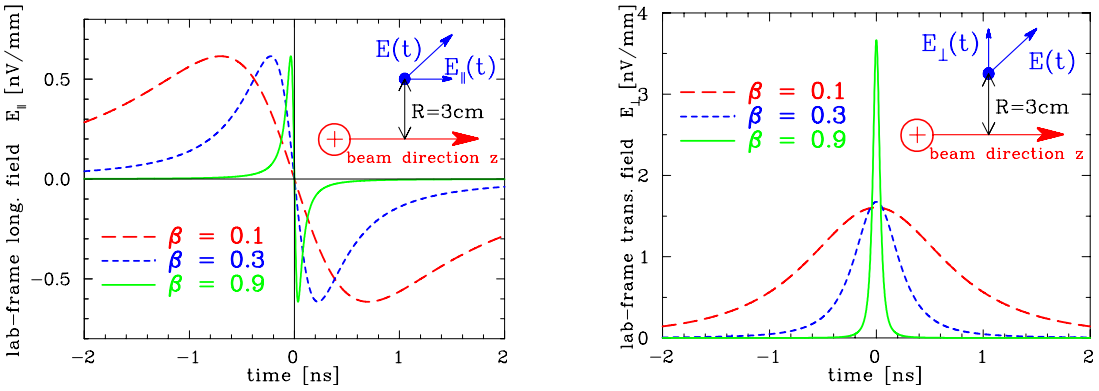


Fig. 98: Effect of advanced electric field at an observation point $R = 30$ mm from the beam axis for different velocities of a point-like charge. Left: longitudinal electric field. Right: Transverse electric field. The closest approach is reached for $t_{\mathrm{rest}} = t_{\mathrm{lab}} = 0$.

7.1 Electric fields of a relativistic charge

Owing to the finite velocity of light, the electric field of a moving charge is modified, as given by the Lorentz transformation. This effect for different velocities β of the moving charge is shown in Fig. 98, which shows the longitudinal electrical field $E_{\parallel}(t)$ in the laboratory frame calculated for a single particle with a charge of e travelling in empty space, according to the time in the laboratory frame, i.e., $t_{\mathrm{lab}} \equiv t$. It is calculated by the relativistic Lorentz boost, including the transformation of the time co-ordinate [18], as

$$E_{\parallel,\mathrm{lab}}(t) = -\frac{e}{4\pi\varepsilon_0} \cdot \frac{\gamma\beta ct}{\left[R^2 + (\gamma\beta ct)^2\right]^{3/2}} , \tag{76}$$

where R is the distance of closest approach and

$$\gamma = \left(1 - \beta^2\right)^{-1/2}$$

is the relativistic Lorentz factor. The transverse field component $E_{\perp}$ in the laboratory frame is given by

$$E_{\perp,\mathrm{lab}}(t) = \frac{e}{4\pi\varepsilon_0} \cdot \frac{\gamma R}{\left[R^2 + (\gamma\beta ct)^2\right]^{3/2}} \quad . \tag{77}$$

These formulae are valid for a free particle and are modified by the boundary condition of the metallic surface, e.g., of the beam pipe or the pick-up plates. A beam bunch consists of many particles distributed within a certain volume; to calculate the accompanying electric field, one must integrate over the bunch volume. For small velocities, the electric field arrives earlier at the pick-up and does not reflect the charge distribution of the bunch; see Section 7.6 for more details.

7.2 Determination of phase position

To match successive linac modules, the right phase relation between accelerating frequency and bunch arrival must be determined using a beam-based measurement. A typical plot is shown in Fig. 99 for a slow MeV/u ion beam recorded by a button pick-up. Owing to the differentiated beam signal, the centre of the bunch is well-determined. There is an arbitrary, but fixed, phase shift between the signal for the pick-up and the accelerating frequency f_{RF}, owing to the different cable lengths. The corresponding quantity concerning the transverse beam position is the centre of mass of a bunch relative to the accelerating RF sine wave.

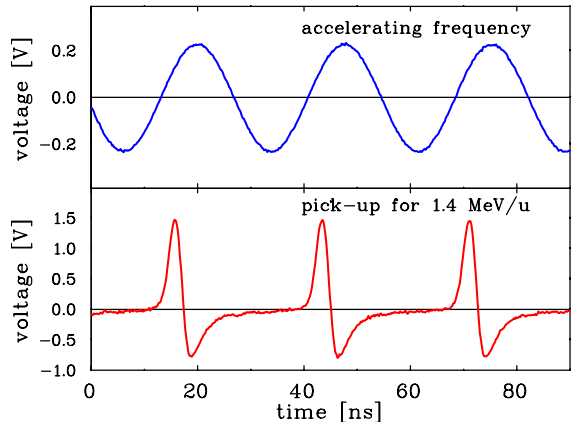


Fig. 99: Top: Linac module accelerating frequency for determination of bunch zero-crossing from a 1.4 MeV/u ion beam. Bottom: Corresponding pick-up signal, recorded using a capacitive pick-up.

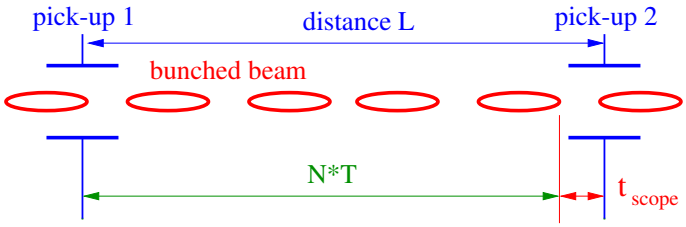


Fig. 100: Principle of time-of-flight measurement using two pick-ups

7.3 Determination of non-relativistic mean energy

The final energy of a linac module is sensitive to the actual setting of the RF amplitude and the phase relative to the preceding modules. Non-relativistic energies can be measured using two pick-ups separated by a distance L and a time-of-flight measurement, see, e.g., Refs. [77, 78, 138]. As displayed in Fig. 100, two pick-ups are installed in a straight section, typically separated by several metres. This distance L must be known precisely, typically to within 1 mm. The centre of mass of the bunch can be read from the oscilloscope, or the correlation function between the two signals can be calculated. A typical measurement is displayed in Fig. 101. As shown in the schematic, there are several bunches between the pick-ups and a coarse estimate of the velocity is needed. This is normally known, but in case of trouble, a third pick-up must be installed much closer to one of the other two. The velocity is calculated for the measured time t_{scope} via

$$\beta c = \frac{L}{NT + t_{\mathrm{scope}}} \tag{78}$$

with N bunches between the pick-ups and a bunch repetition time of $T = 1/f_{\mathrm{RF}}$.

The accuracy of such a velocity measurement $\Delta\beta/\beta$ is given by the uncertainty of the distance ΔL and the scope reading Δt as

$$\frac{\Delta\beta}{\beta} = \sqrt{\left(\frac{\Delta L}{L}\right)^2 + \left(\frac{\Delta t}{NT + t_{\mathrm{scope}}}\right)^2}\,. \tag{79}$$

An accuracy better than 0.1% for the energy spread $\Delta E_{\mathrm{kin}}/E_{\mathrm{kin}} = 2\Delta\beta/\beta$ can be obtained, assuming an accuracy for the distance of $\Delta L = 1$ mm and for the time measurement of $\Delta t = 100$ ps; the parameters used at the GSI linac are summarized in Table 13.

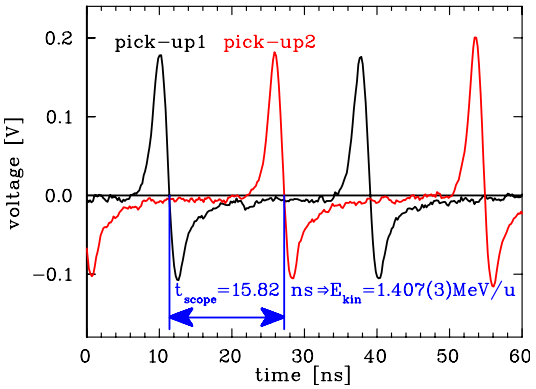


Fig. 101: Time-of-flight signal from two pick-ups with $E_{\mathrm{kin}} \simeq 1.4$ MeV/u at the GSI linac; $f_{\mathrm{RF}} = 36.136$ MHz $\Leftrightarrow$ $T = 1/f_{\mathrm{RF}} = 27.673$ ns. For a distance $L = 1.629$ m between the pick-ups, there are $N = 3$ bunches between the pick-ups. Inserting a time difference $t_{\mathrm{scope}} = 15.82$ ns in Eq. 78, the velocity $\beta = 0.05497(7)$ and corresponding beam energy $E_{\mathrm{kin}} = 1.407(3)$ MeV/u are calculated. The error is given by Eq. 79, inserting uncertainties of $\Delta t = 0.1$ ns and $\Delta L = 1$ mm.

Table 13: Energy resolution for time-of-flight measurement with pick-up distance of 3.25 m for some GSI linac modules (internal linac module names are given). The RF frequency is $f_{\mathrm{RF}} = 36.14$ MHz and the assumed time accuracy is 100 ps.

Physical parameter	Unit	RFQ	IH1	IH2	AL4
Energy, W	MeV/u	0.12	0.75	1.4	11.4
Velocity, β	%	1.6	4.0	5.5	15.5
Total time of flight	ns	677	271	197	70
Bunch spacing, $\beta c/f_{\mathrm{RF}}$	cm	13	33	45	129
Resolution, $\Delta W/W$	%	0.07	0.10	0.12	0.22

The final energy of an ion beam in a linac module depends on the applied amplitude and phase of the accelerating RF. This is depicted in Fig. 102 for an ion beam of 1.400 MeV/u nominal energy. This nominal energy is achieved only in a small interval; for the displayed case, the allowed amplitude range is less than $\pm 2\%$ and the phase should be aligned within $\pm 10^{\circ}$. This corresponds to a relatively small range; it calls for precise beam alignment and a large accuracy in the time-of-flight measurement, as demonstrated by the small error bars in Fig. 102. Matching the nominal beam energy of a linac module is an important task, because the energy acceptance of the following module is typically restricted within several percent and a sufficient beam transmission through a chain of linac modules is only achieved if all modules produce the correct energy increase. Moreover, the longitudinal emittance depends strongly on the correct RF amplitude and phase setting.

Even though the time-of-flight method is commonly called energy measurement, it is basically a velocity determination. This method cannot be applied for electron beams, because electrons are relativistic after the first linac modules. Acceleration of electrons is related more to an increase in their momentum; therefore, energy determination by the bending of a dipole magnet is more appropriate in this case. Because of the dependence of the bending radius on the particle momentum Δp with respect to a reference particle of momentum p_0, the momentum distribution is transferred to a distribution in the horizontal plane. This is described by the optical parameter dispersion $D(s)$, as the additional beam offset is $\Delta x(s) = D(s) \cdot \Delta p/p_0$. However, the resulting profile is a mixture of the transverse and longitudinal parameters and transverse collimation is required at the entrance of the optical system, see discussions in textbooks on beam optics, e.g., Ref. [137].

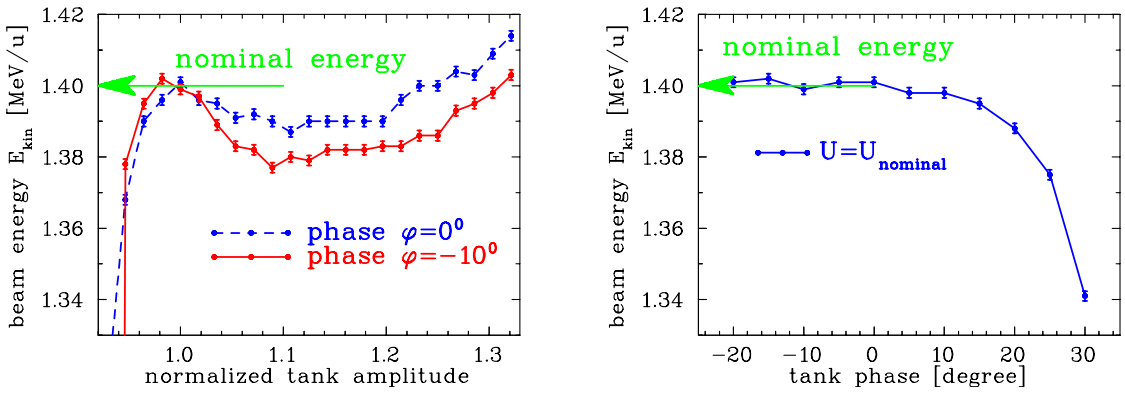


Fig. 102: Left: Output energy of the GSI linac, depicted as a function of the applied RF amplitude inside the tank, as displayed in Fig. 101. Right: Phase for nominal amplitude. The required energy of this module is 1.400 MeV/u. The distance between the pick-ups in this case is $L = 1.97$ m with $N = 4$ bunches between pick-ups.

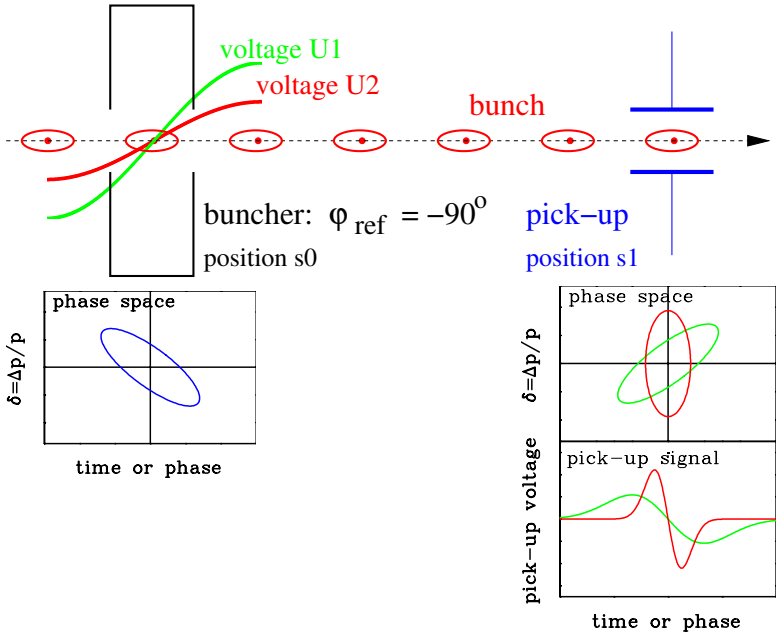


Fig. 103: Longitudinal emittance measurement using a buncher and one bunch shape monitor

7.4 Longitudinal emittance from linear transformation

Longitudinal emittance can be determined from linear transformation, as in the transverse case described in Sections 5.3 and 5.4. The quadrupole variation corresponds to the change in longitudinal focusing performed by a buncher, see Fig. 103. The phase of a buncher is adjusted to 0°, so no acceleration is present. By varying the amplitude, the bunch can be longitudinally focused, i.e., compressed at the measurement location or de-focused, i.e., lengthened. The action is described in the longitudinal sub-space $(l, \delta = \Delta p/p)$, which is a part of the full six-dimensional phase space describing both transverse and longitudinal planes. (The six-dimensional phase space is spanned by the vector $\vec{x}^{\,\mathrm{t}} = (x, x', y, y', l, \delta)$; the linear behaviour of the accelerator, including the coupling between the horizontal, vertical, and longitudinal degree of freedom, can be described with the help of the 6×6 transport matrices $\mathbf{R}$. For the formula used here, the bunch length l is given in units of spatial length (metres) to have the same units for all three degrees of freedom.)

The general transfer matrices $\mathbf{R}$ for the transverse and longitudinal phase space have 6×6 elements, as does the beam matrix $\boldsymbol{\sigma}$. Disregarding any transverse–longitudinal coupling (e.g., by a dipole

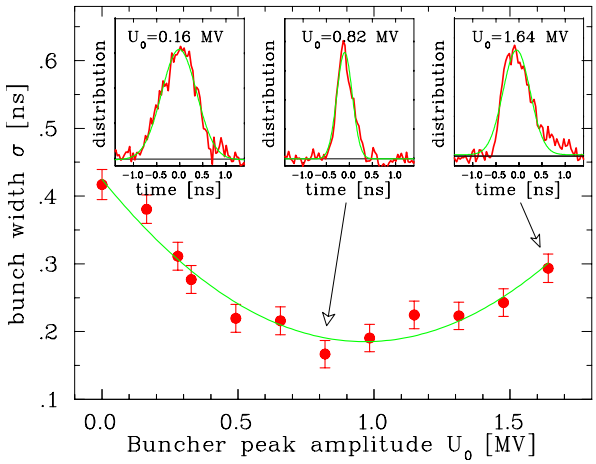


Fig. 104: Measurement of bunch width (one standard deviation) as a function of the buncher voltage 31 m upstream. The bunch shape is not measured by a pick-up, but by a specially designed monitor for low-energy beams [139].

magnet), the longitudinal phase space corresponds to the 'lower right' part of the full 6×6 matrices, having the indices 5 and 6. The reduced 2×2 transfer matrix of a buncher, $\mathbf{R}_{\mathrm{buncher}}$, is given in a thin-lens approximation by

$$\mathbf{R}_{\mathrm{buncher}} = \begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix}, \tag{80}$$

for the focal length f

$$\frac{1}{f} = \frac{2\pi q U_0 f_{\mathrm{RF}}}{A p v^2} \tag{81}$$

and the voltage amplitude U_0 of the accelerating frequency f_{RF} for a particle of charge q and mass number A with velocity v and momentum p. The transfer matrix for a drift space of length L is given by

$$\mathbf{R}_{\mathrm{drift}} = \begin{pmatrix} 1 & L/\gamma_{\mathrm{rel}}^2 \\ 0 & 1 \end{pmatrix}, \tag{82}$$

where γ_{rel} is the relativistic Lorentz factor. For the transverse case, the emittance can be determined by an amplitude variation U_0 of the buncher, leading to different longitudinal focusing. The bunch length must be measured at the pick-up location s_1 for at least three settings of the buncher. The beam matrix $\boldsymbol{\sigma}(0)$ must be determined at the buncher location s_0. For bunch width $l_{\max}(1, f_k) = \sqrt{\sigma_{55}(1, f_k)}$ at pick-up location s_1, a system of linear equations for $\sigma_{ij}(0)$ of the form

$$\begin{aligned}
\sigma_{55}(1,f_1) &= R_{55}^2(f_1)\cdot\sigma_{55}(0) + 2R_{55}(f_1)R_{56}(f_1)\cdot\sigma_{56}(0) + R_{56}^2(f_1)\cdot\sigma_{66}(0)\,, \quad \text{focusing } f_1 \\
\sigma_{55}(1,f_2) &= R_{55}^2(f_2)\cdot\sigma_{55}(0) + 2R_{55}(f_2)R_{56}(f_2)\cdot\sigma_{56}(0) + R_{56}^2(f_2)\cdot\sigma_{66}(0)\,, \quad \text{focusing } f_2 \\
&\vdots \\
\sigma_{55}(1,f_n) &= R_{55}^2(f_n)\cdot\sigma_{55}(0) + 2R_{55}(f_n)R_{56}(f_n)\cdot\sigma_{56}(0) + R_{56}^2(f_n)\cdot\sigma_{66}(0)\,, \quad \text{focusing } f_n
\end{aligned} \tag{83}$$

is obtained for $k = 1, 2, \ldots, n$ measurements with $f_k = f_1, f_2, \ldots, f_n$ different settings of the buncher. This redundant system of linear equations must be solved, as discussed in Section 5.3. A measurement is shown in Fig. 104. For the method, it is of no importance which device is used for determination, as long as it reflects the actual bunch shape of the beam.

As discussed for the transverse case, the restrictions are the presence of simple, e.g., Gaussian, phase space distributions, linear beam manipulations, and the lack of non-linear space charge forces.

In addition, one should emphasize that this method is only used at linacs; in most cases, the pick-up resolution is insufficient and special detectors are used, as discussed in Section 7.6 for proton beams and Section 7.7 for electron beams.

A method corresponding to the 'three-grid method' for the transverse emittance (Section 5.4), i.e., determination of the bunch shape at several locations, is seldom applied. This is because a special monitor for bunch length determination must be used at proton linac facilities, and, in most cases, is only installed at a few locations.

For beams extracted from a synchrotron, the variation of the bunch length within the beam transfer line is negligible in most cases, owing to the small momentum spread $\Delta p/p$ of the circulating beam and the short flight time t_{drift} to the target of a further synchrotron, as the increase of the bunch length is $\Delta t = \Delta p/p \cdot t_{\mathrm{drift}}$. As an example, the following numbers can be inserted. For a relativistic beam of velocity $\beta \simeq 1$ and a distance of $L = 100$ m from the synchrotron to the target, the drift time $t_{\mathrm{drift}} = L/\beta c = 333$ ns; assuming a momentum spread of $\Delta p/p = 10^{-3}$, the resulting lengthening of the bunch is $\Delta t = \Delta p/p \cdot t_{\mathrm{drift}} = 0.3$ ns, which is shorter than the typical bunch length of $\sigma_{\mathrm{bunch}} > 10$ ns in a proton synchrotron. Therefore, measurement inside the synchrotron is sufficient and the current transformer mounted in the synchrotron can be used for monitoring.

7.5 Longitudinal emittance using tomographic reconstruction

In a proton synchrotron, the bunch length is much longer than the typical size of a pick-up or a current transformer. This enables a true image of the bunch shape to be recorded. Observation and control of the longitudinal behaviour at injection, acceleration, and possibly bunch manipulations, such as combining or splitting of bunches, can be performed.

An innovative analysis method has been developed at CERN [140–142] to determine the longitudinal phase space distribution and hence the emittance from a measurement of the bunch shape on a turn-by-turn basis. The idea is based on the well-known tomographic reconstruction used as an imaging technique in medicine for X-ray tomography and nuclear magnetic resonance tomography and described in textbooks on medical physics.

The underlying principle of tomography is to combine the information from a sufficiently large number of projections to unambiguously construct a more complete picture with the extra dimension reinstated. The application of tomography to longitudinal phase space in an accelerator becomes obvious once it is realized that a bunch of particles performing synchrotron motion is analogous to a patient rotating in a stationary body scanner. On each turn around the machine, a longitudinal pick-up provides a 'snapshot' of the bunch projected at a slightly different 'angle', i.e., the oscillation of the bunch shape, as given by the inverse of the synchrotron frequency f_{synch}, is, in nearly all cases, much smaller than the revolution frequency $f_{\mathrm{rev}} \gg f_{\mathrm{synch}}$. It suffices to combine such profiles tomographically to obtain a two-dimensional picture of the phase space density. Non-linear emittance growth can also be included [140–142]. A general, well-fulfilled condition is that the synchrotron frequency is much smaller than the revolution frequency. At least half a synchrotron period must be recorded, corresponding to a 180° image of the patient body in the medical application.

The general idea behind tomography is based on algebraic reconstruction techniques (described in textbooks on medical physics), as depicted in Fig. 105 from Refs. [140–142]. The contents of the bins of a one-dimensional histogram are redistributed over the two-dimensional array of cells that comprises the reconstructed image. Assuming no a-priori knowledge of the original two-dimensional distribution, the contents of one bin are spread over all the cells that could have contributed to that bin. The back-projection of all bins of all profiles yields a first approximation of the original distribution. Back-projection of the bin-by-bin difference between the original set of profiles and a new set yields an improved approximation. Further iterations converge more rapidly if any cell whose contents have become negative is reset to zero. A more detailed description is found in Refs. [140–142].

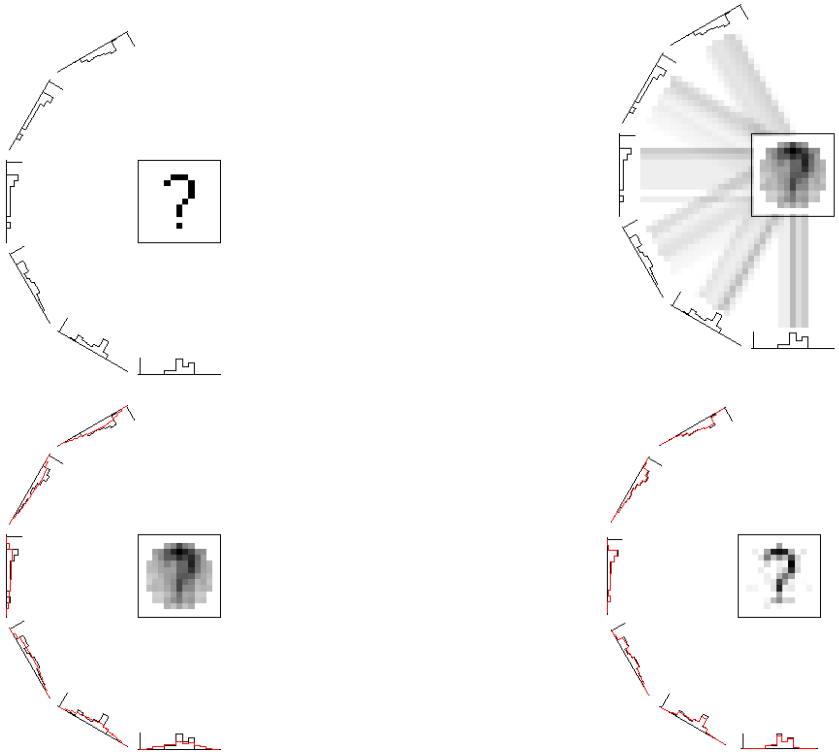


Fig. 105: Schematic plots for tomographic reconstruction [140–142]. Top left: Original two-dimensional image and one-dimensional projections. Top right: First back-projection. Bottom left: First iteration step for the generation of differences in the projected data. Bottom right: After ∼50 iterations, the original image can be reconstructed.

Two examples demonstrate the analytical power of tomography [142, 143]. The first, illustrated in Fig. 106, is the recording of a standard, constant bunch shape using a digital scope. Reconstruction of the first time slice shows good filling of the bucket. The second, illustrated in Fig. 107, is related to a special bunch manipulation. In phase space (as shown for a small bunch width), the distribution rotates at the synchrotron frequency, resulting in varying bunch shapes as a function of time. This type of short bunch is created to fulfil the demands of the user, who wants to use a short bunch for dedicated experiments. Reconstruction of the beam shape inside the synchrotron gives the advantage of observing the entire creation process of the so-called bunch rotation and enabling correct trigger timing for extraction of the bunch.

7.6 Bunch structure for non-relativistic energies using secondary electrons

For beams of non-relativistic velocities, the electric field of a bunch has a significant longitudinal component, as calculated for Fig. 98. This longitudinal field is seen by the pick-up, leading to a smearing of the signal. The structure of a bunch cannot be monitored with a pick-up, as demonstrated in Fig. 108 for a 1.4 MeV/u ion beam corresponding to $\beta = 5.5\%$. In addition, the required resolution of 1° in phase, or less than 50 ps in time, corresponds to a bandwidth of more than 10 GHz, which is hardly achieved by RF technologies. Other methods must be applied for such a measurement. The result is important for the adjustment of linac modules, including bunchers for injection to subsequent accelerators, and the comparison of measurements and beam dynamics calculations, as, e.g., required during a commissioning phase of a new accelerator. The bunch width and structure enter numerical codes in the same manner as the transverse parameters, e.g., for the estimation of emittance growth by space charge forces. Moreover, as is shown next, a Gaussian distribution is not a good approximation of the bunch structure at these low energies.

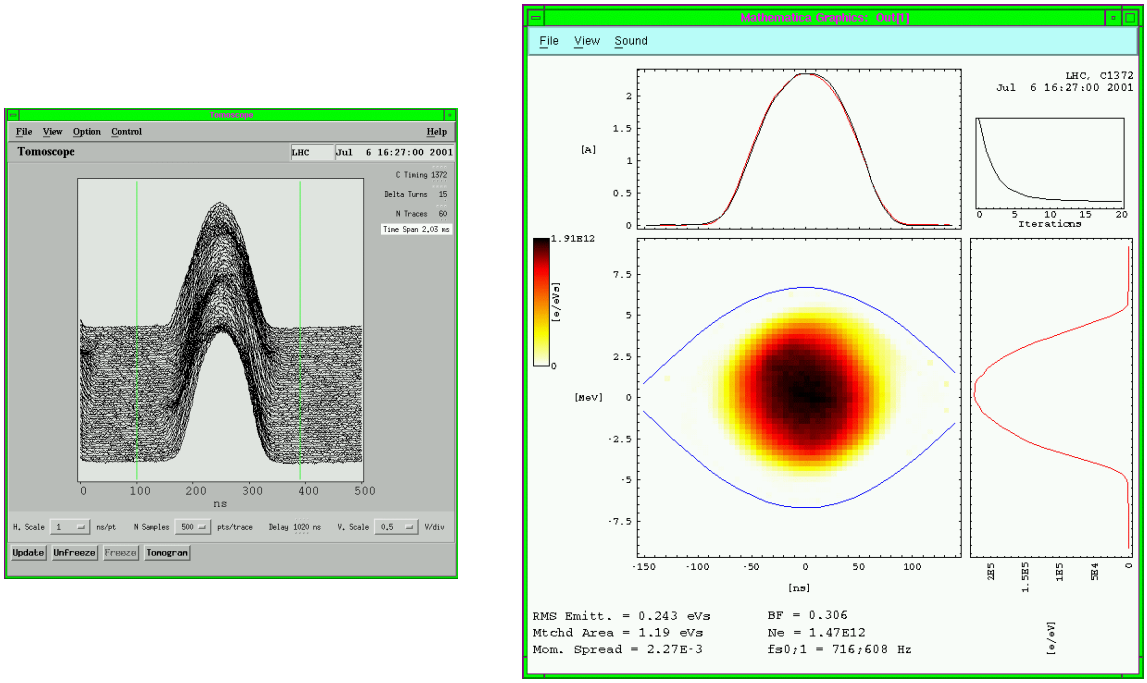


Fig. 106: Left: Bunch shape recorded at CERN PS using a resistive wall current monitor for 500 turns. Right: Reconstructed longitudinal phase space. The bucket size is shown in the phase space and the projections to the time axis (top) and energy (right) are displayed [142, 143].

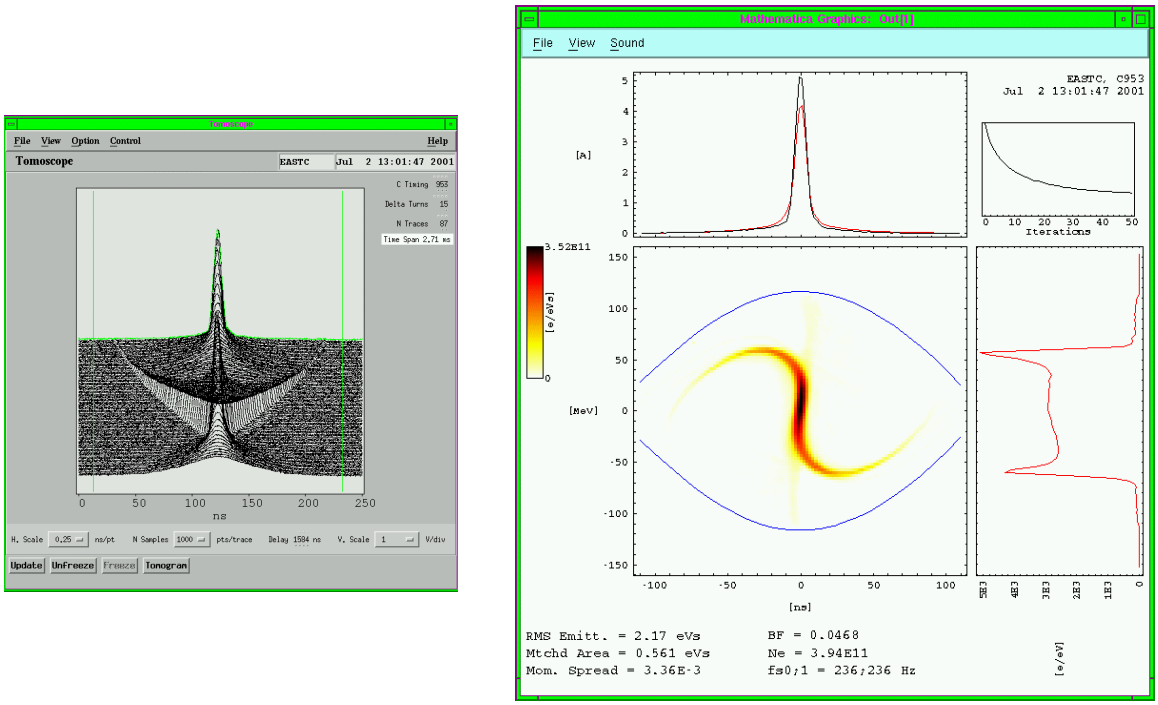


Fig. 107: Figure 106 except for a mismatched bunch. Left: Complex rotation of the bunch width for a specific time slice. Right: Reconstructed phase space density for the same time slice. The filamented distribution rotates to yield varying bunch shapes [142, 143].

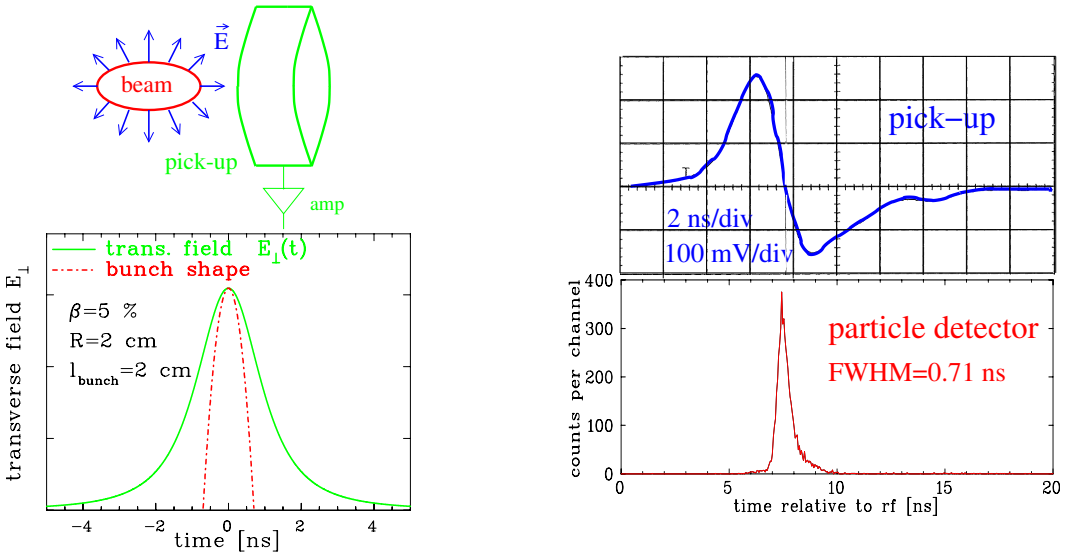


Fig. 108: Left: Simulation of time evolution of transverse electric field, as seen by the pick-up, for a bunch with $\cos^2$ distribution. Right: Comparison of pick-up signal with bunch structure, as measured using a particle detector at the GSI linac with 1.4 MeV/u. Broadening of the pick-up signal due to the longitudinal field component is clearly visible.

At most low-energy proton or ion linac facilities, a device for bunch structure observation, the so-called bunch structure monitor, is installed, based on the observation of secondary electrons liberated from a single wire intersecting the beam, as shown in Figs. 109 and 110 and reviewed in Refs. [144, 145]. The beam hits a wire, which is held at a negative potential of typically -10 kV. The liberated secondary electrons are accelerated by the related electric field towards a thin slit in the side of the vacuum chamber. Here, they pass a pair of plates, across which an RF field with the same frequency as the accelerating RF, or a higher harmonic of it, is applied. This field deflects the electrons; the differences in arrival times are transformed via varying deflection angles to differences in space, i.e., this deflector acts as a 'time-to-space converter'. (The same principle is used in an analogue oscilloscope.) After a drift of typically 0.5 m, the electrons are detected using a Faraday cup or secondary electron multiplier with a small slit in front. For a fixed slit position, the phase of the deflecting RF relative to the accelerating RF can be changed to transmit electrons representing different time slots of the beam bunches, giving a full image of the bunch structure.

The resolution of this device is better than 1° in phase, independent of the frequency of the RF, up to several hundred megahertz. It is known that the emission of secondary electrons is a fast process, with a time difference of less than 10 ps between the hit of the beam particle and electron emission, i.e., much faster than the required resolution. The RF deflector is built as a resonator to produce an electric field high enough for typically 100 mrad deflection to have an appropriate separation at the detector of $\sim$0.5 m. In addition, the electrons must be focused transversally. A constant d.c. voltage applied to the deflector plates serves as an electrostatic lens.

A typical measurement behind a low-energy linac module is shown in Fig. 111, which also acts as an example of the probable non-Gaussian shape typically created in high-current proton or ion linacs. The two measurements for high (top) and low (bottom) currents show quite different bunch shapes. A larger width, corresponding to a filamented phase space distribution, is observed for the low-current case. Particle-tracking calculation for this case confirms the strong ion current dependence of the longitudinal emittance. The applied RF power in the cavity counteracts the space charge force for a high-current beam. For a low current, filamentation occurs as the result of the missing damping by the space charge, which must be corrected by a smaller RF power. This shows that matching of the amplitude and phase for

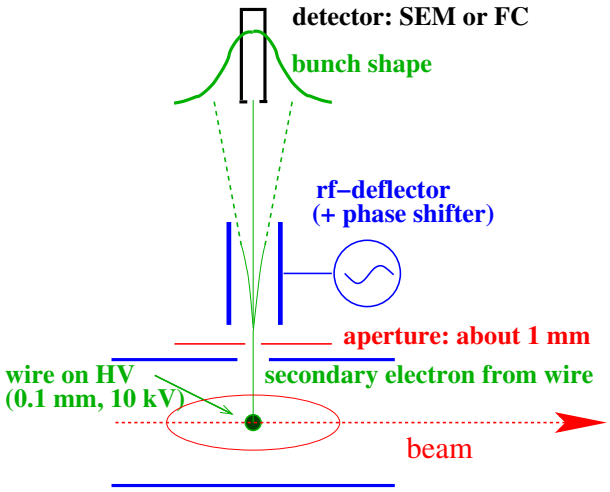


Fig. 109: Principle of deflector technique for bunch structure observation by bunch shape monitor. FC, Faraday cup; HV, high voltage; SEM, secondary electron multiplier.

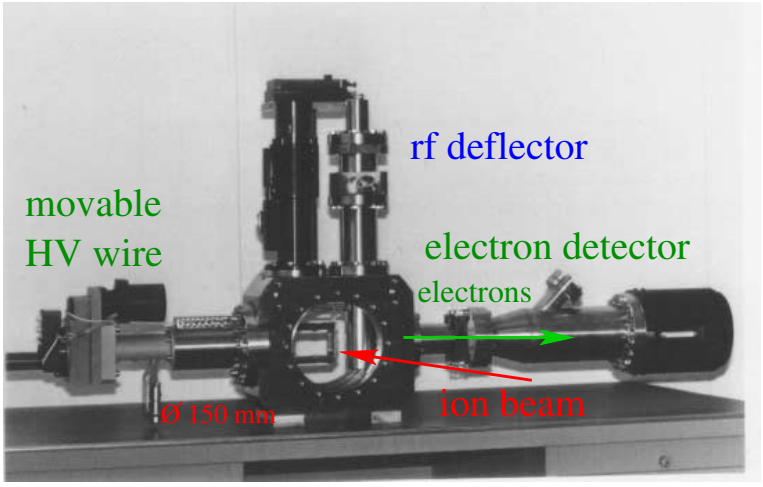


Fig. 110: Bunch structure monitor installed at CERN linac 2 [144, 145]. The vacuum feed-through on the left-hand side is the support for the high-voltage wire. The RF deflector is installed on the vertical flange. The detector is mounted at the right.

a linac cavity is of great influence on the longitudinal beam emittance and must therefore be measured and controlled to preserve beam quality.

7.7 Bunch structure for relativistic electron beams

In most cases, pick-ups are not used for longitudinal diagnostics at electron accelerators, although, owing to the relativistic velocities, this would be possible. However, the bunch length, e.g., at a synchrotron light source, is typically only of the order of several tens of picoseconds, so that even a pick-up bandwidth of several gigahertz is insufficient to reproduce the detailed bunch structure. Here we profit from the emission of synchrotron light at a dipole or, preferably, from an insertion device. A review of this technique is given in Ref. [147].

The principle of such a measurement is shown in Fig. 112. It uses a streak camera as a commercially available device [149], which allows visible light observations with a time resolution down to, typically, 1 ps. Synchrotron light in the optical to ultraviolet wavelength range emitted by the bent elec-

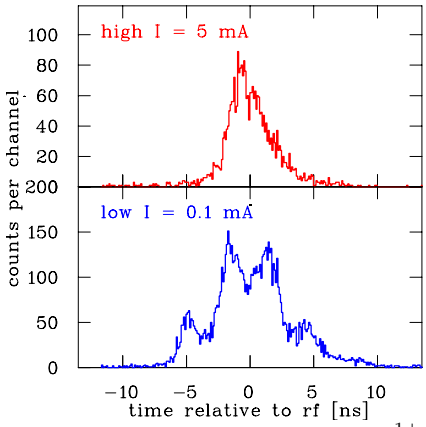


Fig. 111: Bunch structure measured at the GSI linac with a 120 keV/u Ar^{1+} beam for different currents inside the RFQ [146]; one RF period is shown.

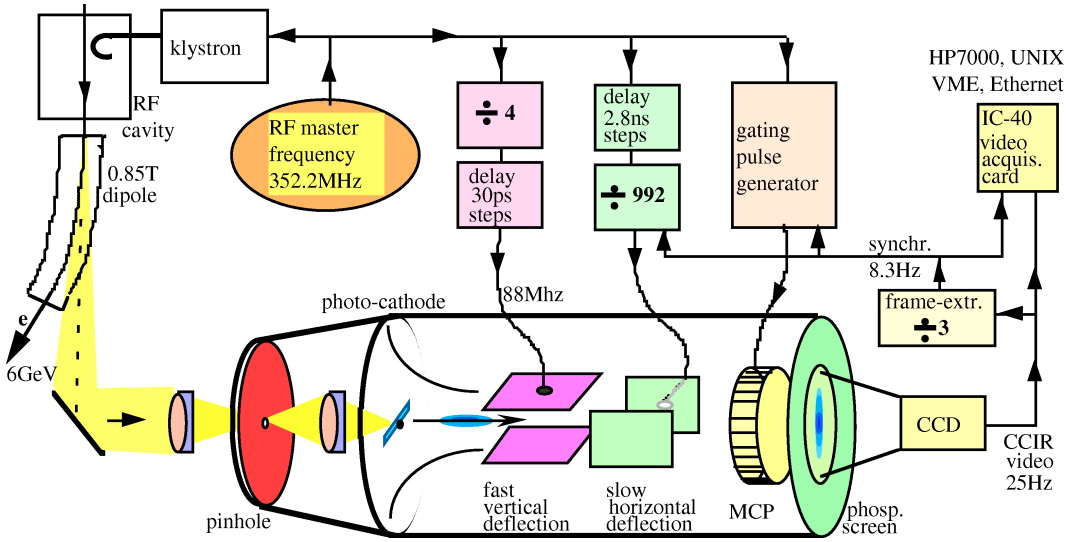


Fig. 112: Principle of streak camera measurement for bunch width determination at ESRF electron accelerator [148].

tron beam is used. It is focused and scraped by a pinhole onto the photocathode of a streak camera. The secondary electrons from the photocathode are accelerated and pass a fast deflector driven by a frequency in the range 50–300 MHz, which is locked to the accelerating frequency for synchronization. The general scheme is similar to the 'time-to-space converter' discussed in Section 7.6. After a certain drift space, the electrons are amplified by a multichannel plate and converted to photons by a phosphor screen. The image from the phosphor is captured using a standard CCD camera. The different arrival times of the synchrotron light, and therefore the bunch structure, are converted to differences in space, with a typical full scale of several 100 ps and a resolution in the picosecond range. A second, perpendicular, pair of plates is driven at a much lower frequency to separate the images from the individual bunches.

An example of a bunch length measurement, as collected using synchrotron radiation from a dipole, is depicted in Fig. 113 [150]. The slow scan for the separation of the individual bunches is displayed along the horizontal axis. The vertical axis has a full scale of 360 ps; the bunch width is only $\sigma \simeq 40$ ps, demonstrating the high resolution of such a system. Short bunches are needed by synchrotron radiation users for time-resolved spectroscopy. A large voltage is applied to the synchrotron RF cavities to produce and maintain these short bunches. In Fig. 114, the bunch width is plotted for different RF voltages. As displayed, the achievable bunch length depends on the stored electron current. Counter-

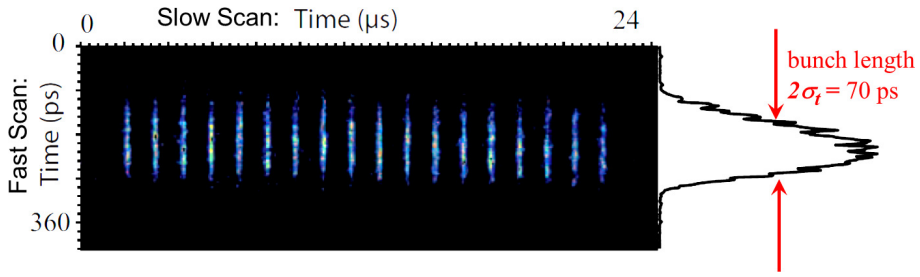


Fig. 113: Bunch length measurement with a streak camera, using synchrotron light from individual bunches emitted by passage through a dipole at SOLEIL [150]. The horizontal axis scaling is 24 μs full scale for the bunch repetition (slow scan direction); the vertical axis is 360 ps full scale for the bunch structure (fast scan direction).

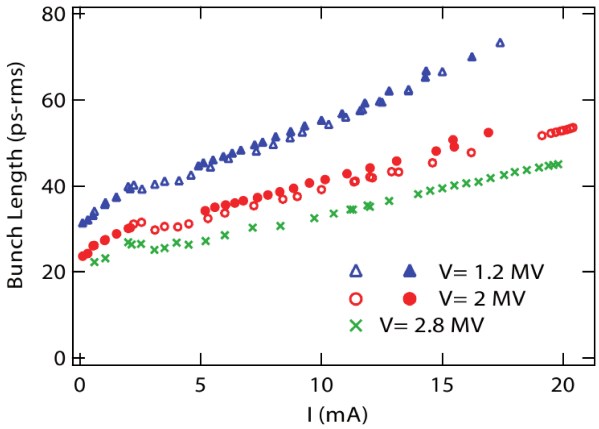


Fig. 114: Example of the use of a streak camera at the SOLEIL synchrotron light facility achieved from measurements like the one depicted in Fig. 113. For different accelerating voltages, the bunch length is recorded as a function of the beam current stored in the synchrotron [150].

acting the RF bunching force, the incidence of non-linear processes, such as wakefields and instabilities, increases for higher currents. Such effects as head–tail oscillations or longitudinal bunch oscillation due to coupled bunch instabilities are also made visible with the help of streak camera measurements; see, e.g., Ref. [147].

The longitudinal properties of the bunch injected to the storage ring from the booster ring of a synchrotron light facility can be monitored using a streak camera to enable adequate correction. The results of such a measurement, made at ALS (Berkeley) [151, 152], for injection into an empty storage ring, are depicted in Fig. 115, with injection energy $E_{\mathrm{kin}} = 1.5$ GeV, revolution time $t_{\mathrm{rev}} = 656$ ns, accelerating frequency $f_{\mathrm{acc}} = 499.7$ MHz and, for the chosen RF amplitude, synchrotron frequency $f_{\mathrm{synch}} = 11.4$ kHz, corresponding to a synchrotron tune $Q_{\mathrm{synch}} = t_{\mathrm{rev}} \cdot f_{\mathrm{synch}} = 0.0075$, i.e., a longitudinal repetition of the oscillations every $n = 1/Q_{\mathrm{synch}} = 133$ turns. The longitudinal damping time is about 10 ms, i.e., much longer than the depicted observation. In the first displayed case, a single bunch is injected with a phase error, i.e., the bunch centre does not pass the RF cavity during the zero-crossing of the sine wave. The particles execute coherent synchrotron oscillations, which correspond to a rotation in the longitudinal phase space that is measurable via a variation of the bunch centre. The projection along the time axis of the bunch distribution is a function of the longitudinal forces; this is measured as a variation in bunch length. The same movement occurs when the bunch is injected with an incorrect energy, except that a phase change of the synchrotron oscillation of 90° is visible for the bunch central position (middle part of Fig. 115). The remedy for these mismatches is a proper adjustment of the RF phase relation between the booster and the storage ring in the first case and a realignment of the storage ring energy in the second case. This is important to prevent emittance enlargement by filamentation in

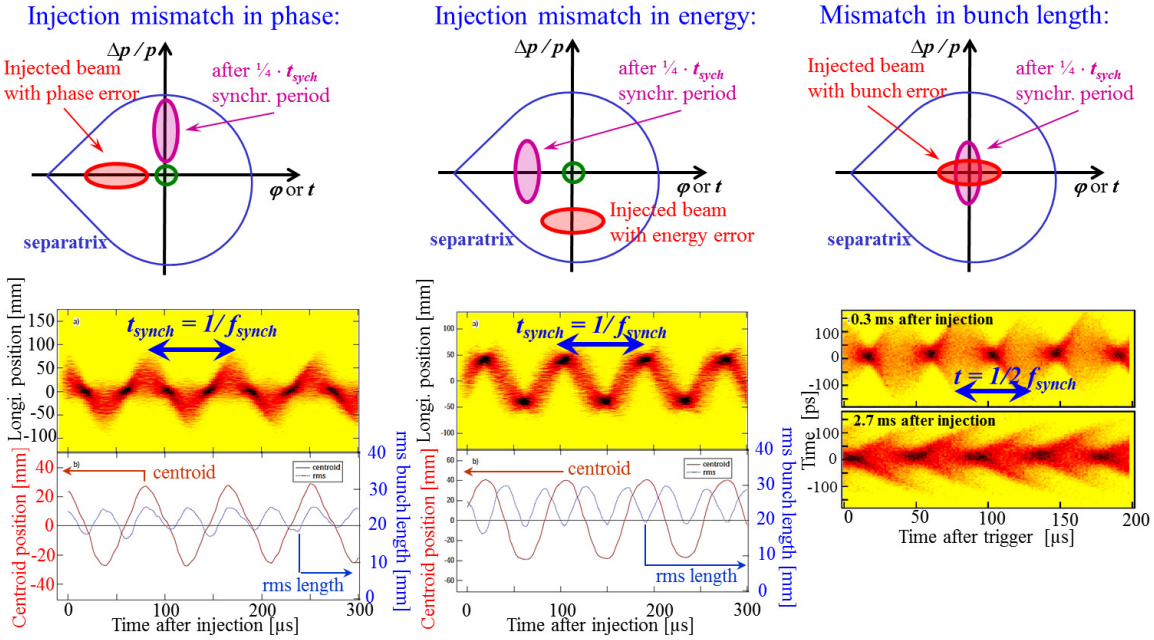


Fig. 115: Measurement of longitudinal injection mismatch at the ALS light source. Top: Schemes of injection mismatch in phase, energy, and bunch length. Bottom: Bunch observation using streak camera with evaluated bunch centre and width for each case. Data and plots are from Refs. [151, 152].

phase space for long storage times, which is not completely compensated by longitudinal damping and influences the circulating beam properties if a top-up injection scheme is used. The right-hand panel of Fig. 115 shows the case of an injection with correct phase and energy but unmatched bunch length. The synchrotron movement leads to variation of the symmetrical bunch distribution except with double frequency, $2 \cdot f_{\mathrm{synch}}$. After some storage time, the bunch shape changes, owing to non-linear forces created by the RF wave, leading to filamentation in phase space, as shown for the observations at 0.3 ms and 2.7 ms after injection. More details on this instructive example, as well as a comparison with simulations, are given in Refs. [151, 152]. A streak camera is used to monitor the short bunches present at a synchrotron light facility, in the display case of $\sigma_{\mathrm{bunch}} = 30$–$100$ ps. For a hadron synchrotron with a bunch length of typically $\sigma_{\mathrm{bunch}} = 1$–$100$ ns, a comparable behaviour could be observed using a fast current transformer.

8 Conclusion

The functionality of various beam monitoring instruments has been described and the applicability for different beam parameters has been discussed. For regular operation of a transfer line, current, transverse profile, and beam position measurements are mainly used for possible manual or automated beam alignment. It is an advantage if the measurement is non-destructive for the beam, as this enables simultaneous measurement at several locations. This is, e.g., realized by current transformers, if the beam parameters allow for their usage. For smaller currents, e.g., for beams extracted slowly from a synchrotron, destructive detectors must be used; these are based on the energy loss of the particles in a gas or metal. For profile measurements, various techniques are applied; scintillation and optical transition radiation screens are favoured techniques for regular operation, owing to their reproduction of the two-dimensional distribution. Both SEM grids and wire scanners are also often used, owing to their robustness and large dynamic range. For bunched beams, the transverse centre position can be recorded using BPMs, which serve as the most frequently used signal source for beam alignment. Owing to their principle, BPMs are non-destructive but require a minimum beam current and are not suited for unbunched beams, e.g.,

for slow extraction. At transport lines behind a linac or a cyclotron, more complex parameters must be determined, in particular, transverse and longitudinal emittances. The related measurements are, in many cases, made by experts during dedicated machine development times; this allows for more complex instruments and methods. Alignment in the injection process involves using beam instrumentation installed in the synchrotron. An example is the control of proper timing by observing the beam current and its time structure, in the transfer line and the stored beam, using current transformers. A further example is the observation of transverse or longitudinal profiles and their evolution during storage. The long observation time during storage enables precise determination of several beam parameters and the control of such injection settings as steering, focusing by transfer line magnets and the timing properties of the injection devices. The beam's parameters are also determined prior to extraction of a beam from a synchrotron. In many facilities, beam loss monitors are installed to ensure correct accelerator functionality on an online basis; an interlock is generated if there is any malfunction. Owing to their sensitivity, beam loss monitors can also be used to optimize beam transmission during regular operation, serving as an alternative to current transformers.

Acknowledgements

Valuable discussions with A. Reiter and R. Singh from GSI, Darmstadt, and M. Schuh from KIT, Karlsruhe, as well as the careful review of the paper by S. Mirza from GSI, are warmly acknowledged.

References

[1] V. Smaluk, *Particle Beam Diagnostics for Accelerators: Instruments and Methods* (VDM Verlag Dr. Müller, Saarbrücken, 2009).

[2] Beam diagnostics for accelerators, in Proceedings of the CAS-CERN Accelerator School: Course on Beam Diagnostics, Dourdan, France, 28 May-6 Jun 2008, edited by D. Brandt, CERN-2009-005, (CERN, Geneva, 2009), https://doi.org/10.5170/CERN-2009-005.

[3] P. Strehl, *Beam Instrumentation and Diagnostics* (Springer Verlag, Berlin, 2006).

[4] M.G. Minty and F. Zimmermann, *Measurement and Control of Charged Particle Beams* (Springer Verlag, Berlin, 2003).

[5] S-I. Kurokawa *et al.* (Eds.), Proc. School on Beam Measurement, Proceeding Montreux (World Scientific, Singapore, 1999).

[6] J.M. Month and S. Turner (Eds.), *Frontiers in Particle Beams; Observation, Diagnosis and Correction* (Springer Verlag, Berlin, 1988).

[7] P. Forck *JUAS Lecture Notes on Beam Instrumentation and Diagnostics*, www.gsi.de/work/gesamtprojektleitung_fair/commons/beam_instrumentation/research_and_development_lobird/veroeffentlichungen.htm

[8] D. Lipka *et al.*, Proc. IBIC 16, Barcelona, 2016, p. 14.

[9] V. Schlott *et al.*, Proc. IBIC 15, Melbourne, 2015, p. 12.

[10] H. Loos, Proc. IBIC 14, Monterey, 2014, p. 475.

[11] www.bergoz.com

[12] D. Belohrad, Proc. DIPAC 11, Hamburg, 2011, p. 564.

[13] H. Reeg, Proc. DIPAC 01, Grenoble, 2001, p. 120.

[14] R.C. Webber, Proc. BIW 94, Vancouver, 1994, p. 3.

[15] K. Unser, *IEEE Trans. Nucl. Sci.* **NS-28** (1981) 2344, https://doi.org/10.1109/TNS.1981.4331686.

[16] W.R. Leo, *Techniques for Nuclear and Particle Physics Experiments: A How-to Approach* (Springer Verlag, Berlin, 1994).

[17] G.F. Knoll, *Radiation Detection and Measurement* (John Wiley, New York, 1999).

[18] J.D. Jackson, *Classical Electrodynamics* (John Wiley & Sons, New York, 1998).

[19] F. Ziegler *et al. The Stopping and Range of Ions in Solids (Stopping and Range of Ions in Matter, vol. 1)* (Pergamon Press, New York, 1985).

[20] www.SRIM.org

[21] lise.nscl.msu.edu/lise.html

[22] http://physics.nist.gov/PhysRefData/Star/Text/ESTAR.html

[23] A.F.D. Morgan, Proc. DIPAC 05, Lyon, 2005, p. 51.

[24] E.J. Sternglass, *Phys. Rev.* **108** (1957) 1, https://doi.org/10.1103/PhysRev.108.1.

[25] C.G. Drexler and R.D. DuBois, *Phys. Rev.* **A53** (1996) 1630, https://doi.org/10.1103/PhysRevA.53.1630.

[26] P. Finocchiaro *et al.*, Proc. DIPAC 97, Frascati, 1997, p. 53.

[27] P. Forck *et al.*, Proc. DIPAC 97, Frascati, 1997, p. 165.

[28] P. Forck *et al.*, Proc. BIW 96, Argonne, 1996, p. 422.

[29] P. Lecoq *et al.*, *Inorganic Scintillators for Detector Systems* (Springer Verlag, Berlin, 2006).

[30] F. Sauli, Principle of operation of multiwire proportional and drift chambers, CERN-77-09 (CERN, Geneva, 1977), https://doi.org/10.5170/CERN-1977-009.

[31] H. Reeg, Proc. DIPAC 99, Chester, 1999, p. 141.

[32] C. Schömers *et al.*, *Nucl. Instrum Methods Phys. Res.* **A795** (2015) 92, https://doi.org/10.1016/j.nima.2015.05.054.

[33] C. Schömers *et al.*, Proc. IPAC 11, San Sebastian, 2011, p. 2851.

[34] C. Schömers *et al.*, Proc. IPAC 13, Shanghai, 2013, p. 2944.

[35] G. Ferioli and R. Jung, Proc. DIPAC 97, Frascati, 1997, p. 168.

[36] K. Wittenburg, Proc. EPAC 02, Paris, 2002, p. 109.

[37] K. Wittenburg, in Proceedings of the CAS-CERN Accelerator School: Course on Beam Diagnostics, Dourdan, France, 28 May-6 Jun 2008, edited by D. Brandt, CERN-2009-005, (CERN, Geneva, 2009), pp. 249-280, https://doi.org/10.5170/CERN-2009-005.249.

[38] A. Zhukov, Proc. BIW 10, Santa Fe, 2010, p. 553.

[39] R.E. Shafer, Proc. BIW 02, Brookhaven AIP 648, 2002, p. 44.

[40] A.H. Sullivan, *A Guide to Radiation and Radioactivity Level Near High Energy Particle Accelerators* (Nuclear Technology Publishing, Ashford, 1992).

[41] S. Roesler *et al.*, in *Handbook of Acceleration Physics and Engineering*, Eds. A.W. Chao *et al.* (World Scientific, Singapore, 2013), p. 767.

[42] P. Forck and T. Hoffmann, Proc. DIPAC 01, Grenoble, 2001, p. 129.

[43] D. Kramer *et al.*, Proc. DIPAC 07, Venice, 2007, p. 313.

[44] D. Kramer, Ph.D. thesis, Technical University of Liberec, 2008.

[45] R.E. Shafer *et al.*, Proc. Intern. Conf. High Energy Acc., Batavia, 1983, p. 609.

[46] D. Gassner *et al.*, Proc. BIW 00, Cambridge, 2000, p. 392.

[47] B. Dehning, Proc. DIPAC 05, Lyon, 2005, p. 117.

[48] E.B. Holzer *et al.*, Proc. EPAC 08, Genoa, 2008, p. 1134.

[49] M. Stockner *et al.*, Proc. DIPAC 07, Venice, 2007, p. 328.

[50] http://wwwinfo.cern.ch/asd/geant4/

[51] A. Zhukov, Proc. PAC 13, Pasadena, 2013, p. 714.

[52] R. Jung *et al.* Proc. DIPAC 03, Mainz, 2003, p. 28.

[53] R. Jung, Proc. DIPAC 93, Montreux, 1993, p. 4.

[54] B. Walasek-Höhne *et al.*, Proc. DIPAC 11, Hamburg, 2011, http://www-

bd.gsi.de/ssabd/index.html

[55] E. Gütlich *et al.*, Proc. BIW 08, Lake Tahoe, 2008.

[56] E. Gütlich *et al.*, Proc. DIPAC 11, Hamburg, 2011, p. 179.

[57] G. Kube *et al.*, Proc. PAC 12, New Orleans, 2012, p. 2119.

[58] U. Iriso *et al.*, Proc. DIPAC 09, Basel, 2009, p. 200.

[59] R. Ischebeck *et al.*, Proc. IBIC 14, Monterey, 2014, p. 259.

[60] R. Ischebeck *et al.*, *Phys. Rev. S.T. Accel. Beams* **18** (2015) 082802, https://doi.org/10.1103/PhysRevSTAB.18.082802.

[61] T.J. Shea *et al.*, Proc. PAC 97, Vancouver, 1997, p. 2215.

[62] P. Forck *et al.*, Proc. IPAC 14, Dresden, 2014, p. 3482.

[63] A. Lieberwirth *et al.*, *Nucl. Instrum. Methods Phys. Res.* **B365** (2015) 533. https://doi.org/10.1016/j.nimb.2015.07.111

[64] P. Forck *et al.*, Proc. DIPAC 11, Hamburg, 2011, p. 170.

[65] R. Haseitl *et al.*, Proc. PCaPAC 2008, Ljubljana, 2008, p. 180.

[66] M. Plum, Proc. BIW 04, Knoxville, 2004, p. 23.

[67] S. Burger *et al.*, Proc. DIPAC 03, Mainz, 2003, p. 122.

[68] C. Fischer *et al.*, Proc. EPAC08, Rome, 1988, p. 1081.

[69] U. Raich, Proc. DIPAC 05, Lyon, 2005, p. 1.

[70] R.J. Colchester and R. Jung, Proc. PAC 97, Vancouver, 1985, p.1917.

[71] C. Field, *Nucl. Instrum. Methods Phys. Res.* **A360** (1995) 467, https://doi.org/10.1016/0168-9002(95)00077-1.

[72] M. Matoba *et al.*, *IEEE Trans. Nucl. Sci.* **NS-32** (1985) 541, https://doi.org/10.1109/TNS.1985.4336890.

[73] P. Forck, Proc. IPAC 10, Kyoto, 2010, p. 1261.

[74] P. Forck *et al.*, Proc. DIPAC 05, Lyon, 2005, p. 221.

[75] T. Giacomini *et al.*, Proc. BIW 04, Knoxville,2004, p. 286.

[76] T. Giacomini *et al.*, Proc. DIPAC 05, Lyon, 2005, p. 150.

[77] P. Forck and P. Strehl, Proc. DIPAC 99, Chester, 1999, p. 42.

[78] P. Forck *et al.*, Proc. BIW 00, Cambridge AIP 546, 2000, p. 606.

[79] R. Connolly *et al.*, Proc. BIW 00, Cambridge AIP 546, 2000, p. 330.

[80] G. Ferioli *et al.*, Proc. DIPAC 01, Grenoble, 2001, p. 201.

[81] M. Benedikt *et al.*, Proc. DIPAC 01, Grenoble, 2001, p. 189.

[82] C. Bovet *et al.*, Proc. EPAC 98, Stockholm, 1998, p. 1488.

[83] K. Satou *et al.*, Proc. High Brightness Hadron Beam Conf. HB 10, Morschach, 2010, p. 506.

[84] K. Satou *et al.*, Proc. EPAC 08, Genoa, 2008, p. 1275.

[85] T. Schottmann, Thesis DESY HERA 93-09, 1993.

[86] K. Satou, Proc. IBIC 17, Grand Rapids, 2017.

[87] D. Vilsmeier *et al.*, Proc. IBIC 17, Grand Rapids, 2017.

[88] M. Sapinski *et al.*, Proc. IBIC 16, 2016, p. 520.

[89] A. Allisy (Ed.), Secondary electron spectra from charged particle interaction, International Commission on Radiation Units and Measurement Report No. 55 (1996).

[90] L. Wartski *et al.*, *J. Appl. Phys.* **46** (1975) 3644, https://doi.org/10.1063/1.322092.

[91] J. Bosser *et al.*, *Nucl. Instrum. Methods Phys. Res.* **A 238** (1985) 45, https://doi.org/10.1016/0168-9002(85)91025-3.

[92] C. Bovet *et al.*, Proc. DIPAC 99, Chester, 1999, p. 90.
[93] V.E. Scarpine *et al.*, Proc. PAC 07, Albuquerque, 2007, p. 2639.
[94] G. Kube, Proc. DIPAC 07, Venice, 2007, p. 6.
[95] M. Wilke, Proc. BIW 94, Vancouver, 1994, p. 128.
[96] C. Bovet *et al.*, Proc. PAC 91, San Francisco, 1991, p. 1160.
[97] B.X. Yang *et al.*, Proc. PAC 97, Vancouver, 1997, p. 2215.
[98] B.K. Scheidt, Proc. DIPAC 05, Lyon, 2005, p. 24.
[99] M. Pont *et al.*, Proc. IPAC 11, San Sebastian, 2011, p. 3023.
[100] G. Benedetti *et al.*, Proc. IPAC 11, San Sebastian, 2011, p. 2059.
[101] T. Mitsuhashi, Beam size measurement by use of SR interferometers, Proc. School on Beam Measurement, Montreux, p. 399 (World Scientific, Singapore, 1999).
[102] T. Mitsuhashi, Proc. BIW 04, Knoxville, 2004, p. 3.
[103] J.M. Stockli, Proc. BIW 06, Batavia, 2006, p. 25.
[104] M. Reiser, *Theory and Design of Charged Particle Beams* (J. Wiley Publishing, New York, 1994).
[105] K.T. McDonald and D.P. Russell, in *Frontiers of Particle Beams; Observation, Diagnosis and Correction*, Eds. M. Month and S. Turner (Springer Verlag, Berlin, 1988), p. 122.
[106] G. Riehl *et al.*, Proc. DIPAC 95, Travemünde, DESY M9507, 1995, p. 6.
[107] G. Penco *et al.*, Proc. EPAC 08, Genoa, 2008, p. 1236.
[108] M.E. Schulze *et al.*, Proc. LINAC 00, Monterey, 2000, p. 575.
[109] V.A. Dimov *et al.*, Proc. High Brightness Hadron Beam Conf. HB 16, Malmö, 2016, p. 433.
[110] R.E. Shafer, Proc. BIW 89, Upton, 1989, p. 26, www.bergoz.com/en/mx-bpmf.
[111] S.R. Smith, Proc. BIW 96, Argonne AIP 390, 1996, p. 50.
[112] P. Forck, in Proceedings of the CAS-CERN Accelerator School: Course on Beam Diagnostics, Dourdan, France, 28 May-6 Jun 2008, edited by D. Brandt, CERN-2009-005, (CERN, Geneva, 2009), pp. 187-228, https://doi.org/10.5170/CERN-2009-005.187.
[113] M. Wendt, Proc. DIPAC 11, Hamburg, 2011, p. 18.
[114] G. Kube and M. Werner, Proc. DIPAC 07, Venice, 2007, p. 105.
[115] B.G. Pine, CARE-ABI Workshop, Lüneburg, 2006, adweb.desy.de/mdi/CARE/Lueneburg/ABI-Lueneburg.htm.
[116] R.E. Shafer, Proc. BIW 93, Santa Fe, 1993, p. 303.
[117] C. Simon *et al.*, Proc. IBIC 14, Monterey, 2014, p. 303.
[118] M. Almaki *et al.*, Proc. IBIC 14, Barcelona, 2016, p. 319.
[119] C. Boccard, CARE-ABI Workshop, Lüneburg, 2006, adweb.desy.de/mdi/CARE/Lueneburg/ABI-Lueneburg.htm.
[120] E. Calvo-Giraldo *et al.*, Proc. DIPAC 03, Mainz, 2003, p. 187.
[121] P. Kowina *et al.*, Proc. DIPAC 05, Lyon, 2005, p. 114.
[122] R. Lorenz, Proc. BIW 98, Stanford AIP 451, 1998, p. 53.
[123] G. Vismara, Proc. BIW 00, Cambridge AIP 546, 2000, p. 36.
[124] G. Vismara, Proc. DIPAC 99, Chester, 1999, p. 11.
[125] M. Wendt, Proc. IBIC 14, Monterey, 2014, p. 468.
[126] R. Garoby, Low level RF building blocks, CERN Acc. School, CERN 93-03, 1992, p. 428.
[127] www.i-tech.si
[128] J.M. Belleman, Proc. DIPAC 05, Lyon, 2005, p. 137.
[129] J. Belleman, in Proceedings of the CAS-CERN Accelerator School: Course on Beam Diagnostics,

Dourdan, France, 28 May-6 Jun 2008, edited by D. Brandt, CERN-2009-005, (CERN, Geneva, 2009), pp. 281-316, https://doi.org/10.5170/CERN-2009-005.281.

[130] V. Schlott *et al.*, Proc. EPAC 00, Vienna, 2000, p. 1809.

[131] R. Jones, Proc. BIW 10, Santa Fe, 2010, p. 22.

[132] J.D. Fox and E. Kikuzani, Bunch feedback system and signal processing, Proc. School on Beam Measurement, Montreux (World Scientific Singapore, 1999), p. 579.

[133] R. Steinhagen, in Proceedings of the CAS-CERN Accelerator School: Course on Beam Diagnostics, Dourdan, France, 28 May-6 Jun 2008, edited by D. Brandt, CERN-2009-005, (CERN, Geneva, 2009), pp. 317-360, https://doi.org/10.5170/CERN-2009-005.317.

[134] J. Wenninger, in Proceedings of the CAS-CERN Accelerator School: Course on Beam Diagnostics, Dourdan, France, 28 May-6 Jun 2008, edited by D. Brandt, CERN-2009-005, (CERN, Geneva, 2009), pp. 361-376, https://doi.org/10.5170/CERN-2009-005.361.

[135] K. Fuchsberger *et al.*, Proc. IPAC 10, Kyoto, 2010, p. 462.

[136] M. Meddahi *et al.*, Proc. IPAC 10, Kyoto, 2010, p. 1680.

[137] H. Wollnik, *Optics of Charged Particles* (Academic Press, Orlando, 1987).

[138] J.C. Dooling *et al.*, Proc. LINAC 00, Monterey, 2000, p. 193.

[139] P. Forck *et al.*, Proc. DIPAC 05, Lyon, 2005, p. 48.

[140] S. Hancock *et al.*, *Phys. Rev. Special Topics Accel. Beams* **3** (2000) 124202, https://doi.org/10.1103/PhysRevSTAB.3.124202.

[141] CERN-PS-2000-068-OP (CERN, Geneva, 2000).

[142] http://cern.ch/tomography

[143] S. Hancock *et al.*, CERN PS/RF Note 2001-010 (CERN, Geneva, 2001).

[144] A. Feschenko, Proc. PAC 01, Chicago, 2001, p. 517.

[145] A. Feschenko, Proc. RuPAC 12, Saint Petersburg, 2012, p. 181.

[146] P. Forck *et al.*, Proc. LINAC 00, Monterey, 2000, p. 166.

[147] K. Scheidt, Proc. EPAC 00, Vienna, 2000, p. 182.

[148] K. Scheidt, Proc. EPAC 96, Sitges, 1996, p. 1624.

[149] www.hamamatsu.com

[150] M. Labat *et al.*, Proc. DIPAC 07, Venice, 2007, p. 241.

[151] J.M. Byrd *et al.*, Proc. PAC 99, New York, 1999, p. 382.

[152] J.M. Byrd and S. De Santis, Longitudinal injection transients in an electron storage ring, 2001, http://escholarship.org/uc/item/8k6677nw.

Proceedings of the CAS–CERN Accelerator School: Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018)

Low-Energy Beam Transport Design

R. Baartman
TRIUMF, Vancouver, Canada

Abstract

There is a trick to designing beam transport systems. In most cases, we are more interested in assuring that higher-order aberrations are small than we are in finding accurate values for all the transfer map elements. I show a simple way to do this, using an envelope technique and modifying the usual optimization parameter.

Keywords

Solenoid; einzel lens; electrostatic; quadrupole; aberration.

1 Introduction

In most cases, the goal of a design is to transport a well-matched particle beam with no losses and no decrease in quality. 'Quality' here, of course, means emittance, and maintaining it means that higher-order effects are negligible. As 'higher-order' is difficult to calculate, requiring seemingly much more input, such as Enge coefficients or field maps, the tendency is to design 'linear' first and then, if at all, run it through a time-consuming higher-order calculation using a multiparticle code and field maps or Enge coefficients. At this point, the elements have usually already been designed and there is resistance to changing them.

I intend to show that there is a better way. 'Better' means not only that it is quicker to calculate but also that, being very quick, it is a simple matter to optimize acceptance, usually bringing the emittance growths down to the level of negligible. The technique uses simple formulae that find the emittance growths at each element and simply add them in quadrature to that arising from mismatch, minimizing this sum while varying gross features, such as quadrupole lengths and solenoid apertures.

A major issue with the usual approach is that field maps are found, and perhaps summarized, by fitting their form in the 'fringe' region to a high-order polynomial only after the elements have already been designed. They are not designed as part of the optimization process. Very often, they are designed using folkloric rules-of-thumb regarding aperture occupation or field flatness, in place of a real understanding of the origin of non-linear forces. This has the effect of perpetuating such rules.

The standard beam transport elements for low energy are: dipoles, both electrostatic and magnetic; quadrupoles, also of both types, and solenoids and einzel lenses.

History: the mathematical formalism for describing charged-particle beam transport is only 60 years old. First came the understanding that particle optics of individual particles could be described by transfer matrices [1], then applied to ensembles of particles with the σ-matrix formalism developed by Brown [2]. The first useful codes were `TRANSPORT` [3], `TRACE3D` [4],

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-005.491

and GIOS [5], whose first versions were written in the mid-1960s. These used highly idealized 'hard-edged' elements and so were often not sufficiently accurate. The result was that beamlines designed in this way were expected to work only approximately and allowed to depart from 'theoretical' settings, sometimes by wide margins. There developed a prejudice that the σ-matrix or 'envelope' technique was not accurate. Further, it was thought that this technique could never handle space charge very well. Both these turned out to be incorrect but the prejudice remains.

When computing power grew by many orders of magnitude, designers tended towards codes that integrated motions of millions of individual particles. So many particles, that even with today's computing power, it can take minutes or hours of CPU time on a personal computer to simulate beam transport. In making this step, from simple matrix-based codes that calculated 'beam envelopes', to codes that integrate millions of particles through stored magnetic and electric fields, the 'baby' (envelope technique) was thrown out with the 'bathwater' (hard-edged elements with no fringe-field effects).

The thinking behind this is that if we calculate through well-developed real fields, we automatically take all relevant effects (fringe fields, space charge, time-varying fields) into account. This is true, but it comes at a cost, namely, a loss of any feeling for basic effects, such as changing aperture or shaping fringe fields.

A second, well-regarded technique is to calculate transfer maps to high order, using, for example, differential algebraic tools, as in COSY-∞ [6]. Again, this is an extremely helpful tool but its use comes at the cost of loss of 'feel': for any given system, one can find accurate maps and even minimize the worst higher-order effects but the reverse problem, of knowing how small the aberrations need to be and then finding an optimum transport system, is not tractable.

So both techniques (integrating through stored fields and higher-order transfer maps) give good accurate and trustworthy results 'after the fact', i.e., after a beamline design has already been completed, perhaps even built, but are not used in guiding a beamline design from scratch.

The key is to realize that, except in special cases like spectrometers, where higher-order aberrations are inevitable, we don't actually 'need' exact higher-order maps, we only need to be assured that their effects are 'sufficiently small'. As I will show, the minimization techniques to make this assurance are surprisingly simple.

Let us first determine what is meant by 'sufficiently small'.

The key parameter is, of course, the emittance. A higher-order distortion is shown in Fig. 1. This may be a cubic (as is common for focusing devices where the fields are odd functions of transverse co-ordinates) or parabolic in the case of the dominant error from dipoles. (Early versions of TRANSPORT extended it to only second-order and so missed this effect entirely. More on this later.)

The order of the dominant effect is not particularly important for our purposes. Due to the higher orders the emittance ellipse starts to deform, Fig. 1 red curve. In case of a misalignment of the beam this effect can lead to an additional outspreading and deformation of the ellipse, which is called filamentation [Edwards, "An introduction to the physics of high energy accelerators", chap.7 "Emittance Preservation"]. The order of the dominant effect is not particularly important for our purposes; only the location of the 'edge' of the desired emittance matters. If desired, the

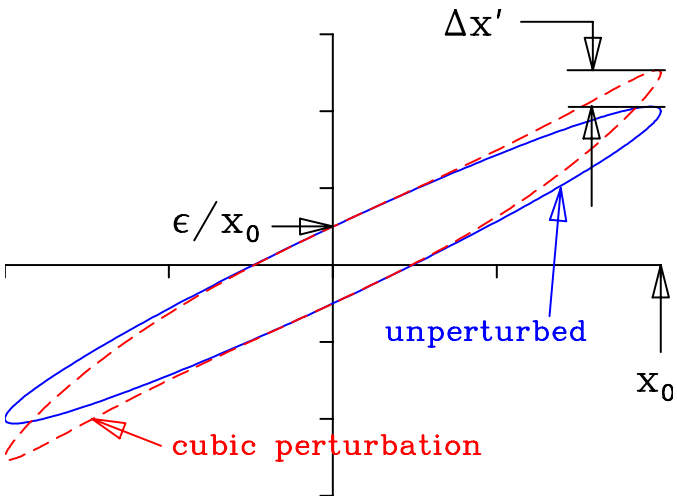


Fig. 1: Phase-space ellipse with cubic distortion

expected misalignment can be included in the 'size'. A high-quality beam may have negligible emittance growth if it is aligned to the beamline symmetry axis, but not so when misaligned.

Let us call the integrated higher-order error $\Delta x'$. Figure 1 makes clear that this x' is to be compared not with the beam divergence $\hat{x'}$, but with the local divergence $\epsilon/\hat{x}$, which may be significantly smaller: the two versions agree only at a waist.

The effective fractional growth in emittance is

$$\frac{\Delta\epsilon}{\epsilon} \sim \frac{\Delta x'\,\hat{x}}{\epsilon}, \tag{1}$$

and we want this to be small compared with one. How small depends on context: typically 1% is sufficient but clearly if there are thousands of elements, we would want this to be much smaller than 1%. Most often, the emittance growth is dominated by one or two focusing elements; the ones that create the final focus.

In any beamline design, there are many constraints besides minimizing emittance growth. The main constraint is to achieve a desired match or 'Twiss' parameters at the final focus. This is often done by calculating the mismatch factor, which also constitutes an emittance growth. Clearly, it matters not whether growth comes from mismatch or from aberrations so these two are simply added in quadrature and the sum is minimized. This is the whole 'trick'.

But it remains to find $\Delta x'$. This turns out to be surprisingly simple: for any of the standard beamline elements, simple formulae can be found.

2 Transport elements

Solenoid third-order aberration depends on aperture, or more specifically, fringe-field 'hardness'. Quadrupole third-order aberrations do not depend on fringe field, but fifth-order aberrations do. Dipole bender second-order aberrations do not depend on fringe-field hardness, but third-order aberrations do.

2.1 Solenoid

A solenoid's focal strength scales as the square of the ratio of magnetic field to particle momentum, while in quadrupoles the scaling is linear, so solenoids are much weaker focusing elements than quadrupoles. For this reason, they are most commonly used for low-energy electron transport. They have the great convenience that they focus in both transverse directions.

It is useful to know that the magnetic field from a solenoid is completely given by the following expansions:

$$B_z(r,z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!^2} \left(\frac{r}{2}\right)^{2n} \left(\frac{\mathrm{d}}{\mathrm{d}z}\right)^{2n} B_z(0,z)\,, \tag{2}$$

$$B_r(r,z) = \sum_{n=0}^{\infty} \frac{(-1)^{n+1}}{(n+1)!n!} \left(\frac{r}{2}\right)^{2n+1} \left(\frac{\mathrm{d}}{\mathrm{d}z}\right)^{2n+1} B_z(0,z)\,. \tag{3}$$

These follow directly from Maxwell's equations plus the symmetry and establish that the fields can be completely derived in all of space, knowing only the on-axis field function.

For small r, we have

$$B_z = B_0 - \frac{r^2}{4} B''\,, \tag{4}$$

$$B_r = -\frac{r}{2} B' + \frac{r^3}{16} B'''\,, \tag{5}$$

where $B_0(z)$ is the on-axis field $B_z(0,z)$ and B' etc. are derivatives of $B_0(z)$.

To the same order,

$$B_z^2 = B_0^2 - \frac{r^2}{2} B_0 B''\,. \tag{6}$$

The equation of motion through the solenoid has the radial part:

$$r'' + K B_z^2\, r = 0\,, \tag{7}$$

where K is a constant containing the magnetic rigidity: $K = 1/(2B\rho)^2$.

Expanding to cubic force order gives

$$r'' + K\left(B^2\, r - \frac{BB''}{2}\, r^3\right) = 0\,. \tag{8}$$

The non-linear term gives an r' error, denoted $\Delta r'$,

$$\begin{aligned}\Delta r' &= \frac{K}{2}\int r^3 BB''\mathrm{d}z = \frac{K}{2} r^3 \int BB''\mathrm{d}z \text{ (thin-lens approximation)} \\ &= -\frac{K}{2} r^3 \int B'^2 \mathrm{d}z\,.\end{aligned} \tag{9}$$

Thus we see that optimizing a solenoid by minimizing the off-axis focal deviation is exactly the same as minimizing the mean-squared value of B'. Solenoids' spherical aberrations

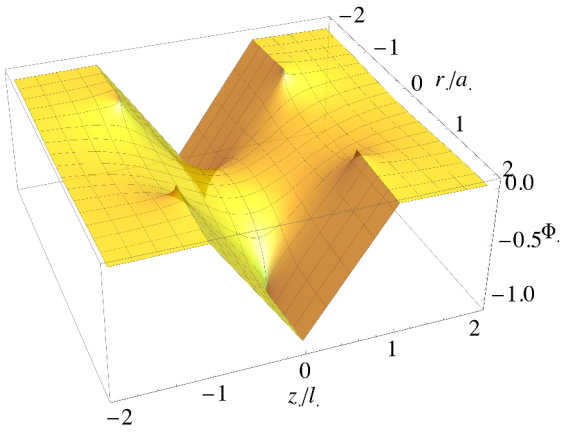


Fig. 2: Potential function plot (r, z) of three-aperture einzel lens, where apertures are at $z = -1, 0, 1$ in units of aperture radius.

diverge in the hard-edge limit. This puts to rest the notion that solenoids are improved by flattening their fields. In fact, the opposite is the case: the softer the edges, the smaller the aberrations. For a long solenoid, it is beneficial to have more turns at the centre than at the entrance and exit. Again, this is contrary to common practice. An example is the Rutherford front end test stand [7]. If they had not flattened their solenoid fields, and had put extra turns near the centre rather than near the ends, the third-order aberration would have been roughly only half as large. See also Biswas [8] on this point.

Clearly, the integral (Eq. (9)) depends on the form of the fringe field. However, if we can characterize it as having an effective width w, then

$$\Delta r' \sim \frac{r^3}{fwL} = \frac{r^3}{4w\rho^2}, \tag{10}$$

as the focal length for solenoids is given by

$$\frac{1}{f} = \frac{\int B^2 \mathrm{d}z}{(2B\rho)^2} = \frac{L}{4\rho^2}\,.$$

Usually, if no special shaping is done and there is a uniform number of turns per unit length, $w \sim 3a$, where a is aperture radius.

Roughly speaking,

$$\frac{\Delta\epsilon}{\epsilon} = \frac{\Delta r'}{\epsilon/\hat{r}} \sim \frac{\hat{r}^4}{\epsilon f w L}\,. \tag{11}$$

2.2 Einzel lens

As with solenoids, the axial symmetry of einzel or aperture lenses means that the field in all space can be determined if the on-axis field is known. In fact, the equations for the electric field are identical to those for the magnetic field in a solenoid (Eqs. 2 and 3). Integrating the latter

equation as the radial force, in the thin-lens approximation, results in zero focal effect and zero aberrations. It is only when the particle's changing speed is taken into account that there is a net effect.

An einzel lens can be constructed from as few as three apertures; a potential plot in $\Phi(r, z)$ is shown in Fig. 2. Near the axis, the potential is parabolic, but near the aperture edge, it is clearly not. This can be investigated approximately by considering one aperture at a time. The grounded conducting circular aperture separating two regions that have asymptotically constant electric fields on either side is a textbook boundary-value problem, for example by Jackson [9]. From this we find that in the plane of the aperture, the potential is actually elliptical, given by

$$\Phi(r, 0) = \frac{E_1}{\pi}\sqrt{a^2 - r^2}\,, \tag{12}$$

where E_1 is the electric field on the non-ground side and a is the aperture radius. Thus, the radial electric field diverges at the aperture edge; this, in itself, suggests that the beam should not be allowed to fill the aperture. A quadrupole's focusing force, for example, is not singular in this way.

One can find the focal length of the single aperture by using Eq. 5 as $E_r = -rE'/2$ and integrating in the thin-lens limit:

$$\Delta r' = \int r'' \mathrm{d}z = \int \frac{1}{v^2}\frac{\mathrm{d}^2 r}{\mathrm{d}t^2}\mathrm{d}z = \int \frac{q}{mv^2}E_r \mathrm{d}z \approx -\frac{r}{4V_0}E_1, \tag{13}$$

where $qV_0 = mv^2/2$ and V_0 is the voltage through which the beam particle of charge q has been accelerated. This a focal length f:

$$\frac{1}{f} = \frac{E_1}{4V_0}. \tag{14}$$

Clearly, this only applies in the non-relativistic regime: einzel lenses are not used at relativistic energies.

Two such apertures, separated by a distance l, the first kept at the source potential V_0 and the second at $V_1 + V_0$, create the field $E_1 = V_1/l$. This constitutes an accelerator column and the focal power is known:

$$\left.\frac{1}{f}\right|_{\text{acc.column}} = \frac{3\kappa^2(2+\kappa)}{8l(1+\kappa)^2}\,, \tag{15}$$

where

$$\kappa = \sqrt{\frac{V_1 + V_0}{V_0}} - 1$$

is the change in momentum relative to initial momentum. This can be derived from three transfer matrices: entrance aperture, with f given by Eq. 14 ($f = 4V_0 l/V_1$); body of column where $E'_z = 0$; and exit aperture, with $f = -4V_1 l/V_0$ [10].

We create an einzel lens by placing two acceleration columns back-to-back, bringing beam energy back to source energy. The focal length of this is found to be

$$\left.\frac{1}{f}\right|_{\text{einzel}} = \frac{3\kappa^2(4-\kappa^2)}{8l(1+\kappa)}. \tag{16}$$

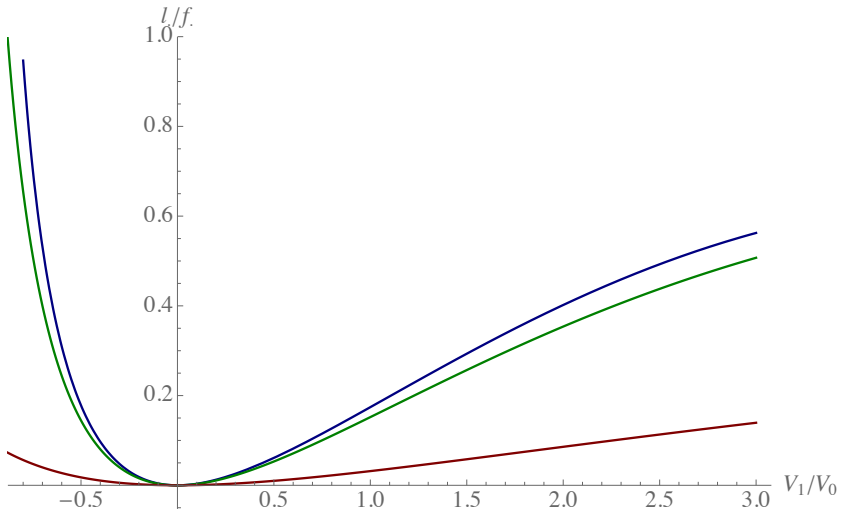


Fig. 3: Focal power in units of $1/l$ for einzel lenses. From top down: blue, infinitesimal aperture; green, $a = l/10$; red, $a = l$, as drawn in Fig. 2.

The plot is the blue curve shown in Fig. 3.

Unfortunately, this expression is only accurate for cases where $l \gg a$, while in common practice, $l/a \sim 4$ or less. The reason is that the superposition of apertures is not valid for small aspect ratios. This is clear from Fig. 2. The three apertures are not actually equipotential surfaces, and the on-axis potential does not reach $V_1 + V_0$ at the lens centre.

The green and red curves of Fig. 3 were calculated using the `CEA` element of `COSY-`∞ [6]. This code is an invaluable tool for calculating higher-order effects in general, and in particular for realistic einzel lenses, where $a \sim l$. `COSY-`∞ uses the single-aperture on-axis potential as given by Jackson [9],

$$\Phi(0, z) = \frac{E_1 a}{\pi}\left[1 - \frac{z}{a}\cot^{-1}\left(\frac{z}{a}\right)\right], \tag{17}$$

and adds apertures together to form the lens.

In spite of there being no closed-form expression for focal length f, we can still write the scaling for higher-order effects, in terms of f. For comparison with other lenses, we write in terms of the total length $L = 2l$. From `COSY`, we find

$$\Delta r' \approx \frac{r^3}{\eta f a L}, \tag{18}$$

where η varies according to aspect ratio: for $L \gg a$, $\eta = 3/2$, but for example for $L = a$, $\eta = 1/2$. Unsurprisingly, the aberrations scale similarly to solenoids as $(faL)^{-1}$, but for einzel lenses the proportionality factor is two to six times worse.

2.3 Quadrupole

Like solenoids, the lowest-order aberration is cubic or third-order. But unlike solenoids, this aberration cannot be reduced by shaping the field strength function, i.e. softening the edges. It turns out that the third-order aberration is independent of fringe-field shape or extent. (This is not, however, true of fifth and higher order.)

The electrostatic quadrupole potential field

$$V(x,y) = \frac{k}{2}(x^2 - y^2) \tag{19}$$

is a solution to Laplace's equation, but only if the quadrupole is infinitely long ($k =$ constant). For finite quadrupoles, we use the expansion

$$V(x,y,z) = \frac{k}{2}(x^2 - y^2) - \frac{k''}{24}(x^4 - y^4) + \frac{k''''}{720}(x^6 - y^6) - \cdots . \tag{20}$$

The quartic term gives a cubic force term, which leads to the following focusing error,

$$\Delta x' = \frac{-1}{f^2 L_\mathrm{Q}} \left(\frac{7}{6}x^3 - \frac{1}{2}xy^2 \right), \tag{21}$$

where L_Q is the quad length and f the focal length. It is important to note that this is independent of aperture size or fringe-field hardness: indeed, the aberration is not affected by changing the fringe-field shape.

For the Hamiltonian technique used to derive these, see Ref. [11]. I further refine electrostatic for the relativistic case [12]:

$$\Delta x' = \frac{-1}{f^2 L_\mathrm{Q}} \left(\frac{7 - 3\beta^2}{6}x^3 - \frac{1 - \beta^2}{2}xy^2 \right). \tag{22}$$

The formula for magnetic quadrupoles is similar:

$$\Delta x' = \frac{-1}{f^2 L_\mathrm{Q}} \left(\frac{1}{3}x^3 + xy^2 \right), \tag{23}$$

So electrostatic quadrupoles are not really worse than magnetic quadrupoles; they might actually be better, e.g. for ribbon beams.

Roughly speaking,

$$\frac{\Delta\epsilon}{\epsilon} = \frac{\Delta x'}{\epsilon/\hat{x}} \sim \frac{\hat{x}^4}{\epsilon f^2 L_\mathrm{Q}} . \tag{24}$$

This formula has an obvious consequence. In any transport system where optics strengths and beam sizes are given, the only way to reduce emittance growth due to aberrations is to lengthen the 'worst-offending' quadrupoles.

In a periodic transport channel with β-function β_T, $x^2 = \epsilon\beta_\mathrm{T} \sim \epsilon L_\mathrm{c}$ and $f \sim L_\mathrm{c}$, so somewhat surprisingly the optics length scale cancels and

$$\frac{\Delta\epsilon}{\epsilon} \sim \frac{\epsilon}{L_\mathrm{Q}} . \tag{25}$$

Even if the emittance is as large as 100 μm, this gives the easily-satisfied constraint $L_{\rm Q} \gg 0.1\,$mm. (Commonly, 100 μm is quoted as 100π mm-mrad. I use the convention that emittance is phase-space area divided by π [13].)

However, this is for a regular FODO transport case. Matching sections often have much tighter constraints. An example is when matching to an RF quadrupole. Since the RF quadrupole requires a very small beam, the matching system must transform a relatively large beam with small divergence to a small beam with large divergence. The 'worst-offending' quadrupole in these situations is always the next-to-last quadrupole where the beam is much larger in the defocusing plane of the last quadrupole than in the focusing plane. This forces the last two quadrupoles close together and requires a very short last quadrupole, but a long next-to-last quadrupole.

2.3.1 Quadrupole match example

We apply this to the example shown in Fig. 4. The RF quadrupole requires matched $\beta_{\rm T} = 6.125\,$cm and the beam comes from a FODO section, where $\beta_{\rm T} = 84.6\,$cm. With final quadrupoles of effective length $L = 8.4\,$cm, we get $\hat{x} = 1.8\,$cm, $f = 9.7\,$cm, with desired acceptance $\epsilon = 0.005\,$cm. [1] We get

$$\frac{\Delta\epsilon}{\epsilon} \sim \frac{1.8^4}{0.005 \times 9.7^2 \times 8.4} = 2.6\,. \tag{26}$$

To improve the emittance growth, we shorten the final quadrupole in order to decrease the size of the beam in the next-to-final quadrupole. In this case (Fig. 5), the beam size is only one-half as large, so the result is $1/16$ or

$$\frac{\Delta\epsilon}{\epsilon} \sim 0.17. \tag{27}$$

These calculations were verified with COSY-∞ [6] and gave $\Delta\epsilon/\epsilon \approx 0.1$. Final phase spaces are shown in Fig. 6.

It is important to understand that codes like TRANSPORT and TRACE3D can result in aberrations as large as that of the blue curve in Fig. 6, without any hint that there is a problem.

2.4 Dipole bender

Dipole magnets or electrostatic benders are similar to quadrupoles in that their lowest-order aberrations are insensitive to field falloff shaping and extent. The difference is that the lowest order is quadratic rather than cubic force. See my Snowmass talk [14] for derivations.

For the sector magnet the non-linear kicks are as follows ($L = \rho\theta$, the trajectory length):

$$\Delta x' = -\frac{L}{2\rho^3}y^2 \text{ and } \Delta y' = -\frac{L}{\rho^3}xy\,. \tag{28}$$

Often for dipoles the length L is not small compared with the focal length, so the thin-lens approximation used in estimating the aberration is not very good. In that case the calculation can simply split the dipole once or twice.

[1] Commonly: '50π mm-mrad'.

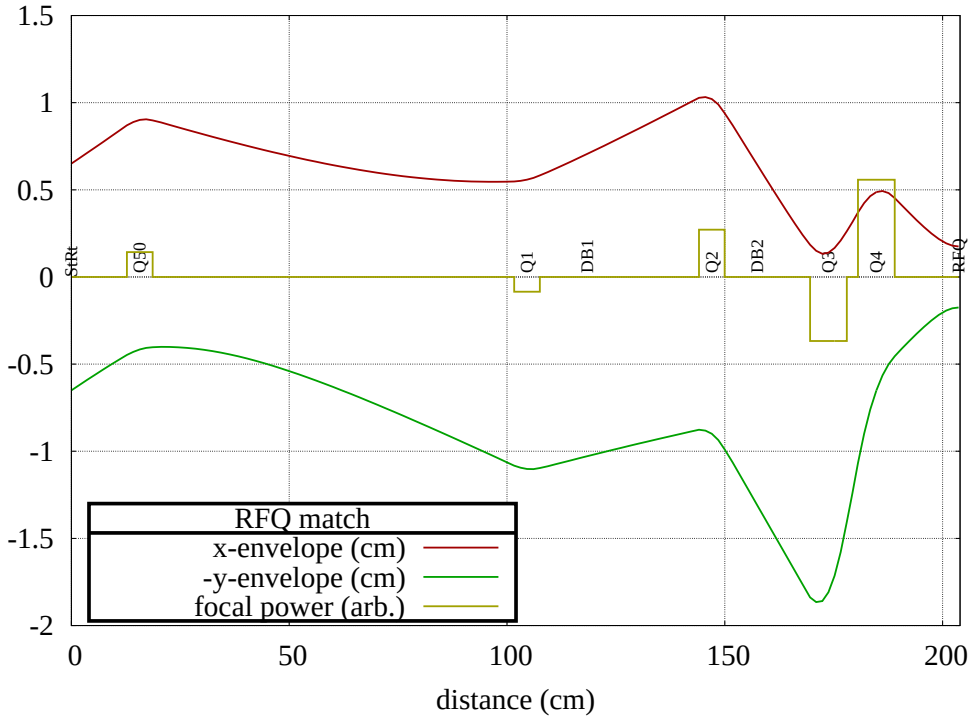


Fig. 4: First-order beam envelopes for match to RF quadrupole. Emittance is 50 μm. Red (top) is x-envelope, green (bottom) is y-envelope, plotted for clarity as if it is negative. Yellow (middle) is quadrupole strength function. This is for exact first-order match, ignoring the third-order aberrations. Emittance growth is very bad, as can be seen in the blue curve of Fig. 6.

For the electrostatic dipole bender, I have solved the general toroidal case. Here, c is the ratio of electrode curvature in the bend plane to the non-bend plane. This means that $c = 0$ is a cylindrical bend and $c = 1$ is a spherical one:

$$\Delta x' = \frac{L}{\rho^3}\left[\left(-4 + \frac{7}{2}c - c^2\right)x^2 + \left(-\frac{1}{2}c + c^2\right)y^2\right], \tag{29}$$

$$\Delta y' = \frac{L}{\rho^3}\left(-c + 2c^2\right)xy\,. \tag{30}$$

3 Beam transport optimization technique

Emittance growth factors are found for each transport element and combined with the emittance growth due to mismatch, to form a function to be minimized. We simply add (in quadrature) together, and minimize this as we vary such parameters as quadrupole position, length, or strength.

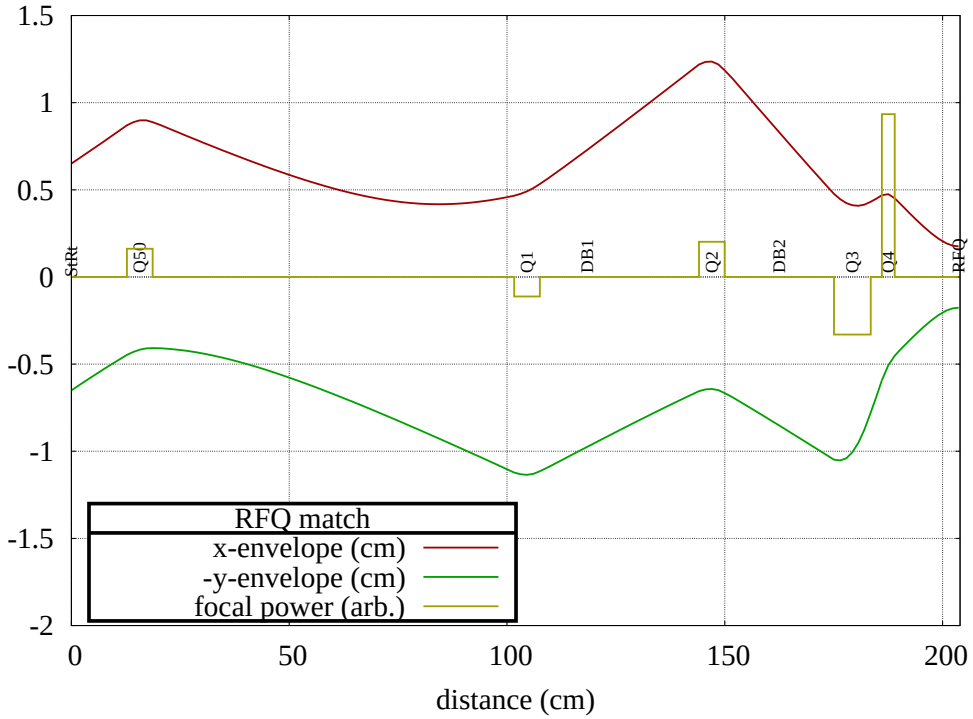


Fig. 5: Optimized RF quadrupole match, taking into account third-order aberrations, according to Eq. 21

3.1 Growth from higher order

Lowest higher-order kicks are used to estimate emittance growth. See Table 1, which is not meant to be exhaustive.

3.2 Growth from mismatch

The mismatch factor is the standard definition from, e.g., Bovet *et al.* [15], see Fig. 7. The ratio of area of the ellipse that is similar but encapsulates the mismatched ellipse to the area of the mismatched ellipse is $D + \sqrt{D^2 - 1}$, where $D \equiv (\beta_2\gamma_1 + \beta_1\gamma_2)/2 - \alpha_1\alpha_2$ and α, β, γ are the 'Twiss' parameters of the mismatched ellipse (subscript 1) and the matched ellipse (subscript 2). As $D \ll 1$, it is sufficient to summarize this as $\Delta\epsilon/\epsilon \approx \sqrt{2(D-1)}$.

This does not pretend to be a higher-order calculation; indeed, it is possible (though highly unlikely) that the higher order of one element is compensated by another; this would only show up in an actual higher-order calculation. As stated, the intention is not to find accurate higher-order effects, only to ensure that they are negligible.

3.3 Optimization engine

There are many optimization techniques for designing and tuning beam transport lines. Some are built into the transport codes themselves. Almost all of these work on the basis of reducing

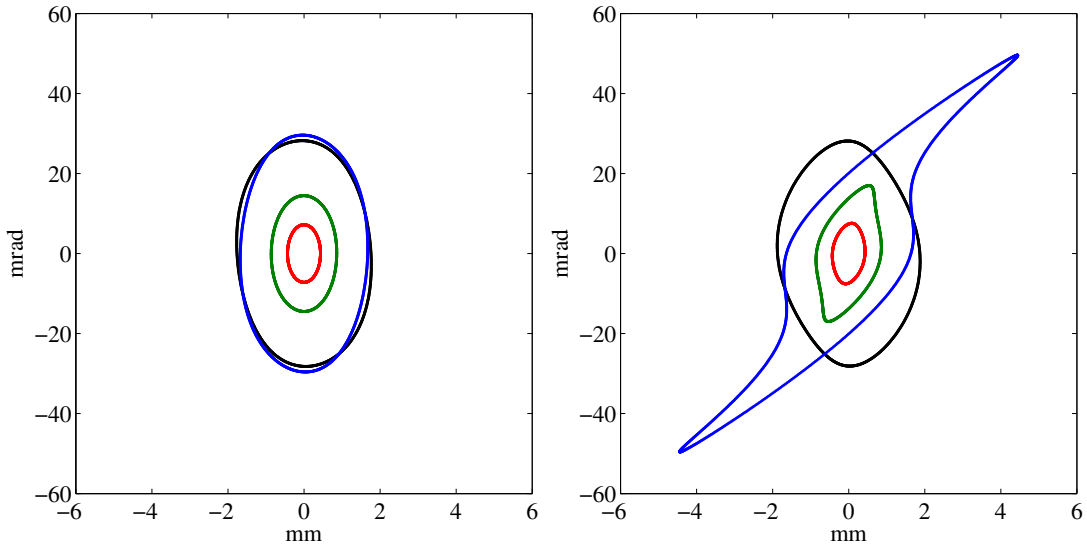
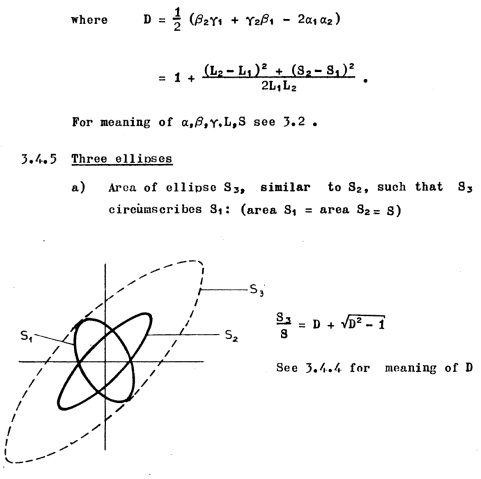


Fig. 6: Emittance contours (i.e. constant action) in phase space for previous envelope figures, calculated from the COSY-∞ transfer maps. Left, x-P_x; right,s y-P_y. Black is the optimized case for the design acceptance $\epsilon = 50$ μm, where the long next-to-last quadrupole is brought forward by shrinking the final quadrupole, as shown in Fig. 5. From outside in, blue, green, and red are for the non-optimum configuration: blue, matched beam $\epsilon = 50$ μm; green, $\epsilon = 12.5$ μm; red, $\epsilon = 3$ μm.

Table 1: Higher-order kicks for a few chosen common elements

Element	$\Delta x'$	$\Delta y'$
Solenoid	$-\frac{x(x^2+y^2)}{fwL}$	$-\frac{y(x^2+y^2)}{fwL}$
$w =$ fringe-field length, $f =$ focal length, $L =$ eff. length		
Einzel lens	$-\frac{x(x^2+y^2)}{\eta faL}$	$-\frac{y(x^2+y^2)}{\eta faL}$
$a =$ aperture radius, $f =$ focal length, $L =$ total length, $1/2 < \eta < 3/2$		
Magnetic quadrupole	$\frac{-1}{f^2L}\left(\frac{1}{3}x^3 + xy^2\right)$	$\frac{-1}{f^2L}\left(\frac{1}{3}y^3 + yx^2\right)$
$f =$ focal length, $L =$ eff. length		
Electric quadrupole	$\frac{-1}{f^2L}\left(\frac{7}{6}x^3 - \frac{1}{2}xy^2\right)$	$\frac{-1}{f^2L}\left(\frac{7}{6}y^3 - \frac{1}{2}yx^2\right)$
Magnetic sector bend	$-\frac{\theta y^2}{2\rho^2}$	$-\frac{\theta xy}{\rho^2}$
$\theta =$ bend angle, $\rho =$ bend radius		
Electric cylindrical bend	$-\frac{4\theta x^2}{\rho^2}$	0
Electric spherical bend	$\frac{\theta(-3x^2+y^2)}{2\rho^2}$	$\frac{\theta xy}{\rho^2}$



where $D = \frac{1}{2}(\beta_2\gamma_1 + \gamma_2\beta_1 - 2\alpha_1\alpha_2)$

$= 1 + \frac{(L_2 - L_1)^2 + (S_2 - S_1)^2}{2L_1L_2}$.

For meaning of $\alpha, \beta, \gamma, L, S$ see 3.2 .

3.4.5 Three ellipses

a) Area of ellipse S_3, similar to S_2, such that S_3 circumscribes S_1: (area S_1 = area S_2 = S)

$\frac{S_3}{S} = D + \sqrt{D^2 - 1}$

See 3.4.4 for meaning of D

Fig. 7: Excerpt from Bovet *et al.* [15]

an error to zero by finding local derivatives of the error with respect to the parameters. I use a downhill simplex method. It is fast and robust.

It is also easily modified to incorporate simulated annealing for more than three free parameters. I use routines from the book *Numerical Recipes* by Press *et al*. [16]. They give the following description.

> Offered a succession of options, a simulated thermodynamic system was assumed to change its configuration from energy E_1 to energy E_2 with probability $p = \exp[(E_2 - E_1)/(kT)]$. Notice that if $E_2 < E_1$, this probability is greater than unity; in such cases the change is arbitrarily assigned a probability $p = 1$, i.e., the system always took such an option. This general scheme, of always taking a downhill step while sometimes taking an uphill step, has come to be known as the Metropolis algorithm. To make use of the Metropolis algorithm for other than thermodynamic systems, one must provide the following elements:
>
> 1. A description of possible system configurations.
> 2. A generator of random changes in the configuration; these changes are the options presented to the system.
> 3. An objective function E (analogue of energy) whose minimization is the goal of the procedure.
> 4. A control parameter T (analogue of temperature) and an annealing schedule which tells how it is lowered from high to low values, e.g., after how many random changes in configuration is each downward step in T taken, and how large is that step.

Applied to the design of a transport channel, we can identify these four points as:

1. a set of parameters, such as quadrupole strengths, locations, and their allowed ranges;
2. use the vertices of a 'simplex' in the space of all possible parameters, and change these randomly, independently according to T; for $T = 0$, the technique reduces to the standard 'simplex' method of minimization;
3. E = the sum of all possible effects to be minimized, such as mismatch or emittance growth, with their appropriate weights;
4. use T, such that $T = 1$ implies that parameters are varied through their entire ranges, specify N, the number of random changes at each temperature, and $\alpha(< 1)$, the factor with which T is multiplied at each step.

For as few as three parameters, α can be 0.88, with $N = 8$, resulting in only about 100 evaluations of the beamline to achieve 10^{-4} accuracy. But this number of evaluations increases exponentially with the number of parameters. For six parameters, we need ten times slower cooling $\alpha = 0.98$, $N = 50$, and this results in roughly 100 times more evaluations ($10^4 - 10^5$) before convergence is obtained. An empirical relation found from many such runs is that for n variable parameters, we need $N = 2^n$, $\alpha = 1 - \frac{1}{N}$. Still, the calculations are simple so that millions of beamline calculations can be made and an optimum found, for the same amount of computing time as a single million-particle simulation.

3.4 TRANSOPTR: Space charge and other complications

TRANSOPTR [17] was originally written in 1981. Like the original envelope codes TRANSPORT and TRACE3D, it included only elements whose transfer matrices could be written in closed form and were known. This required all elements to be 'hard-edged'. The extension to other elements and to space charge requires that the envelope code itself numerically integrates through the elements. The exact differential equations of Sacherer [18] were incorporated into TRANSOPTR by Mark deJong [19], but greatly expanded by me to include many other element types [20]. In principle, any linear element can be incorporated, from solenoids and quadrupoles with bell-shaped strength functions, to linear accelerators and bunchers. All that is needed is the Hamiltonian to quadratic order [21].

4 Conclusions

Simple formulae can be found for the lowest-order aberrations and used to estimate emittance growth. A simple strategy for beam transport system design optimization is to add these growths to the growth from mismatch, to form an error function to be minimized. Efficient first-order transport codes (with or without space charge) can be augmented in this way. An example is TRANSOPTR, which uses a minimization engine to vary beamline parameters to find an optimum with negligible emittance growth.

At TRIUMF, many beamlines have been efficiently designed in this way, some with strong space charge effects, and perform as predicted, in spite of never having been simulated with any multiparticle code [22–24].

Acknowledgement

I would like to thank Dr Thomas Planche for his careful reading and helpful comments.

References

[1] S. Penner, *Rev. Sci. Instrum.* **32** (1961) 150, https://doi.org/10.1063/1.1717300.

[2] K.L. Brown, A first-and second-order matrix theory for the design of beam transport systems and charged particle spectrometers, Tech. Rep. SLAC-75, Stanford Linear Accelerator Center, CA, 1971.

[3] K.L. Brown *et al.*, Transport: a computer program for designing charged-particle beam-transport systems, Tech. Rep. SLAC-R-91-REV-3, Stanford Linear Accelerator Center, CA, 1980.

[4] K.R. Crandall and D. Rusthoi, *Trace 3-D Documentation* (Los Alamos National Laboratory, Los Alamos, 1987), https://doi.org/10.2172/6290515.

[5] H. Wollnik *et al.*, `GIOS`—a program for the design of general ion optical systems, ii, Physikalisches Institut, Justus-Liebig-Universität, Giessen, 1988.

[6] M. Berz, *Nucl. Instrum. Methods Phys. Res. A* **298** (1990) 473, https://doi.org/10.1016/0168-9002(90)90649-Q.

[7] J. Back *et al.*, Commissioning of the low energy beam transport, MOPEC078, Proc. IPAC, 2010 vol. 10.

[8] B. Biswas, *Pramana*, **86** (2016) 1299, https://doi.org/10.1007/s12043-016-1197-7.

[9] J.D. Jackson, *Classical Electrodynamics* (John Wiley & Sons, New York, 2007).

[10] A. Galejs and P. Rose, in *Focusing of Charged Particles*, Ed. A. Septier (Academic Press, Orlando, 1967), vol. ii, pp. 297–326, https://doi.org/10.1016/B978-0-12-636902-1.50014-7.

[11] R. Baartman, Intrinsic third order aberrations in electrostatic and magnetic quadrupoles, Proc. Particle Accelerator Conf., 1997, (IEEE, 1997), vol. 2, p. 1415.

[12] R. Baartman, Quadrupoles, to third order, *arXiv:1508.01576*, 2015.

[13] R. Baartman, Emittance convention, Tech. Rep. TRI-BN-15-07, TRIUMF, 2015.

[14] R. Baartman, End effects of beam transport elements, Snowmass, 2001. http://lin12.triumf.ca/text/Talks/2001Snowmass/FF.pdf

[15] C. Bovet *et al.*, A selection of formulae and data useful for the design of Ag synchrotrons, Tech. Rep. (CERN, Geneva, 1970).

[16] W.H. Press *et al.*, *Numerical Recipes* (Cambridge University Press, Cambridge, 1996).

[17] E. Heighway and R. Hutcheon, *Nucl. Instrum. Methods Phys. Res.* **187** (1981) 89, https://doi.org/10.1016/0029-554X(81)90474-2.

[18] F.J. Sacherer, *IEEE Trans. Nucl. Sci.* **18** (1971) 1105, https://doi.org/10.1109/TNS.1971.4326293.

[19] M. deJong and E. Heighway, *IEEE Trans. Nucl. Sci.* **30** (1983) 2666, https://doi.org/10.1109/TNS.1983.4332918.

[20] R. Baartman, TRANSOPTR: changes since 1984, Tech. Rep. TRI-BN-16-06, TRIUMF, 2016. http://lin12.triumf.ca/text/design_notes/b2016_06/TRI-BN-16-06_TRANSOPTR.pdf

[21] R. Baartman, Fast envelope tracking for space charge dominated injectors, 28th Linear Accelerator Conf. (LINAC'16), East Lansing, MI, 25–30 September 2016, pp. 1017–1021 (JACOW, Geneva, 2017).

[22] T. Planche *et al.*, Commissioning and early operation of the ARIEL e-Linac, 28th Linear Accelerator Conf. (LINAC'16), East Lansing, MI, 25–30 September 2016, pp. 12–16 (JACOW, Geneva, 2017).

[23] R. Baartman, ISAC LEBT, ISAC and ARIEL: The TRIUMF Radioactive Beam Facilities and the Scientific Program, pp. 69–77 (Springer, 2013), https://doi.org/10.1007/978-94-007-7963-1_7.

[24] R. Baartman, TRIUMF 300 keV vertical injection line, European Cyclotron Progress Meeting, 2015. http://lin12.triumf.ca/text/Talks/2015ECPM/ISIS.pdf

Proceedings of the CAS–CERN Accelerator School: Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018)

Transfer Lines

W. Bartmann
CERN, Geneva, Switzerland

Abstract
Transfer lines are used to connect accelerators or deliver beam to fixed target experiments. In the following sections the design steps of a transfer line are introduced, starting from back-of-the-envelope calculations. The second part focuses on matching tools for optics design and related hardware specifications. The third part collects special topics which can appear in beam lines.

Keywords
CAS; transfer lines; optics matching; error and correction studies.

1 What is a Transfer Line?

The difference between rings and transfer lines was introduced in [1]. Let us recall that in a ring the condition of periodicity defines a unique solution for the Twiss parameters at any location along the accelerator. In a transfer line instead, without any condition of periodicity, the Twiss parameters are defined by the initial conditions and are propagated according to the transfer matrices of each element along the line. While the change of a single element in a ring, e.g. the tuning the strength of a quadrupole, will affect the optics all around the ring, the change will only be visible downstream of this element in a transfer line. Also, since beam passes only once a transfer line, hardware specifications with regard to field homogeneity, power converter stability and alignment are usually relaxed with respect to rings. On the other hand, it becomes more difficult and time consuming to measure beam parameters and align elements with beam. These uncertainties of the beam optics have to be taken into account in conservative margins for beam acceptances as detailed in the following sections.

2 Paper Studies

This section covers the initial design steps of a transfer line which are based on analytic calculations. They can be performed on paper and, most importantly, serve as an efficient crosscheck of studies performed with optics and tracking codes.

2.1 Geometry - estimate of bend angles

In order to calculate the geometry of a transfer line, coordinates and angles of the points to be linked have to be defined in a common coordinate system. While this is obvious for short lines of several meters, this has to be carefully performed for large scale accelerator systems, e.g linking CNGS to Gran Sasso in Italy where the CERN reference frame had to be connected to the global systems of Switzerland and Italy [2]. As soon as the geometry is defined, the required bending angles can be calculated from the magnetic rigidity

$$B\rho = \frac{p}{q} \qquad \rightarrow \qquad B\rho\,[Tm] = 3.3356\,\frac{A}{n}\,p\,[GeV/c] \tag{1}$$

where A is the atomic mass number, n the charge state and p the average momentum per nucleon. In case of electrostatic elements, the rigidity is written as

$$E\rho = \frac{pv}{q} \qquad \rightarrow \qquad E\rho\,[kV] = \frac{\gamma+1}{\gamma}\,\frac{A}{n}\,T\,[keV] \tag{2}$$

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-005.507

where T is the average kinetic energy per nucleon. The respective deflection angles are then:

$$\theta = \frac{Bdl}{B\rho} \quad or \quad \frac{Edl}{E\rho} \tag{3}$$

While the magnetic rigidity is proportional to the particle momentum, the electric rigidity is proportional to the kinetic energy. This behaviour, together with the technological limits of magnetic and electric bends, defines the energy reach of an accelerator. Electric devices are limited by the applied voltage. For reasonable accelerator apertures, several tens of kV can be assumed as limit. Magnets on the other hand are limited by the field quality at low fields. Due to the strong dependence on material properties, remnant fields become important and measuring the field becomes a challenge. A practical example where electric devices can be used to bend and focus the beam are the 100 keV antiproton beam lines for the antimatter factory at CERN. Here, electrostatic quadrupoles with a 60 mm diameter require applied voltages of below 10 kV, electrostatic bends of up to 30 kV. What are the criteria to choose between magnetic and electric elements if you are in this energy gray zone? Electric devices are cheap in production, have cheap power supplies and cables, they are mass-independent, have no hysteresis effects hence easy operation, no power consumption hence no cooling and the transverse field shape is easy to optimize. On the other hand, the field shape and effective length of electric devices is difficult to measure. Also the diagnosis of faulty connections can pose problems. In case these devices are inside vacuum, they present a large outgassing surface area, are vulnerable to dirt inside vacuum and require interlocking for sparking and safety. Any repair requires opening the vacuum and insulators have to be vacuum and bake-out compatible. Many of these arguments related to vacuum apply also to magnets.

Estimating the required bending angles can be straightforward as in the above mentioned example of 100 keV antiproton beam lines where the beam deflection in the is handled by single elements over a length of 0.5 m. Highly rigid beams require often long transfer lines due to the limit of maximum reachable magnetic fields. Here the bending is distributed over the full line and in first paper estimates one has to account also for focusing, diagnostic and vacuum elements. A reasonable dipole filling factor is around 70%.

So far we discussed technological limits of the bending fields. A few 10 kV can be applied on electric devices due to sparking. For magnets one can assume 2 T for normal conducting, iron dominated magnets and around 9 T for superconducting ones. But there are limits also coming from the beam itself. When H^- ions are transferred through a line with magnetic fields, the moving ions will see the magnetic field in its rest frame. The Lorentz transform of this magnetic field gives an electric field which can be written:

$$E\left[\frac{MV}{cm}\right] = 3.197 \cdot p\left[\frac{GeV}{c}\right] \cdot B\,[T] \tag{4}$$

The electric field can detach the extra electron with 0.755 eV binding enery which leads to a definition of the ion's lifetime:

$$\tau = \frac{A}{E}\exp^{\frac{C}{E}} \qquad A = 7.96 \cdot 10^{-14}\ s\ MV/cm,\ C = 42.56\ MV/cm \tag{5}$$

For the example of a 4 GeV H^- injection a reasonable fractional loss of less than 10^{-4} limits the magnetic field to 0.13 T, see Fig. 1. In case of the Fermilab Project-X study, it was considered to inject a 8 GeV H^- beam into the recycler [4]. While usually power loss limits in lines are assumed to about 1 W/m, the corresponding activation was found to be not acceptable for 8 GeV ions and therefore the power loss maximum reduced to 0.05 W/m. In this low loss regime, black body radiation appeared to give a relevant contribution. In Fig. 2, the H^- photo-detachment cross section is shown together with the black body spectrum at different energies. Figure 3 shows the power loss due to black body radiation as a function of the beam pipe temperature. Another effect relevant for very low loss levels is rest gas stripping for which the power loss per length l can be calculated as

$$P = E\,I\,\frac{d\sigma}{d\Omega}\,\rho\,l \tag{6}$$

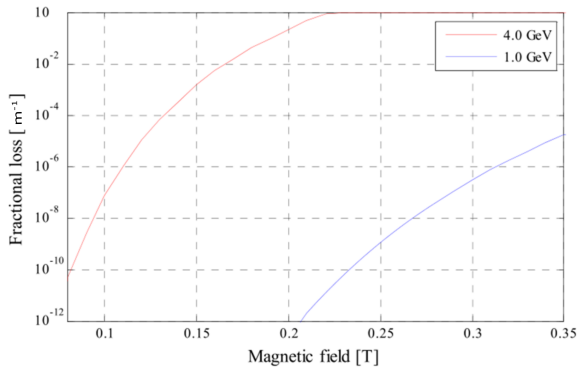


Fig. 1: Fractional loss of 1 and 4 GeV H^- as a function of the magnetic field

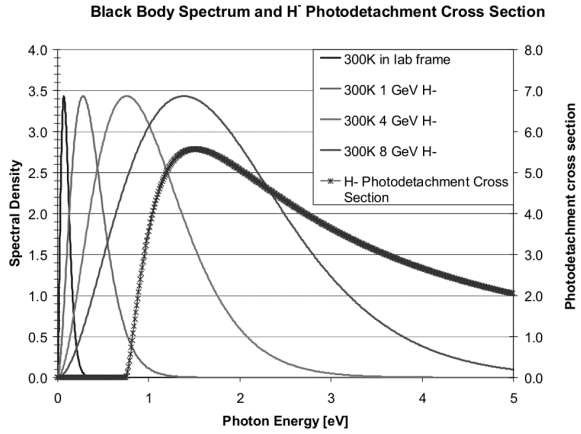


Fig. 2: Black body spectrum and H^- photo-detachment cross section as a function of the photon energy [4]

where E and I denote the beam energy and density, respectively. ρ is the gas density which is a function of temperature and pressure. The gas dependent stripping cross sections show an electron loss behaviour proportional to $1/\beta^2$:

$$\begin{aligned} \frac{d\sigma}{d\Omega} &= \frac{7 \cdot 10^{-19}}{\beta^2} cm^2 \qquad \text{per atom of nitrogen or oxygen} \\ \frac{d\sigma}{d\Omega} &= \frac{1 \cdot 10^{-19}}{\beta^2} cm^2 \qquad \text{per atom of hydrogen} \end{aligned} \tag{7}$$

All three effects, Lorentz stripping, black body radiation causing H^-photo-detachment and rest gas stripping exceeded the acceptable loss level for the 8 GeV H^- beam transport study. As a solution a cold beam screen (77 K) was assumed to reduce the photo-detachment cross section and the rest gas stripping. In order to limit Lorentz stripping, the magnetic field had to be below 0.05 T.

2.2 Focusing structure - estimate of quadrupole gradients

In case beam is extracted from a synchrotron, the focusing structure of the respective transfer line can be well defined already on paper. The cell length of a ring is usually optimized for the dipole filling at high energy. Therefore this cell length can be assumed well optimized also for the transfer line. Another assumption can be taken on the focusing structure and phase advance of each cell. Often a 90^o FODO cell is used since it has a good ratio of minimum and maximum betatron functions, features equal aperture

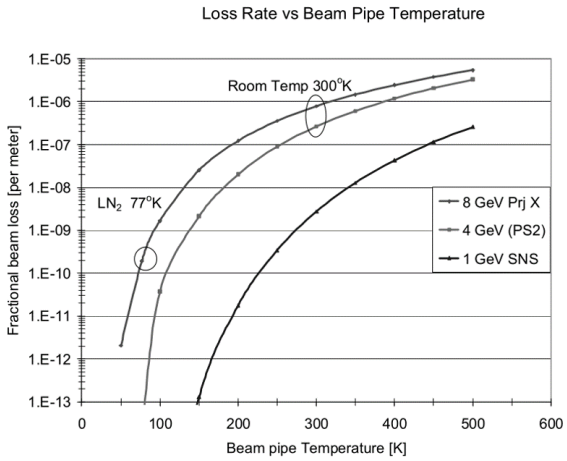


Fig. 3: Loss rate due to black body radiation for different beam screen temperatures [4]

requirements in both planes, provides good locations for trajectory correction and beam instrumentation and it provides the ideal phase advance for injection and extraction equipment. With the above taken assumptions on cell length and phase advance, the number of transfer line quadrupoles and their main parameters can be fully defined. The main quadrupole parameters are the pole tip field and radius, and

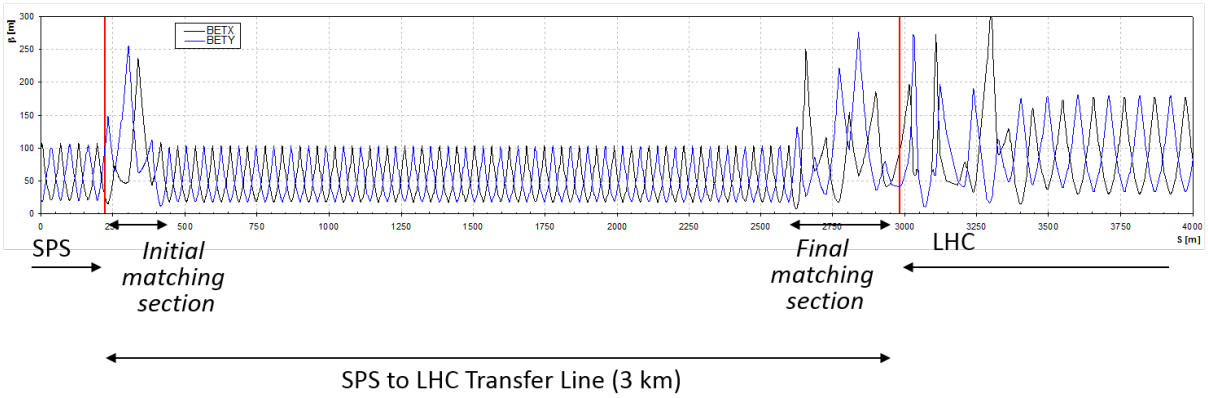


Fig. 4: Typical optics of a transfer line with a regular FODO structure and two matching sections on either side to connect to the respective ring optics'.

the gradient:

$$B = g \cdot x \qquad B_{poletip} = g \cdot a \tag{8}$$

Our assumptions on cell length and phase advance allow us to calculate the required quadrupole gradient. The transfer matrix for a 90^o FODO cell, from centre to centre of two successive defocusing quadrupoles

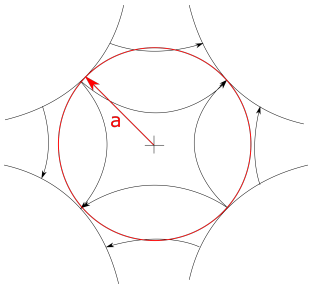


Fig. 5: Quadrupole pole tip radius

as shown in Fig. 6, can be written as:

$$R = \begin{pmatrix} 1 & 0 \\ 1/2f & 1 \end{pmatrix} \begin{pmatrix} 1 & L/2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix} \begin{pmatrix} 1 & L/2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1/2f & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 - L^2/8f^2 & L - L^2/4f \\ -L/4f^2 - L^2/16f^3 & 1 - L^2/8f^2 \end{pmatrix} \tag{9}$$

$$= \begin{pmatrix} \cos\mu + \alpha \sin\mu & \beta \sin\mu \\ -\gamma \sin\mu & \cos\mu - \alpha \sin\mu \end{pmatrix}$$

From Equ. 9 we get

$$\frac{L}{f} = 4 \sin\frac{\mu}{2} \tag{10}$$

Stability of such a FODO cell is given for

$$f > \frac{L}{4} \tag{11}$$

which is visualized in Fig. 6. For a 90^o phase advance the focal strength becomes $f = \frac{L}{2\sqrt{2}}$. From

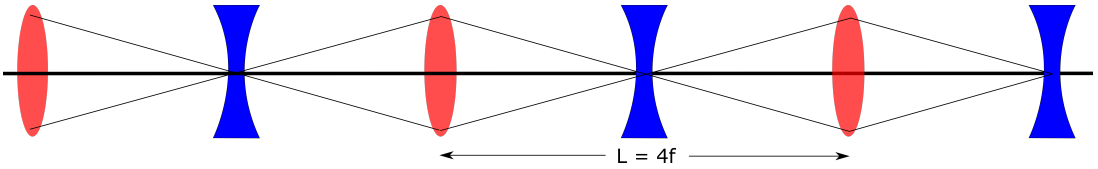


Fig. 6: FODO cell with stability condition [9]

the cell length, phase advance and the stability criterion we can define the quadrupole gradient and the maximum betatron functions:

$$\beta = \left(L + \frac{L^2}{4f}\right) / \sin\mu \tag{12}$$

In order to also estimate the pole tip field, the minimum inscribed radius inside the poles is needed. Specifying the acceptance of hardware elements in a transfer line requires to take into account certain margins in addition to the pure beam size. The following aperture formula for the radius can serve as example:

$$A_{x,y} = n_\sigma \sqrt{k_\beta\, \beta_{x,y}\, \frac{\epsilon_{x,y}}{\beta\gamma_{rel}}} + k_\beta\, D_{x,y}\, \frac{\Delta p}{p} + CO\sqrt{\frac{\beta_{x,y}}{\beta_{max}}} + alignment \tag{13}$$

The first two summands account for betatron and dispersive beam sizes, respectively. Those contributions are added linearly to stay more conservative when specifying hardware acceptances. The quadratic addition should be chosen when beam sizes are calculated for collimators and beam stoppers where the energy density on the material is critical. For the momentum spread it is advisable to use twice the rms value on the radius to include a reasonable part of this one-dimensional distribution. The third summand accounts for trajectory variations. Since their cause is often a dipole kick as $\Delta x = \Theta\sqrt{\beta_1\,\beta_2}\;sin\Delta\mu$, the constant CO is scaled with the betatron function. The last contribution stands for misalignment or material roughness and is $\mathcal{O}(mm)$. The safety factor k_β stands for beta beating or more generally uncertainty in the optics. As a typical optics uncertainty 20% can be assumed. Depending on the beam energy and consequences of beam impact, the constant n_σ is often chosen to be around 3 in low energy beam lines and around 6 in high energy beam lines. With this aperture calculation the input is complete to specify fields and pole distances of dipoles and quadrupoles.

2.3 Alignment and field errors

Another important ingredient for designing magnets are maximum tolerances on the field errors. The impact of field errors on aperture requirements is usually very low, but they are very relevant for beam quality and emittance growth. There is a rigorous derivation of field errors in [5] of which only the results are quoted here:

$$\epsilon_2 = \epsilon_1 + \frac{1}{2}\left((\Delta y)^2\,\frac{1+\alpha^2}{\beta} + \left(\Delta y'\right)^2\beta\right) \tag{14}$$

ϵ_2 denotes the resulting emittance after taking into account magnet misalignments Δy and field errors $\Delta y' = \frac{\Delta Bl}{B\rho}$. Gradient errors from quadrupoles $k = -\frac{\Delta Gl}{B\rho}$ lead to an emittance growth as:

$$\epsilon_2 = \frac{1}{2}\left(k^2\beta^2 + 2\right)\epsilon_1 \tag{15}$$

For a given emittance growth budget, these estimates allows to specify the maximum tolerable field errors. Since there are usually many error sources, one has to average of these distributed errors:

$$\left\langle y^2\right\rangle = \frac{\beta(s)}{2}\sum_n \beta_n\left\langle\delta_n^2\right\rangle \tag{16}$$

with δ_n denoting the dipole and gradient errors as:

$$\delta_n = \frac{\Delta Bl}{B_0\rho_0} \qquad and \qquad \delta_n = -lk\Delta y \tag{17}$$

which can be averaged using the rms error to:

$$\left\langle y^2\right\rangle = \frac{\beta_{aver}\;n\;\delta_{rms}}{2} \tag{18}$$

It should be noted that this approach assumes uncorrelated errors along the transfer line. While this is a reasonable assumption for misalignments, errors related to powering are often strongly correlated, as can be seen in Fig. 5 where most of the quadrupoles are connected to the same power supply. Defining tolerances for misalignments and field errors is therefore more efficient via algorithms which will be shown in the following section. In order to roughly estimate the outcome of a correction study some ballpark figures are given below:

- Number of monitors and required resolution
  - Every $1/4$ betatron wavelength
  - Grid resolution: 3 wires/σ

- Number of correctors and strength
  - Every $1/2$ betatron wavelength in H - same for V
  - Displace beam by few betatron σ per cell
- Dipole and quadrupole field errors
  - Integral main field known to better than 1 - 10^{-4}
  - Higher order field errors < 1-10^{-4} of the main field
- Dynamic errors from power converter stability
  - 1 - 10^{-5}
- Alignment tolerances
  - 0.1 - 0.5 mm
  - 0.1 - 0.5 mrad

It should be noted that these values strongly vary depending on energy, intensity and machine purpose. They can be used as good guess starting point for error and correction studies as treated in the following section.

3 Matching tools

In the previous section we estimated bending angles, quadrupole gradients, cell length, phase advance and apertures. In this section we will go into more detail by using MADX as tool for optics matching and error studies.

3.1 Optics and survey matching

In order to calculate the survey coordinates of a transfer line, we need start and end points in all three coordinates x, y, z and the three angles θ, ϕ, ψ as defined in Fig. 7. Before getting a matching routine to

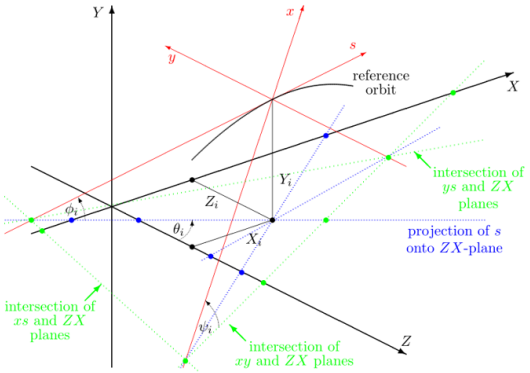


Fig. 7: MADX reference system [14]

calculate survey coordinates, a sequence of magnets needs to be established which is based on the paper estimates shown in the previous section. An example of such a line is shown in Fig. 8. This line has a central part with a FODO lattice and two matching sections on either side. There are two combined horizontal and vertical achromats. The first achromat is placed in the 90^o FODO structure where the dispersion is canceled automatically if bends with the same angle are placed with 180^o phase advance. In this example the dispersion of the second achromat is deliberately not fully canceled. Dispersion at the stripping foil is used to compensate for the Doppler spread particles with different momenta would

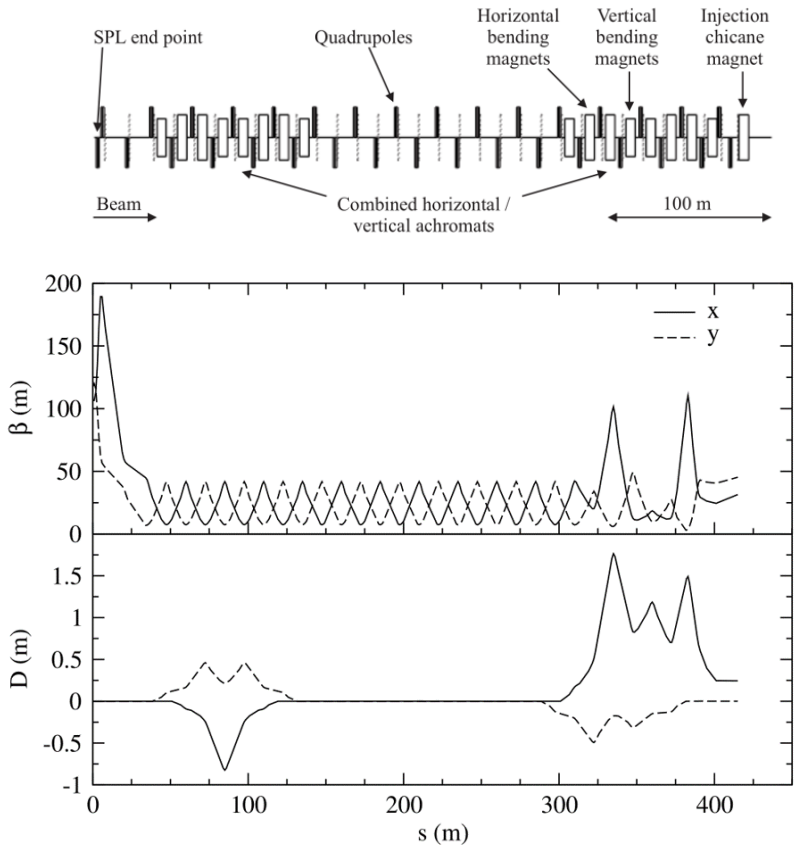


Fig. 8: Study of a transfer line between SPL and PS2 [6]

see and therefore not be excited by the laser light, explained in more detail in [1]. The related survey matching of this line is shown in Fig. A.1. While matching geometries the path length - and with it the target location - is changing. Therefore, in this example the location of the stripping foil has to be matched simultaneously. In order to match the optics, a basic FODO cell is defined as shown in the following code example. The strength initialization of the quadrupoles is according to Eq. 10.

The matching from the SPL optics to the FODO structure and from the FODO to the PS2 optics is shown in Fig. A.3. The matching itself is straightforward as long as there are enough knobs, i.e. independently powered quadrupoles available. It is advisable to start generously with the number of variables to find an initial solution, which then can be optimized by reducing the number of quadrupoles used for matching. The matching should not only target a single optics solution, but provide tunability of the line in case of optics uncertainty or changes. Also, it is not always straightforward to get the optics constraints right. It is good practice to use mechanical handover points of which the positions are clearly defined. For the optics conditions one should make sure that the single turn optics at injection or extraction with all injection or extraction equipment (kickers and bumps) is calculated to correctly take the dispersion created by these elements into account.

3.2 Final focus matching

Matching a final focus is required for fixed target experiments and often requires a high flexibility in tuning the optics parameters at the target. While a FODO structure provides good properties to transfer the beam along a line, other structures like a doublet or triplet give more optics flexibility at the focal

point. The transfer matrix of a doublet with distance d between the quadrupoles can be written as [8]:

$$\begin{pmatrix} 1 & 0 \\ -1/f_2 & 1 \end{pmatrix} \begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1/f_1 & 1 \end{pmatrix} = \begin{pmatrix} 1+d/f_1 & d \\ 1/f_1 - 1/f_2 - d/f_1 f_2 & 1-d/f_2 \end{pmatrix} \tag{19}$$

For equal quadrupole strength the matrix reduces to

$$R = \begin{pmatrix} 1+d/f & d \\ -d/f^2 & 1-d/f \end{pmatrix} \tag{20}$$

which gives the effective focal length of a doublet:

$$f^* = \frac{f^2}{d} \tag{21}$$

A doublet can be used to focus particle rays from point to parallel as shown in Fig. 9 For this condition

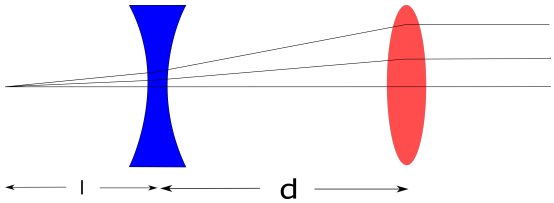


Fig. 9: Focusing particle rays point to parallel with a doublet structure

the matrix elements R_{22} and R_{44} need to be zero which gives

$$f_1\ f_2 = l\ d \qquad \frac{f_1}{f_2} = \frac{l}{l+d} \tag{22}$$

and for the magnetic fields:

$$f_1 = l\sqrt{\frac{d}{l+d}} \qquad f_2 = \sqrt{d(l+d)} \tag{23}$$

Now we can use two doublets to focus from point to point as shown in Fig. 10. By placing two mirrored

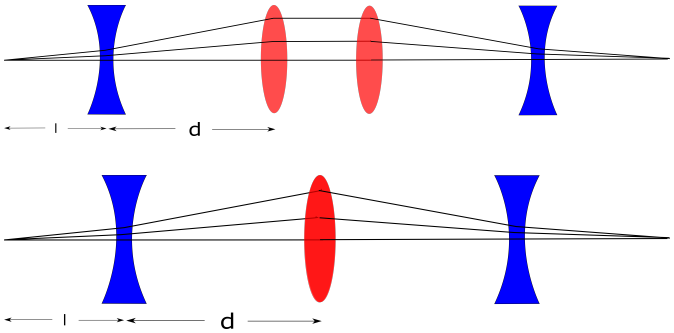


Fig. 10: Focusing particle rays point to point with a pair of doublets (top) which can be simplified to a triplet structure (bottom) [8].

doublets and combining the central quadrupoles into one we create a triplet structure. The corresponding transfer matrix is written as

$$R = \begin{pmatrix} 1 - 2\frac{d^2}{f^2} & 2d\left(1 - \frac{d}{f}\right) \\ -2\frac{d}{f^2}\left(1 + \frac{d}{f}\right) & 1 - 2\frac{d^2}{f^2} \end{pmatrix} \tag{24}$$

with the effective focal length

$$f^* = \frac{f^2}{2d}\left(1 + \frac{d}{f}\right) \tag{25}$$

With a triplet we can simultaneously control the focusing in both planes and create round spots on a target. Note that we could focus also with a single doublet point-to-point, but not symmetrically in both planes. The high optics flexibility of a triplet makes it the most common choice for final focus matching. In Fig. 11 we show the optics of a doublet and a triplet structure, respectively. Compared to a FODO, the doublet provides longer drift spaces for special equipment. Its betatron functions have steep asymmetric slopes. Triplets have distinct locations of high betatron functions but then allow to focus them down over long distances. These low betatron functions in the long drifts allow for small dipole apertures. Triplets also feature a high phase advance, e.g. for closing dispersion bumps. A typical example of a final

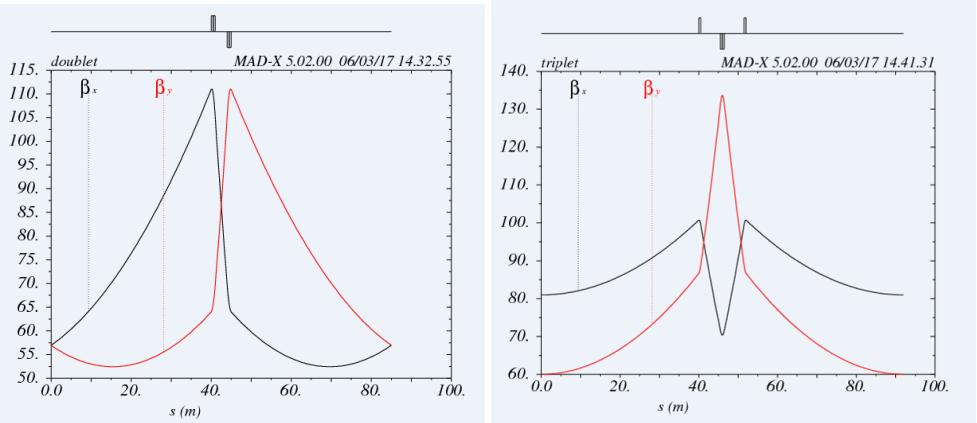


Fig. 11: Optics comparison of doublet (left) and triplet (right)

focus is shown in Fig. 12. The triplet structure upstream the experimental area not only allows to focus both planes to a round spot at the target location, but also the final focus can be shifted longitudinally. Also typical for the triplet optics are the large betatron peaks in the quadrupoles which often requires a dedicated large aperture magnet design.

3.3 Achromats

For many fixed target experiments it is beneficial that the trajectory be independent of the particle momenta and thus, dispersion is usually matched to zero at the focal point. Also for beam transfer between synchrotrons, dispersion in the injection and extraction regions is limited to $\mathcal{O}(10cm)$ to keep beam sizes small and in case of high energy beam transfer, protection system settings independent of momentum errors. In the following the optics conditions of an achromat are introduced. Let us consider two dipole magnets enclosing a quadrupole as shown in Fig.13. The transfer matrix of this system can be written as

$$M = \begin{pmatrix} 1 & 0 & 0 \\ -hs_2 & 1 & s_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & L_2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -1/f & 1 & s_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & L_1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -hs_1 & 1 & s_1 \\ 0 & 0 & 1 \end{pmatrix} \tag{26}$$

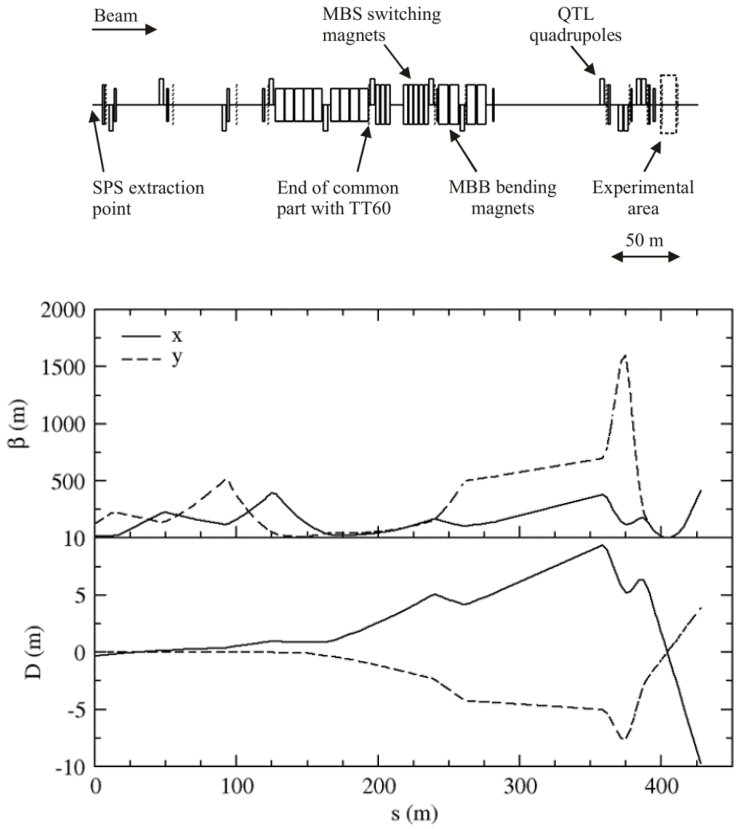


Fig. 12: Final focus optics of the material test facility HiRadMat at CERN [7]

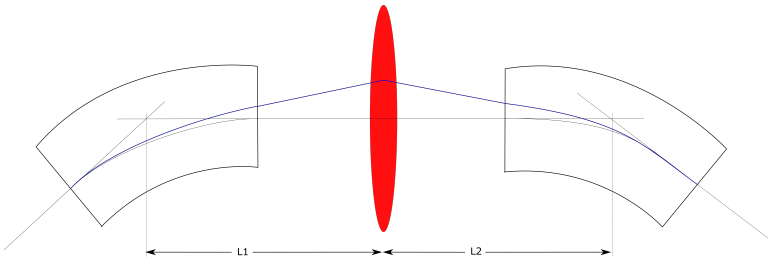


Fig. 13: Basic cell for definition of achromatic condition

with the following relations for the matrix of a single dipole:

$$R_{bend} = \begin{pmatrix} 1 & 0 & 0 \\ -hs & 1 & s \\ 0 & 0 & 1 \end{pmatrix} \quad \ldots \quad h = \frac{B}{B\rho},\ s = \sin\alpha \tag{27}$$

By multiplying Equ. 26 we get the relation for the dispersion and its derivative:

$$\begin{aligned} D_x &= L_2 s_1 \left(1 - \frac{L1}{f}\right) + L_1 s_1 \\ D'_x &= L_1 s_1 \left(-h s_2 - \frac{1}{f} + \frac{h s_2 L_2}{f}\right) + s_1 \left(1 - h s_2 L_2\right) + s_2 \end{aligned} \tag{28}$$

For an achromat, both must vanish which defines the strengths and lengths parameters in our system to:

$$\begin{aligned} \frac{1}{f} &= \frac{1}{L_1} + \frac{1}{L_2} \\ L_1 s_1 &= L_2 s_2 \end{aligned} \tag{29}$$

Several achromat examples with point-to-point focusing are shown in Fig. 14. They are mostly used

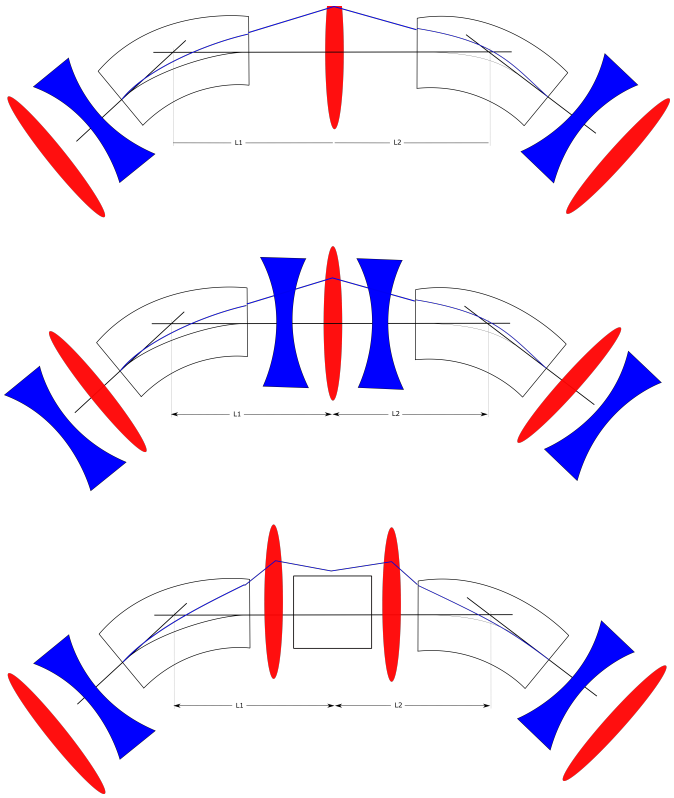


Fig. 14: Achromat examples with point-to-point focusing mostly used in high brightness machines. The double bend achromat (DBA) or Chasman-Green lattice (top), an expanded DBA (middle) and a triple bend achromat (bottom) [10].

in synchrotron light sources. For the Chasman-Green double bend achromat, the central quadrupole is focusing the dispersion down to zero at the outer ends of the dipoles and contributes little to the overall focusing. The quadrupoles outside the dipoles do not alter the dispersion behaviour, they solely provide focusing in both planes. A target placed in the centre can be used as a spectrometer. In contrary to the Chasman-Green lattice, there is strong focusing in the dispersive area for the expanded DBA and the triple bend achromat. For the latter two, one should consider installing sextupoles for chromaticity correction.

3.4 Error and correction studies

After having designed a first order optics for our beam transfer, we need to go a step deeper and evaluate this design against different kinds of errors. In a first step we need to define tolerable error levels in the transfer line. This can be a loss level in W/m, percentage of beam lost or activation level, the trajectory offset in position and angle at the point of delivery, the shot-to-shot and long term variation of the trajectory and the optics, rotation or energy mismatch at the point of delivery. As second step we need to assume reasonable errors for the hardware and study their impact. The error input needs then to be iterated until an acceptable level is reached, which allows then to specify the hardware accordingly.

Usually the error assignment in a transfer line starts from the input error which is defined by the preceding machine. In Fig. A.4 an example is given on how to distribute an assumed or measured orbit offset at extraction according to the optics at the transfer line start. As a next step, magnets in the transfer line need to get errors assigned as shown in Fig. A.5. There are different ways of defining a field error, and in the following we will discuss the multipole expansion of the static magnetic field used in MADX and most other optics codes:

$$B_y + iB_x = B_{ref} \sum_{n=1} (b_n + ia_n) \left(\frac{x+iy}{R_{ref}^{n-1}} \right)^{n-1} \tag{30}$$

where b_n, an, are normal and skew multipole coefficients. R_{ref} is the reference radius, and B_{ref} an appropriately chosen normalization. For a long magnet, the vector potential A in the magnet aperture has only the longitudinal component $A(0, 0, Az)$, with:

$$A_z = -B_{ref} \sum_{n=1} (b_n + i\ a_n) \frac{Re\,(x+iy)^n}{n\ R_{ref}^{n-1}} \tag{31}$$

The Hamiltonian then directly contains the multipole coefficients:

$$H(x, y, z, t) = q\phi + \sqrt{\left(\frac{p_z - qA}{1 + k_x x} \right)^2 + p_x^2 + p_y^2 + m^2 c^4} \tag{32}$$

The multipole expansion is analytical and satisfies automatically the equations for the static field in 2D. It is based on a a complete and orthogonal set of basis functions which means that the values of the low order multipoles do not depend on the number of terms included in the expansion. This is important for the precision of dipole, quadrupole and sextupole fields which do not change if the series contains 8 or 10 terms for example. The expansion guarantees convergence for all radii within the reference radius. On a more practical term, the expansion on a circle is well matched to the rotating coil technique which allows for low measurement and low data treatment errors. If the aperture is very asymmetric as in classical dipoles, the multipole expansion can be performed on an elliptical boundary. Circular and elliptical multipoles are related by a linear transformation. The term "Good Field Region" is a concept related to iron-dominated dipole magnets, with an implicit assumption that the field deviations are largest at the borders of the region near the pole. Here field plots are often fitted with a polynomial, which is however not an orthogonal set and does not necessarily satisfy the field equations. While this polynomial representation can be very useful for magnet design, the multipole expansion should be used to translate field errors into beam impact in optics codes. Which field errors do we need to consider in transfer lines? For the linear terms the dipole and quadrupole strengths are relevant, including power converter errors, transverse misalignments and rotations. As first non-linear term, the normal sextupole can be included. These errors affect the emittance growth, rms orbit errors, and chromatic properties of the transfer lines. The field in the measured or calculated aperture should be obtained on the largest possible reference radius R_{ref}. Ballpark figures as mentioned in Section 2.3 can be used as a good guess starting point. As soon as these distributed errors are assigned to all relevant elements

in the line, the impact on the beam can be simulated as shown in Fig. 16. The level of the trajectory distribution before correction is relevant for first setup of a beam. Trajectory correction is then based on small independently powered dipoles (correctors) and instrumentation to measure the position response downstream of the corrector. Horizontal and vertical correction elements are separated according to

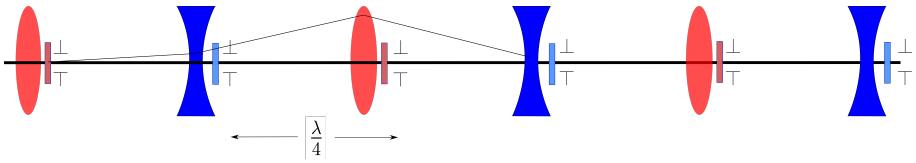


Fig. 15: Setup of correctors and monitors in a FODO lattice

locations of high betatron functions while position pick-ups should be located every quarter of a betatron wavelength as shown in Fig. 15. Special care has to be taken to avoid π-bumps. In case a correction is calculated with many disabled monitors and therefore a poor sampling of phase space, a large trajectory excursion can be created by the correction algorithm. Sufficient instrumentation is therefore essential for trajectory correction. The required correction level is depending on the location in the line. A global correction can be used to minimize the rms offsets at the monitors using all or some of the available correctors. Steering in matching sections and injection and extraction regions requires particular care due to large optics and dispersion functions, limited aperture, losses at collimators and injection offsets being detrimental for performance, i.e. emittance growth. A correction example in MADX is shown in Fig. A.6. One should be careful when running through a loop of many corrections that the corrector kicks are not adding up. As output of such a study it is useful to display trajectories before and after correction and compare to what one is including as trajectory contribution $CO \cdot \sqrt{\frac{\beta}{\beta_{max}}}$ in the aperture specification in Equ. 13. While trajectory offsets in a transfer line are acceptable as long as the beam

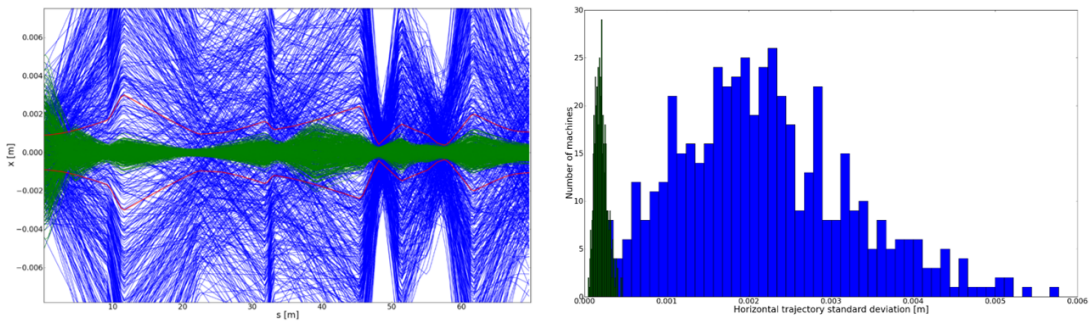


Fig. 16: Trajectories before and after correction (left), the red envelope shows the corresponding contribution in the aperture specification. Distribution of trajectories as a function of maximum offset along the line (right).

envelope does not exceed the good field region of magnets, it is more critical at the end of a line when the beam is delivered onto a target or injected into another machine. In order to perform statistics on the error distribution at the handover point, it is useful to express the error in form of the displacement vector D in normalized phase space as shown in Fig. 17. From this the correlated emittance growth due to filamentation can be directly estimated as:

$$\epsilon_2 = \epsilon_1 + \frac{1}{2}D^2 \tag{33}$$

The next step is a sensitivity analysis to identify the most important contributions to the deliver impreci-

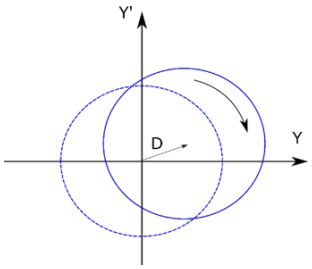


Fig. 17: Displacement vector in normalized phase space

sion and accordingly adjust specifications. This also includes instrumentation errors where the result of the study should define the severeness of reading, scaling and full failure errors. Most synchrotrons use a transverse feedback system to damp injection oscillations. The damping effect of the feedback on the emittance growth can be estimated as:

$$\epsilon_2 = \epsilon_1 + \frac{1}{2}\left((\Delta y)^2 \frac{1+\alpha^2}{\beta} + (\Delta y')^2 \beta\right)\left(\frac{1}{1+\tau_{DC}/\tau_d}\right)^2 \tag{34}$$

where τ_{DC} and τ_d denote the decoherence and damping time, respectively. So far we can specify the peak oscillation amplitude to be damped and the damping time. The required bandwidth of the feedback system needs to be estimated from an FFT analysis of the powering waveforms of the most important error contributions.

4 Special topics in transfer lines

This section focuses on specific topics in transfer lines like particle tracking, perturbation from stray fields, emittance exchange, deflection on a slope and dilution for dump systems.

4.1 Particle tracking

Particle tracking is mostly used to study amplitude dependent impact of non-ideal magnetic fields onto the beam. In particular if special elements are used which are not built-in elements in the code, simulated or measured field maps can be used to track particles through. From the resulting particle distribution the transfer matrix of the special element can be fitted and used as input for an optics code. A real example of this approach is the design of the 100 keV antiproton transfer lines at the Extra Low Energy Antiproton ring at CERN. Due to the low energy transfer lines, electrostatic elements are used to bend and focus the beam. A first order optics was established with MADX and then field maps of strong bending elements were simulated. Particles were tracked through these field maps and the Twiss parameters representing the transfer through the field map were then evaluated from the statistical distributions shown in Fig. 18. In Fig. 19 the optics of a full line tracking is compared to the optics from MADX where quadrupoles are used as built-in elements and bends as transfer matrices. At the end of the line a difference in optics functions can be seen which is due to the hard-edge approximation in MADX which is most visible in the location of the final focus where gradients are strongest.

Another example of particle tracking at beam transfer is the quantification of tail population due to mis-kicked bunches at the SPS injection. The batch spacing required for ions is shorter than the kicker rise time which leads to different deflections of a few bunches at the beginning and end of a batch. To a big part this oscillation can be damped with the transverse feedback and there is no measurable growth of emittance. However, an increased population of the transverse bunch tails is observed by loss

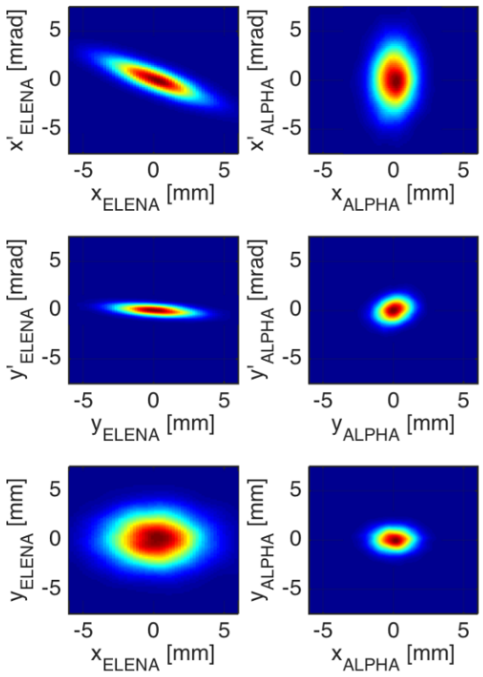


Fig. 18: Phase space distributions of tracked particles at the end of the ELENA ring (left) and the experiment focal point (right) [11].

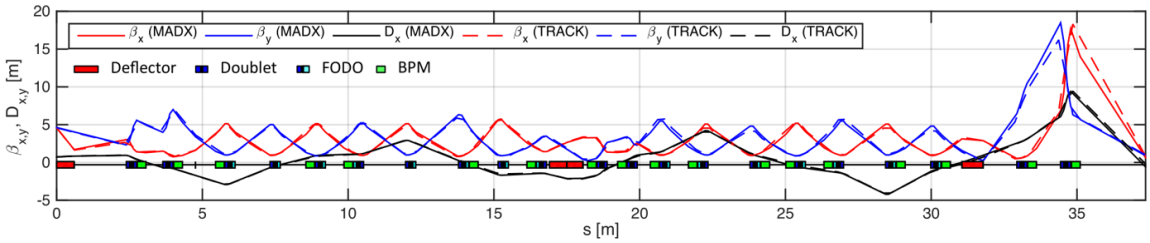


Fig. 19: Optics functions from the ELENA ring to the experimental focal point [11]

measurements on transfer line collimators. In order to simulate this effect, the motion through the SPS ring was described by the effective Hamiltonian formalism [13] where Lie algebra is used to establish a One Turn Map. At turn zero particles get an injection error and in later turns they receive a damping effect from the transverse feedback. In order to establish the coefficients of the effective Hamiltonian, the amplitude dependent detuning needs to be measured in the ring. The chromatic terms and the damper gain are estimated from decoherence measurements as shown in Fig. 20. Figure 21 shows the particle displacement after several thousands of turns for different cases of mis-kick. The gray shaded element stands for the collimator where the displaced particles can be measured as losses.

4.2 Stray fields

A different problem in transfer lines are stray fields which can deteriorate the beam quality or even cause full beam loss. The above mentioned antiproton lines lead a 100 keV beam to several experiments which

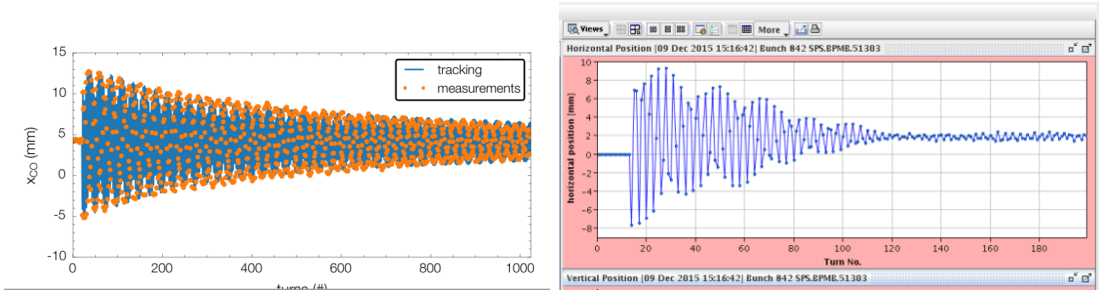


Fig. 20: Measurement and tracking results of amplitude dependent detuning (left) and decoherence measurements (right), courtesy F.M. Velotti.

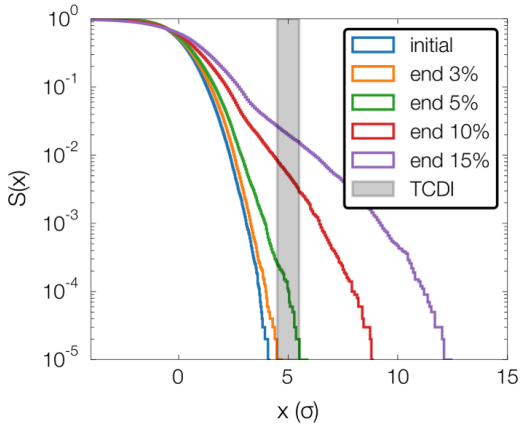


Fig. 21: Particle displacement after several thousands of turns, courtesy F.M. Velotti

have solenoids of up to 5 T field. In order to estimate the impact of these solenoid stray fields onto the beam, the solenoids' magnetic field is modeled with an analytic description of current loops and the integrated transversal fields are translated along the beam line into horizontal and vertical kicks. The resulting effect on the beam trajectory is shown in Fig. 22. The straight lines show the available aperture, which means that trajectory correction is required to stay within the available aperture. Stray fields can be attenuated with magnetic shielding and in Fig. 23 it is shown that a factor 50 attenuation is required to allow for an rms trajectory correction of 10^{-5} m. In case one needs to be independent of different solenoids ramping their fields simultaneously, an attenuation factor of 800 is needed.

The measured field might contain also quadrupolar or higher components. These can be taken into account as multipole kicks in MADX. In case of the PS injection, the stray field of the PS main combined function magnet units was measured and translated into multipole kicks every 2 cm along the last meters before injection.

4.3 Emittance exchange

Acceptances of circular accelerators tend to be larger in the horizontal plane in order to minimize as much as possible the bending dipole gap height. Several multiturn extraction processes produce however beams with emittances larger in the vertical plane. By exchanging the horizontal and vertical planes in a transfer line, we can inject a beam better matched to the acceptance of the downstream accelerator

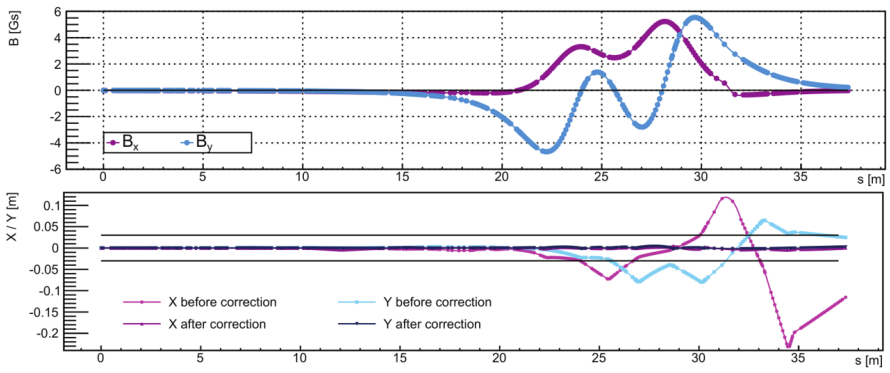


Fig. 22: Stray field level (top) and trajectory displacement (bottom) along the 100 antiproton transfer lines [12].

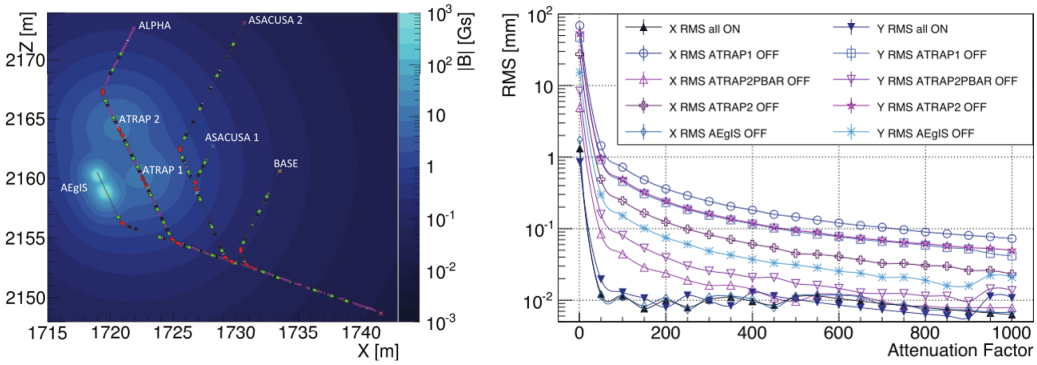


Fig. 23: Magnetic field map in the antiproton decelerator hall at CERN (left) and the respective rms trajectory variation as a function of magnetic shielding attenuation factor [12].

and such, reduce losses at the injection process, see Fig. 24. Skew quadrupoles, which are normal

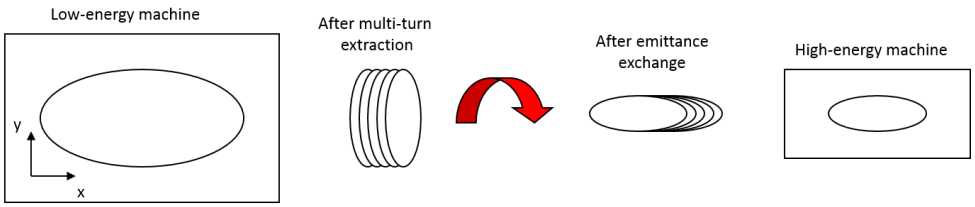


Fig. 24: Emittance exchange to reduce losses in downstream accelerators [3]

quadrupoles rotated by 45^o, introduce a rotation into the beam. A rigorous derivation of this rotation can be found in [5]. The simplest result for a 90^o FODO lattice is given by three skew quadrupoles, each separated by 90^o phase advance, see Fig. 25 and a focal strength of:

$$\delta_{skew} = \frac{1}{\sqrt{\beta_x \beta_y}} \tag{35}$$

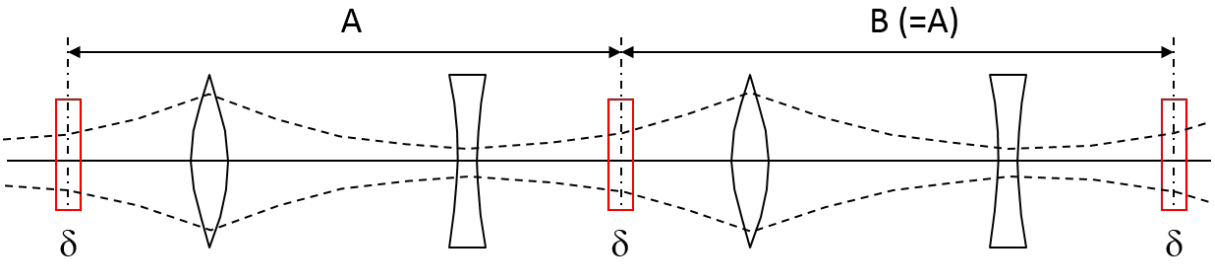


Fig. 25: Emittance exchange insertion with three skew quadrupoles in a 90^o FODO lattice [3]

4.4 Tilt on a slope

In the previous section we have seen that skew quadrupoles can tilt beams around their longitudinal axis. The same effect occurs in a transfer line where e.g. a horizontal dipole is installed on a vertical slope [15]. The resulting tilt amounts to

$$\psi = \alpha\, \phi \tag{36}$$

where α and ϕ denote the bend and the slope angles, respectively. From a single dipole, this effect is small and in the order of misalignment, but it becomes significant when accumulated over many dipoles over a long transfer line. In case of the LHC transfer line TI 8 the roll angle accumulates to 65 mrad. LHC itself is inclined but usually we calculate rings as if they were located in the x-z plane and the survey becomes 2D, thus, tilt and roll angle are zero. If one considers the LHC inclination, the tilt and roll angle oscillate but are perfectly closed after one turn:

$$\phi_{max}\, cos\theta \qquad ... \qquad \phi_{max} = 14.04\, mrad \tag{37}$$

At the injection point, LHC has a local tilt angle of 13 mrad which is not matched to the tilt at the end of TI 8. The corresponding emittance growth is negligible with about 1% for LHC. In general one should match trajectories in all six geometric degrees of freedom. MADX is correctly taking into account the roll angle calculation for dipoles on a slope, however, one needs to make sure to calculate also for a ring the local tilt at the injection point as matching target.

4.5 Dilution

Beam dilution needs to be applied if the deposited energy of the full beam is above the damage limit of dumps or targets. Hydrodynamic tunneling tests at the HiRadMat facility at CERN showed that the LHC beam - if not diluted - would drill a hole of about 35 m length into copper. For an undiluted FCC beam the length of the hole would become around 300 m in copper [16]. For certain targets it is sufficient to blow up the beam size to reduce the energy density in the material. In case of LHC and FCC this would lead to unfeasible absorber dimensions. Instead one is using time-varying dipole fields which give each bunch a different deflection and therefore distribute the particle density on the absorber as shown in Fig. 26. This bunch pattern on the dump is the input to study the shower development inside the block as discussed in [17].

5 Conclusions

In the previous sections we went through the design steps of a transfer line. It is revealing how far such a design can lead on paper. One can define the number of dipoles, quadrupoles, correctors and monitors with great accuracy. Also dipole and quadrupole fields as well as their apertures can be defined. Even

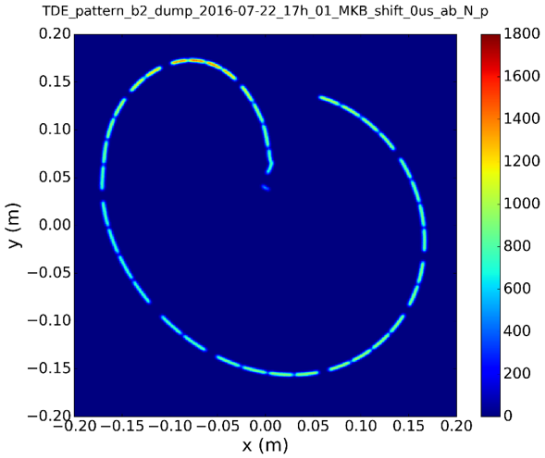


Fig. 26: Bunch pattern on the LHC dump block

a rough estimate of the required field quality and alignment accuracy is possible before switching on a computer.

Error and correction studies, however, show clearly the limits between paper studies and the need of deploying optics codes. We introduced with code examples how to calculate the optics for matching and final focus sections and this allowed to precise the field estimates from paper studies for quadrupoles and dipoles. We went through examples of error and corrections studies to define misalignment tolerances, field homogeneity and ripple, specifications of transverse feedback systems and sensitivity of instrumentation.

For certain cases it is required to dig deeper and track particles, e.g. numerical particle tracking through field maps in case we want to use elements which are not built-in to our codes or for the dedicated problem of transverse tail population due to injection error. We saw problems of transfer lines on the energy extremities as stray field calculation and mitigation in case of 100 keV antiproton lines next to 5 T solenoids or the need for diluting the beam in the high energy dump lines as for LHC and FCC.

6 Acknowledgement

The contributions, discussions and helpful input from R. Baartman, D. Barna, M. Barnes, C. Bracco, P. Bryant, F. Burkart, V. Forte, M. Fraser, B. Goddard, C. Hessler, D. Johnson, V. Kain, T. Kramer, A. Lechner, J. Mertens, R. Ostojic, J. Schmidt, L. Stoel and C. Wiesner are greatly acknowledged.

Appendices

A MADX code examples

References

[1] B. Goddard, Introduction to Beam Transfer, these proceedings, CAS (2017).

[2] N. Ibarrola and M. Jones, Analysis and Evaluation of the CERN Reference Systems, International Workshop of Accelerator Alignment, ((2016)

[3] B. Goddard, Transfer lines, CAS (2004).
`http://cas.web.cern.ch/cas/Baden/PDF/transfer.pdf`

[4] D. Johnson, Challenges associated with 8 GeV H- Transport and Injection for FERMILAB PROJECT X, Proceedings of Hadron Beam 2008, (2008)

[5] P. Bryant, Beam Transfer Lines, CAS, (1992) http://cds.cern.ch/record/261067?ln=en

[6] C. Hessler et al., Transfer Lines to and from PS2, IPAC, Japan, (2010). https://cds.cern.ch/record/1277940/files/sLHC-PROJECT-REPORT-0043.pdf

[7] C. Hessler et al., Beam Line Design of the HiRadMat Test Facility, EuCARD-CON-2009-053, (2009). https://cds.cern.ch/record/1307570/files/EuCARD-CON-2009-053.pdf

[8] D. Carey, The Optics of Charged Particle Beams, (1987)

[9] J. Rossbach and P. Schmueser, Basic Course on Accelerator Optics, CAS, (1992) http://cds.cern.ch/record/247501?ln=en

[10] P.J. Bryant and K. Johnsen, The Principles of Circular Accelerators and Storage Rings, (1993).

[11] M. Fraser, Beam Dynamics Studies of the ELENA Electrostatic Transfer Lines, CERN-ACC-2015-340, (2015). https://cds.cern.ch/record/2141774

[12] J. Jentzsch et al., Beam Dynamics Studies of the ELENA Electrostatic Transfer Lines in the Presence of Magnetic Stray Fields, Proceedings of IPAC2016, Korea, (2016), https://cds.cern.ch/record/2207454/files/thpmb047.pdf

[13] W. Herr, Mathematical and Numerical Methods for Non Linear Beam Dynamics in Rings, CAS, (2013).

[14] L. Deniau, The MAD-X Program, CERN, (2017) http://cern.ch/madx/releases/last-rel/madxuguide.pdf

[15] B. Goddard et al., Geometrical Alignment and associated beam optics issues of transfer lines with horizontal and vertical deflection, LHC Project Report 719, (2004). https://cds.cern.ch/record/733786/files/lhc-project-report-719.pdf

[16] F. Burkart, A. Piriz, R. Schmidt, A. Shutov, N. Tahir, D. Wollmann and M. Zerlauth, Experimental and Simulation Studies of Hydrodynamic Tunneling of Ultra-Relativistic Protons, Proceedings of IPAC2015, USA, (2015). http://inspirehep.net/record/1417473/files/mopty052.pdf

[17] A. Lechner, Particel Interactions with Matter, these proceedings, CAS, (2017)

```
TITLE, 'TTL1';

// read in sequence file
OPTION, -ECHO, WARN, -INFO;
CALL,        FILE = "ttl1-v1.0.0.seq";

// read in beam properties
call, file = "beam.madx";

//----------------------- Survey matching to PS2 --------------------

USE,        SEQUENCE = TTL1;

// read in survey coordinates at the injection point (stripping foil)
// in the PS2 (real beam direction)
call, file = 'out.ps2foil-v8.0.4.survey.ini.mad';

value,        ps2.omk.foil.x0,
        ps2.omk.foil.y0,
        ps2.omk.foil.z0,
        ps2.omk.foil.theta0,
        ps2.omk.foil.phi0,
        ps2.omk.foil.psi0;

// define macro for position matching routine:
m1: MACRO={USE,        SEQUENCE = TTL1;
           SURVEY,
            x0 = -1510.09247,
                        y0 = 2423.58630,
                        z0 = 2278.25946,
                        theta0 = 0.8715909,
                        phi0 = -0.0150853,
                        psi0 = 0.0;};

// match position and angles at injection point (stripping foil):
MATCH, USE_MACRO;

//VARY, name = MB.v1.angle,   STEP = 0.001, LOWER = 0.01,  UPPER = 0.03;
VARY, name = MB.h1.angle,     STEP = 0.001, LOWER =-0.06,  UPPER =-0.01;
VARY, name = MB.h2.angle,     STEP = 0.001, LOWER =0.01,   UPPER = 0.06;
//VARY, name = MB.v2.angle,   STEP = 0.001, LOWER =-0.033, UPPER =-0.01;
VARY, name = ps2.omk.foil.at, STEP = 0.001, LOWER =400,    UPPER = 425;

USE_MACRO, name= m1;

CONSTRAINT,   expr= table(survey,ps2.omk.foil,x) = ps2.omk.foil.x0;
//CONSTRAINT, expr= table(survey,ps2.omk.foil,y)=ps2.omk.foil.y0;
CONSTRAINT,   expr= table(survey,ps2.omk.foil,z)=ps2.omk.foil.z0;
CONSTRAINT,   expr= table(survey,ps2.omk.foil,theta)=ps2.omk.foil.theta0;
//CONSTRAINT, expr= table(survey,ps2.omk.foil,phi)=ps2.omk.foil.phi0;
//CONSTRAINT, expr= table(survey,ps2.omk.foil,psi)=ps2.omk.foil.psi0;

LMDIF, CALLS:=5000, TOLERANCE:=1.E-10;
//JACOBIAN,  CALLS= 500, TOLERANCE= 1.0E-21;
ENDMATCH;
```

Fig. A.1: Example of survey matching code, courtesy C. Hessler

```
// FODO cell properties
l.fodo    = 25;
phadv     = pi/2.0;// 90deg phase advance per cell
k1.theo   = 4.0 / (l.mq * l.fodo) * sin(phadv/2.0);
value, l.fodo, phadv, k1.theo;

kqif.cell = k1.theo;//  5.98643e-02
kqid.cell =-k1.theo;// -5.98462e-02


//----------------------- Prototype Cell ------------------------

FCELL: SEQUENCE, L = l.fodo;

MQF.b:MQH,  AT := l.MQ/4,             K1 := kqif.cell;
MQD.a:MQH,  AT := l.fodo*0.5-l.MQ/4,  K1 := kqid.cell;
MQD.b:MQH,  AT := l.fodo*0.5+l.MQ/4,  K1 := kqid.cell;
MQF.a:MQH,  AT := l.fodo-l.MQ/4,      K1 := kqif.cell;

ENDSEQUENCE;

USE,   SEQUENCE = FCELL;
MATCH, SEQUENCE = FCELL;
  VARY, NAME = kqif.cell, STEP = 0.001, LOWER = 0.0, UPPER = mq.k1.max;
  VARY, NAME = kqid.cell, STEP = 0.001, LOWER = -mq.k1.max, UPPER = 0.0;
  CONSTRAINT, RANGE = #E, MUX = phadv/(2*pi), MUY = phadv/(2*pi);
LMDIF, CALLS:=5000, TOLERANCE:=1.E-10;
ENDMATCH;

SELECT,  FLAG=TWISS, COLUMN=NAME,S,BETX,BETY,ALFX,ALFY,DX,DPX,DY,DPY;
TWISS,   FILE = "fodo.twiss";
```

Fig. A.2: Example of matching a single FODO cell to 90^o phase advance, courtesy C. Hessler

```
//----------------------- Match TTL1 to SPL -----------------------

USE,   SEQUENCE = TTL1;
MATCH, SEQUENCE = TTL1, VLENGTH = TRUE,
   BETX = BETX0, ALFX = ALFX0,
   DX   = DX0,   DPX  = DPX0,
   BETY = BETY0, ALFY = ALFY0,
   DY   = DY0,   DPY  = DPY0;
  VARY,  NAME = kqid.0, STEP = 0.001, LOWER = -0.1, UPPER = 0.0;
  VARY,  NAME = kqif.1, STEP = 0.001, LOWER = 0.0, UPPER = 0.1;
  VARY,  NAME = kqid.1, STEP = 0.001, LOWER = -0.1, UPPER = 0.0;
  VARY,  NAME = kqif.2, STEP = 0.001, LOWER = 0.0, UPPER = 0.1;
  CONSTRAINT,   RANGE = MQF.4, BETX = MQF.betx, BETY = MQF.bety;
  CONSTRAINT,   RANGE = MQF.4, ALFX = MQF.alfx, ALFY = MQF.alfy;
  //CONSTRAINT, RANGE = #E,        DX>-0.1, DX<0.1, DY>-0.1, DY<0.1;
  //CONSTRAINT, RANGE = #E,        DPX=0, DPY=0;
LMDIF, CALLS:=5000, TOLERANCE:=1.E-10;
ENDMATCH;

//----------------------- Match TTL1 to PS2 -----------------------

USE,         SEQUENCE = TTL1;
MATCH,         SEQUENCE = TTL1,         VLENGTH = TRUE,
   BETX = BETX0, ALFX = ALFX0,
   DX   = DX0,   DPX  = DPX0,
   BETY = BETY0, ALFY = ALFY0,
   DY   = DY0,   DPY  = DPY0;
  !VARY,  NAME = kqid.12, STEP = 0.001,   LOWER = -0.1, UPPER = 0.0;
  VARY,   NAME = kqif.13, STEP = 0.001,   LOWER = 0.04, UPPER = 0.1;
  VARY,   NAME = kqid.13, STEP = 0.001,   LOWER = -0.1, UPPER = 0.0;
  VARY,   NAME = kqif.14, STEP = 0.001,   LOWER = 0.04, UPPER = 0.071;
  VARY,   NAME = kqid.14, STEP = 0.001,   LOWER = -0.1, UPPER = 0.0;
  VARY,   NAME = kqif.15, STEP = 0.001,   LOWER = 0.04, UPPER = 0.07;
  VARY,   NAME = kqid.15, STEP = 0.001,   LOWER = -0.1, UPPER = 0.0;
  VARY,   NAME = kqif.16, STEP = 0.001,   LOWER =  0.0, UPPER = 0.1;
  VARY,   NAME = kqid.16, STEP = 0.001,   LOWER = -0.1, UPPER = 0.0;
  !VARY,  NAME = MQ.e10.at, STEP = 0.001,   LOWER = 115,  UPPER = 120;
  !VARY,  NAME = MQ.e11.at, STEP = 0.001,   LOWER = 123,  UPPER = 127;
  CONSTRAINT,  RANGE = ps2.omk.foil,  BETX=BETX1,  BETY=BETY1;
  CONSTRAINT,  RANGE = ps2.omk.foil,  ALFX=ALFX1,  ALFY=ALFY1;
  CONSTRAINT,  RANGE = ps2.omk.foil,  DX=DX1,      DPX=DPX1;
  CONSTRAINT,  RANGE = ps2.omk.foil,  DY=DY1,      DPY=DPY1;
LMDIF, CALLS:=5000, TOLERANCE:=1.E-10;
ENDMATCH;
```

Fig. A.3: Example for the matching sections from SPL to FODO and from FODO to PS2, courtesy C. Hessler.

```
corrmacro(nx): macro={

use, sequence=bt1btp;
sigma:=2;
vert:=TGAUSS(2);

EOPTION,SEED=nx,ADD=false;

eps_x=2e-6/2.9676;
delta_p=1.07e-3;

rand_x=gauss(nx);
sigma_x = sqrt(eps_x * betx0);
x0 = 0.015*rand_x *( sigma_x + (abs(dx0) * delta_p));

rand_px=gauss(nx);
px0 = 0.015*rand_px *((sqrt(eps_x / betx0) - (alfx0 / betx0) * sigma_x))
+ (dpx0 * delta_p);

eps_y=2e-6/2.9676;
//delta_p=1.07e-3;

rand_y=gauss(nx);
sigma_y = sqrt(eps_y * bety0);
y0 = 0.01*rand_y * (sigma_y + (abs(dy0) * delta_p));

rand_py=gauss(nx);
py0 = 0.01*rand_py *((sqrt(eps_y / bety0) -
(alfy0 / bety0) * sigma_y)) + (dpy0 * delta_p);
```

Fig. A.4: Example for orbit correction studies: input error

```
EOPTION,ADD= FALSE, SEED:=nx;
SELECT, flag= error, clear;

SELECT, FLAG= ERROR, PATTERN= "BT.BHZ10";
EFCOMP, ORDER:= 0, Radius:=0.01,
DKNR:={5e-4*TGAUSS(sigma),0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0},
DKSR:={0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0};
EALIGN,  DS:= 0.3e-3*TGAUSS(sigma), DX:= 0.3e-3*TGAUSS(sigma),
DY:= 0.3e-3*TGAUSS(sigma), DPHI:= 0.3e-3*TGAUSS(sigma),
DTHETA:= 0.3e-3*TGAUSS(sigma), DPSI:= 0.3e-3*TGAUSS(sigma);
```

Fig. A.5: Example for orbit correction studies: assigning errors to magnets

```
SELECT,FLAG=ERROR,FULL;
ESAVE,FILE='errors.tfs';

select, flag=twiss, clear;
select, flag= twiss, column= s,x,y,betx,bety,px,py;
TWISS,file="outcorr/test/twissnocorr.nx.tfs", deltap= 0.0,
sequence= bt1btp,BETA0=INITBETA0;

option,echo;

COPTION, PRINT=10;

CORRECT, flag = line, PLANE= x, MODE=svd, COND=0, MONON=1,
MONERROR=1, MONSCALE=0, RESOUT=0, ERROR=1.E-6, CORRLIM=1.0,
CLIST=xcorr.nx.out, MLIST=ymon.nx.out;

CORRECT, flag = line, PLANE= y, MODE=svd, COND=0, MONON=1,
MONERROR=1, MONSCALE=0, RESOUT=0, ERROR=1.E-6, CORRLIM=1.0,
CLIST=ycorr.nx.out, MLIST=xmon.nx.out;

select, flag= corr, column= PX.OLD,PY.OLD,PX.CORRECTION,PY.CORRECTION;

select, flag=twiss, clear;
select, flag= twiss, column= s,x,y,betx,bety,px,py;
TWISS,file="outcorr/test/twisscorr.nx.tfs", deltap= 0.0,
sequence= bt1btp,BETA0=INITBETA0;
};

n=0;
while (n < 200) {
  exec, corrmacro($n);
 n= n+1;
}
```

Fig. A.6: Example for orbit correction studies: correction part

Proceedings of the CAS–CERN Accelerator School: Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018)

Machine Protection and Activation

A. Nordt, R. Andersson, E. Bargalló, S. Kövecses
European Spallation Source ERIC, Lund, Sweden

Abstract
Machine protection can be described as a strategy that modern particle accelerator driven research facilities utilize to achieve high and optimized operational availability. In this context it is important to define what the terms 'operational availability' and 'machine protection strategy' intend to address. Research facilities have become more complex and also more expensive during the past decades, often aiming at unprecedented beam energies or beam power levels. Availability in the context of particle accelerator driven research facilities is often referring to 'availability of beam' that can be used for scientific applications performed by 'users'. This paper provides a general overview of how machine protection can be utilized to increase beam availability of accelerator driven research facilities.

Keywords
CERN report; machine protection; activation; risk management, availability.

1 Machine protection: a term, a system, or a strategy?

In this paper, the authors focus on providing a general overview of what machine protection (MP) is, how to apply machine protection methods and strategies, what criteria to address, and what guidelines potentially to follow when being confronted with the task of implementing or optimizing machine protection at a (complex) accelerator driven research facility. Thus the reader should not expect a detailed description of the mechanisms that can lead to damage, details on the beam physics aspects related to beam-induced damage at high beam energy or high beam power machines, or details on how to mitigate against certain hazards. There are excellent publications on these topics such as from the Joint International Accelerator School on Beam Losses and Accelerator Protection, that took place in 2014, see Ref. [1] and much more, see e.g., Ref. [2].

Instead the reader should be aware that he/she will be guided through the following main topics in a rather strategically inspired way.

- What does high availability of a research facility mean?
- How can we describe availability?
- Why could high availability become more and more important nowadays?
- What could machine protection strategy mean in the context of operational availability?
- How could Machine Protection be implemented?

1.1 What could possibly go wrong?

This section intends to give a brief overview of different types of damage events that have occurred in the past at accelerator driven research facilities. The accidents described are either beam induced or non-beam induced.

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-005.533

In 2004, during extraction tests with 450 GeV/c protons in the SPS[1] extraction line at CERN, Switzerland, a beam with an energy of 2 MJ was deflected with grazing incidence into a vacuum chamber. This happened after the failure of a septum magnet. The vacuum chamber was cut along a length of 25 cm. A magnet further downstream was damaged due to beam losses arising from this incident. Condensed drops of steel were visible on the opposite side of the vacuum chamber and the vacuum chamber and quadrupole magnet had to be replaced, see Ref. [2].

Another accident happened at Tevatron, Fermilab, USA. A roman pot was moved into the beam and the particle showers generated by the Roman pot's movement quenched superconducting magnets located further downstream. What happened was that the beam moved by 0.005 mm/turn, and eventually touched a collimator jaw surface after about 300 turns. The entire beam was then lost, mostly on the collimator, that was damaged, see Ref. [2].

At SNS[2], Oakridge, USA, superconducting radio-frequency (RF) cavities are used to accelerate the beam. This equipment is operated at very high voltage levels and can be very sensitive to even very small beam losses. Beam losses then can lead to a surface quality degradation of the cavity. This means that operation at the same voltage level is not possible anymore and the probability for arcing is increased. At SNS, errant beam losses led to degradation of a superconducting cavity. Beam current monitors (BCMs) measured beam losses of a few microseconds. After such errant beams, sometimes the cavity gradient needed to be lowered. Conditioning after warm up helped in most cases, but in one case a complete cryo-module had to be exchanged because the degradation was too severe. The energy of beam losses in this case is only about 10−100 J and rather difficult to detect. The damage mechanisms are not yet fully understood; it is assumed that some of the beam hitting the cavity desorbs gas or particulates, creating an environment for arcing, see Ref. [2] and Ref. [3].

One more example from SNS is not related to beam-induced damage. In 2014, the water cooling of the absorber of the medium energy beam transfer (MEBT) chopper failed and water entered the MEBT, causing several weeks of downtime. Luckily, this happened during a maintenance period during which all the vacuum gate valves were closed. The water from the damaged MEBT chopper absorber was 'stopped' in the drift tube linac[3] (DTL) section. The absorber was not interlocked and also not checked periodically until this accident happened.

At LHC[4], CERN, an accident happened during test runs without a beam in 2008. A magnet interconnection was malfunctioning and a resulting electrical arc caused the release of a huge helium pressure wave (superconductivity of the helium was lost). Around 600 m of the ring's equipment were damaged and 2 km of the vacuum chamber had to be cleaned. A total of 53 magnets had to be repaired, see Ref. [2] and references therein.

1.2 What if things have gone wrong?

The initiating events and also the consequences of accidents like the ones mentioned in section 1.1 are manifold.

The obviously not so good consequences of such accidents are, from an organizational viewpoint:

- Unplanned downtime/no time for science;
- Additional cost for repair or replacement of damaged equipment;

[1] Super Proton Synchrotron
[2] Spallation Neutron Source
[3] Linear Accelerator
[4] Large Hadron Collider

– Unnecessary beam losses leading to unnecessary activation of equipment;
– Unnecessary beam losses leading to unnecessary stress of equipment and potential early aging.

The good outcome of such accidents however should not be neglected. Certainly, accidents impact on the awareness of the importance of establishing a robust safety and protection culture within an organization. Such culture, if embedded in the daily life of the designers, scientists, and users of research facilities, forms a very robust medium for safe operation and helps drastically in decreasing the probability of accidents happening. When addressing protection aspects it is important to also think out of the box. One should not neglect the obvious and easy parts of a system's design. Often, engineers and designers tend to focus on the challenging aspects of a new system's design. This means that they prefer focusing on the challenges imposed by using a new technology or design approach, rather than also focussing on the 'normal' equipment that is part of the new system. It is however crucial to keep the global system functionality in mind and to put the system into the context of the whole machine and eventually its role in beam production, acceleration, etc. Many accidents in fact happen because of a non-optimal interplay of several systems and lack of awareness of the global picture.

Machine protection tries to cope with this complexity and uncertainty by providing rather flexible functionality that helps to prevent and mitigate the risks of damage events to an acceptable level.

1.3 What is machine protection?

The latest generation of accelerator driven research facilities are very complex and often a high investment must be made to build these machines. As their purpose is to serve external users, these accelerator driven facilities need to fulfil the demands of the users both in terms of machine performance and availability. Numerous operational risks have to be accounted for when operating such machines. Two of the most imminent risks are damage to equipment and the associated downtime of the facility. The increased beam energies and beam power levels lead to increased damage potential of the beams, both from the particle beam and through the increased stress on the equipment. In addition to the unwanted downtime related to damage, there is also a certain cost to consider.

It is therefore necessary for accelerator facilities to have a strategy and systems that address these risks appropriately. Such systems are generally referred to as machine protection systems (MPS), even though the 'S' has become more of a strategy, or system-of-systems, than a single system that operates independently of other systems.

Based on the complexity, cost, and uniqueness of many research facilities but also more and more the demands for machine and beam availability from users, together with the knowledge that 'things can go wrong', the following two goals can describe what machine protection is there for.

Goal 1: Machine protection shall, in this order, prevent and mitigate damage to the machine, be it beam induced or from other sources, in accordance with beam and facility related availability requirements.

Goal 2: Machine protection shall protect the machine from unnecessary beam-induced activation having a potential to cause long-term damage to the machine or increase maintenance times. This shall be addressed in accordance with beam and facility related availability requirements.

It should be noted that machine protection cannot be implemented by a single group of people or a single work package/project. A common effort across the whole organization is needed to ensure the right level of protection. Teamwork is vital for implementing machine protection. Awareness, openness, global thinking, as well as understanding of the impact and consequences of certain decisions on a global (machine wide) level, are highly important.

2 When is machine protection needed?

In order to know how and what to protect against what kind of failures it is vital to set up a failure catalogue of your machine. There are a lot of different risk and hazard analysis methods available, however, it is unlikely that a single one of these could cover all aspects needed for machine protection. These methods are in almost all cases targeted towards safety related analysis. In general it should be avoided to perform a 'silo-system' component failure analysis, but rather focus on performing a functional analysis. Functional analyses can become very complex in the case of an accelerator, where for example the function of accelerating the beam is performed by several rather complex systems. The degrees of freedom for such analyses can easily become very large and then the analysis method itself may become unpractical. The interplay between the different systems performing one function is to be analysed with care.

Since particle accelerators used for research are many times very complex and not necessarily built with only standard off-the-shelf equipment, detailed documentation of e.g., failure modes, may not be available, which makes a standard risk analysis difficult to execute. Most of the research facilities operate 'prototype' machines, using a large amount of state-of-the-art equipment, hence one cannot establish the equipment's failure modes with high confidence nor document them, simply because they are not known. By 'prototypes', we mean here that usually there is only one kind of a specific facility/machine; there are e.g., several colliders around the world (Tevatron, LHC, etc.), however they operate at very different beam energy and beam power levels and use different technologies, hence possible equipment failures can lead to different accidents depending on the machine.

In order to assess whether machine protection is necessary, a large set of criteria are considered. These are, among many other aspects, the damage potential of the beam, the expected beam loss levels at different locations along the accelerator, the delicacy of the equipment (susceptibility for damage, location, environmental aspects, availability of spares, etc.), the beam injection and extraction mechanisms, beam stop capabilities, and requirements on beam and system availability. On top of this there is a need to prepare for unexpected events as modern accelerators take unbeaten paths. In this section we will discuss the criteria to be considered when deciding whether machine protection is needed or not.

2.1.1 Damage potential of the beam

It is important to understand how the beam can damage the different materials of the facility's equipment. To understand heating, melting, and damage mechanisms, one has to know about the particle type, its momentum, the stored energy in the beam, beam power, beam size, beam power/energy density, time structure of the beam, cooling conditions, and more. In order to estimate the order of magnitude of possible damage, the following rule of thumb can be used. One MJ can heat and melt ~ 1.5 kg of copper. One MJ corresponds to the energy stored in ~0.25 kg of Trinitrotoluene (TNT). And one MW during one second corresponds to one MJ, see Ref. [2]. Facilities such as the European Spallation Source, Lund, Sweden (currently under construction) intend to operate at 5 MW average beam power and 125 MW peak power per pulse.

2.1.2 Beam losses

Different types of beam losses lead to different aspects of required protection.

Continuous beam losses are inherent during the operation of accelerators. Usually there is a large effort made during the design phase to optimize continuous beam losses and keep these as low as possible.

Accidental beam losses are transient losses with time scales from nanoseconds to many seconds. These can occur due to a multitude of failure mechanisms. Machine protection protects equipment

from damage, activation, and downtime due to accidental beam losses. Machine protection includes a large variety of systems performing this function. For example: for 1 MW, a loss of 1% corresponds to 10 kW, not to be lost along the beam line to avoid activation of material, heating, etc. Assuming a length of 200 m, such losses would correspond to 50 W/m. However, the 1 W/m rule is a reasonable limit for hands-on maintenance, see Ref. [2].

2.1.3 Activation – hands-on maintenance

An average beam loss of 1 W/m should be a reasonable limit for hands-on maintenance, where 1 W/m corresponds to 6×10^9 protons per [metres times seconds] of energy 1 GeV/c (uniformly distributed). Simulations show that if the irradiation time of a steel pipe with 1 W/m beam losses is 100 days and cool down time is then 4 hours, the effective dose at 30 cm distance from the beam pipe is about 1 mSv per hour, see Ref. [4].

2.1.4 Environmental conditions for equipment

Another criterion to be considered is the environment in which the equipment and its electronic controls are located. It is advisable to e.g., avoid electronics in radiation areas, or areas where high temperatures and high humidity might occur, exposure to water, fire, etc.

It is very important to supply the equipment with rich diagnostics, such as temperature sensors, flow switches, or other sensors, and it is also very important to archive these data in a good way, such that further analysis can be easily performed. Such offline analysis can help to improve the performance of the machine.

2.1.5 Means to inject, extract, stop beam

It is important to understand how the beam is injected into the machine and how it can be extracted or stopped and on what time scales this can be achieved. In linacs it is common to stop the beam by interrupting the source (stop the extraction of plasma), and to use the choppers to deflect the beam to the absorbers. In ring machines usually kicker magnets are used to kick the beam out of the ring into a dump line. Extraction or interruption of beam operation should preferably be possible before melting, or other beam-induced damage with high severity, can occur. Otherwise passive protection needs to be implemented or a re-design undertaken, see Ref. [2].

2.1.6 Availability requirements

Each facility has different requirements on beam and machine availability. This is often because of different sponsoring schemes and budgets, manifold user communities, and scientific goals set up for the specific machines. User and medical facilities usually have very strict availability requirements due to their nature. Many efforts have been made in defining one common availability measurement and analysis method and metrics. This might be very helpful when comparing several machines with each other. This practice works well when comparing facilities of the same type. For example synchrotron light sources. These are strongly user driven and use similar principles and technologies.

2.1.7 Foresee the unforeseeable: how?

Newer facilities nowadays aim towards unprecedented beam energies and beam power levels, and the upcoming failures or incidents are difficult to foresee. This needs to be considered, and planning for rich diagnostics and/or redundancy at an early stage can be crucial to achieve successful operation.

To illustrate what is meant here, we take the example of unknown falling objects (‘UFOs’) detected in the LHC at CERN during 2010 and existing since. These UFOs are probably dust and dirt particles of a few micrometres size only, creating beam losses when ‘falling’ into or through the beam. This phenomenon was not observed before at any other machine and is hard to predict. However, it

was of tremendous advantage to have a very large beam loss monitoring system consisting of 4000 detectors installed around the 27 km long ring. This system allows for a wide coverage and detection of all types of beam losses, from ultra-fast, to fast, to slow. The measurements of this system are taken and archived with high quality and hence it is possible to perform very efficient data analyses. Certain patterns of beam losses caused quenches of magnets and happened very fast (within a few tens of microseconds). The shape of the losses looks similar to the losses created when doing a wire scan, i.e., moving a wire through the beam to measure the beam profile. This is why in the beginning these type of losses were considered to be caused by something falling through the beam. It is very difficult to predict such issues, but reliable beam monitoring systems with a large dynamic measurement range and several integration times are vital to understand them.

2.1.8 Time scales

As mentioned above, failures can be categorized based on severity, but it is also important to categorize them according to other criteria, such as whether the damage is beam induced, non-beam induced, local but impacting on beam, local but not impacting on beam performance. Besides this, it is vital to understand the time scales during which damage can happen. Time evolution scales for damage events can be slow [days, hours, seconds], fast [milliseconds], ultra-fast [nanoseconds, microseconds]. The time leading to a damage event includes several parts: the start time; e.g., a magnet power supply fails and does not provide any voltage to the magnet anymore. Then, there is a certain decay time of the magnetic field during which the effect on the beam will be different. In the beginning, when the magnetic field is only a few per cent lower than anticipated, almost nothing will happen to the beam. At a certain moment in time however, beam losses will occur. Eventually, it can happen that the beam is deflected in a way such that damage to surrounding equipment could happen. Depending on these time scales involved, it is possible to define the timing requirements for the protection functions (PFs). Preferably one can set the goal for machine protection to act in a preventive way first, but always anticipate a layer of protection that can mitigate the event. If damage can happen very quickly, there might be a need to change from active protection to passive protection or to go back and re-design the system, see Ref.[2].

2.1.9 Summary

Machine protection is a facility-wide, availability-driven system of systems, which has equipment protection as its main goal in order to prevent long downtimes of the facility. It should be noted that the primary purpose of machine protection is not to directly handle internal failures of equipment, but to be involved in the prevention of additional damage that can arise, as an extended effect of the equipment failures or system interactions.

3 Means to achieve machine protection

Summarizing the first two chapters, the following ‘means’ help in achieving high operational availability of accelerator driven research facilities.

3.1 General machine protection (MP) requirements

In the following we summarize the most generic requirements for machine protection.

– Designing and operating the equipment under control (EUC) with high inherent reliability and overall low damage potential.

– Minimization of the mean down time (MDT) of EUC by introducing dedicated technical systems preventing and mitigating damage.

– Minimization of the MDT of EUC systems by introducing dedicated operational and preventive maintenance procedures reducing the probability of the need for corrective maintenance.

– Introducing supporting systems dedicated to reducing MDT. These include analysis, management and recovery tools addressing operational activities related to machine protection.

Machine protection functions should be implemented with timing and protection integrity levels (PILs) in accordance with damage risk reduction requirements. This implementation has to be done such that the probability of spurious or false trips is reduced in accordance with the availability and damage risk reduction requirements.

Machine protection shall transmit all necessary information to the responsible staff, allowing them to take adequate action to resume facility operation within a minimum amount of time. All information about detected off-nominal states should be recorded. Performed prevention and mitigation actions should be recorded as well to allow for a posteriori event reconstruction and analysis. Furthermore, machine protection should support operation during all foreseen lifecycle phases of the machine including, but not limited to, assembly and installation, commissioning, tuning, operation, maintenance, and dismantling.

Machine protection should support all foreseen operating modes of the machine, including but not limited to beam up to intermediary targets, beam with reduced beam power, or alternative duty cycles. Machine protection should also support operation in case of a degraded mode of operation of the EUC if required for reaching the availability goals and if compatible with damage risk reduction requirements. Also, it should support operation in case of degraded protection functions, if required for reaching the availability goals, and still be compatible with damage risk reduction requirements.

3.2 The ‘ P^3 ’ principle – tasks of machine protection

The purpose of machine protection, as mentioned before, is to reduce the scientific and economic losses as much as possible. The task is to optimize damage events such that their occurrence is kept at a level that is as low as reasonably achievable. This can be done by simply interrupting beam operation and adjusting relevant parameters to avoid damage and downtime. On the other hand, simply stopping everything in case of ‘any’ errant situation in a complex facility would lead to very little operational time. Instead, machine protection needs to be able to handle minor problems in the background while operation continues.

In addition to protecting equipment and avoiding unnecessary downtime, information on the cause of the stop can be collected for further analysis, to allow for continuous operational improvements. The tasks for machine protection for an accelerator facility can then be summarized as the ‘ P^3 ’ principle, see Ref. [5].

– Protect the equipment.

– Protect the beam.

– Provide the evidence (of what caused the stop).

Protecting the equipment means avoiding damage due to wrong behaviour or a wrong configuration or due to the beam. Protecting the beam means that interruptions of beam operation should be kept at a minimum level. To provide the evidence, a so-called post-mortem system is implemented. This is a system that, in case of a beam stop, collects data on the current machine configuration, time stamps of when the event occurred, and what system of the machine sent the beam stop signal.

3.3 A system of systems approach to managing machine protection

As the systems required for successful research facility operation are typically diverse in both technology and ownership, a traditional approach, where the integrated systems are treated equally, would not handle this diversity satisfactorily. Instead, a system of systems (SoS) approach is more suitable, see Ref. [6], as this appreciates the emergent properties and operational independence of the constituent systems. According to Ref. [7] and Ref. [8], a SoS has the following criteria for the constituent systems, which are found (by the authors) to match well the protection approach of a complex research facility.

– Operational independence.

– Managerial independence.

– Geographical distribution.

– Emergent behaviours.

– Evolutionary behaviours.

The two latter criteria are not apparent if the systems and their interactions are modelled separately. These five criteria are also sometimes referred to as Maier's criteria of a system of systems, see Ref. [8]. Thus, a complex research facility is a typical example of a SoS with many subsystems that jointly perform a task that could not be done by one system alone. It should be noted that SoS is not a tool, method, or practice in itself, but rather a way of thinking. Therefore, it can be well adapted into the functional protection method, see chapter 5. Adapting this way of thinking naturally expands beyond a limiting 'root cause' mentality as the sole solution to protection problems.

4 Machine protection at ESS: a system of systems

This chapter provides an overview on how to apply the SoS approach at a complex research facility such as the European Spallation Source, Lund, Sweden for machine protection purposes.

4.1 Introduction to the European Spallation Source

The European Spallation Source (ESS), currently under construction in Lund, Sweden, will be a user facility for neutron science, aiming to be the brightest neutron source in the world throughout the next decades.

Protons are accelerated to 2 GeV/c along a 600 metres long linac. The linac is operated in pulsed mode and 2.86 millisecond long pulses are injected at a repetition rate of maximum 14 hertz. The average beam power is 5 MW with a peak power of 125 MW per pulse. The 2 GeV/c protons are steered into a rotating, helium cooled, 5 tons tungsten target where the neutron spallation process takes place. Then, the created spallation neutrons are further moderated to appropriate speed and guided towards the neutron science stations where the different experiments can take place. The first proton beam production will take place in 2018 and the first neutron user program will start in 2023.

The ESS proton beam power will be unprecedented and melting of copper or steel can happen within a few (2−10) microseconds, if the beam is deflected perpendicularly into the beam pipe. This is especially an issue in the normal conducting linac (first 50 m), where the energy of the beam is still low (75 keV – 90 MeV).

4.2 ESS availability-driven machine protection concept

Machine protection at ESS has its starting point in the operational goals and user needs for neutron experiments. From Ref. [10], two relevant goals are the following:

– StR-11 – Annual operating periods: “The partnership expects the ESS scientific infrastructure to be available at least 4000 h of each annual operating period when fully operational”.

– StR-12 – Availability: “The partnership expects that the ESS facility will be designed with a goal of 95% availability during its annual operational periods when fully commissioned”.

Based on these, the ESS Neutron Source Reliability and Availability document, see Ref. [10], outlines a strategy to break the overall goals into more concrete and acceptable downtime requirements for the different ESS systems. Machine protection then transfers this into protection needs, plays a key role in its fulfilment, and integrates a facility-wide strategy and analysis effort to reach the ESS goals for neutron availability. This effort will be described in more detail in the next sections and further in chapter 5.

4.3 ESS machine protection as system of systems

At ESS, a single protection function can be performed by different parts of the machine. These parts are managed by different divisions or groups and designed and built by different laboratories around Europe. This requires a way of organizing the responsibilities, which can be achieved by applying the SoS approach. This approach is currently followed at ESS. All systems that belong to the MP-SoS play a role in the protection of the machine.

The ESS project organization is set up to address each of the three sections of the facility through a dedicated division/directorate: the proton accelerator, spallation target, and neutron science systems. Within each division/directorate, there is a set of work packages that deal with specific parts identified in the facility breakdown structure, Fig. 1. The MP-SoS analyses need to merge into the development of the different work packages at the same time as they tie back to the overall goals of ESS. MP-SoS are therefore continuously observing and analysing the progress of the different work packages, in order to cope with the complexity and system interaction that might lead to damage and beam losses that would cause ESS to not fulfil the overall goals connected to operational availability. The selected risk management method has to be flexible enough to handle continuous scope expansion by allowing modular and parallel work for the different work packages, see Ref. [11].

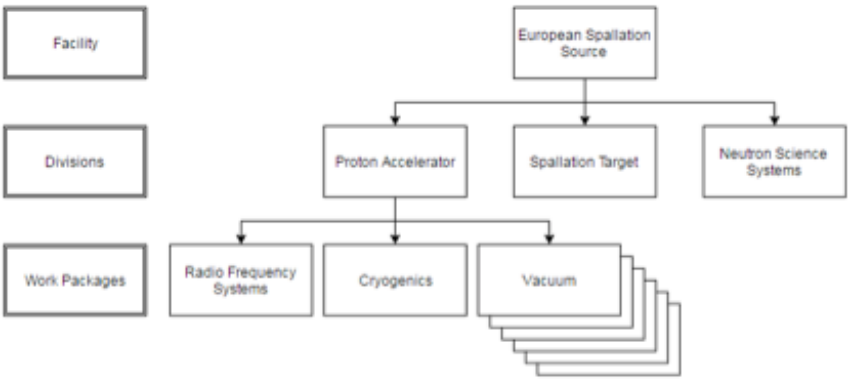


Fig. 1: Part of the facility breakdown structure at ESS

4.4 ESS organizational structure and machine protection responsibilities

The ESS organization has and continues to expand rapidly, hence the culture, communications procedures, and all of the administrative efforts, have to be developed in parallel to the design and construction of the facility. The selected risk management method needs to cope with limited procedures for documentation as well as changing organizational structures internally and externally. In addition this places high demands on flexibility and continuous updates of the information to be

analysed. Traceability then becomes an important aspect already in the early stages, in order to verify that the analyses are matching the latest versions of e.g., the hardware.

The rapid expansion can also be viewed from the perspective of scope, where the MP-SoS includes more systems as the project proceeds, and there is a need to inform and set up discussions with new stakeholders. These stakeholders are themselves required to perform concept designs of their systems at the same time as potential interfaces are changing. It is therefore important for the technical risk managers to recognize that the organizational development affects the technical developments, including systems relevant for machine protection, see Ref. [11].

The protection of the machine occurs at multiple levels. At ESS, three main levels are considered.

The first is at component level, where the manufacturer or the group building it takes care of the protection. Some examples are electronic boards, power supplies, or any stand-alone piece of equipment. This level is quite straightforward to distinguish and it is the responsibility of each system owner.

The second level is related to protection in-between components. This protection generally involves direct relationships between components in the form of services and connections. Some examples of direct relationships are water cooling, controls, power supply, etc. In these cases, the components usually belong to the same system and therefore the protection is the responsibility of the system owner. The protection functions involved in this category are called local protection functions. The example in Fig. 2 shows the local protection for a magnet system where the power supply performs the protection of the magnet. The power supply has to stop delivering power if overheating is detected in the magnet coils. Local protection is different for each system, some of them are quite complex and others are relatively simple.

The third level is focusing on protection functions that are located in-between systems. These functions usually cover damage events that require stopping the proton beam operation to bring the machine to a protected state. In Fig. 2, the second protection function (stop beam operation in case the magnet is not powered since the beam could be deflected in an unwanted and potentially critical way) is performed trough the magnet system, the beam interlock systems (BIS) and the different actuator systems. These together form a so-called global protection function. The ESS MP team is in charge of analysing these functions, retrieving the correct information about the systems that participate in performing these functions, and making sure that the requirements are fulfilled.

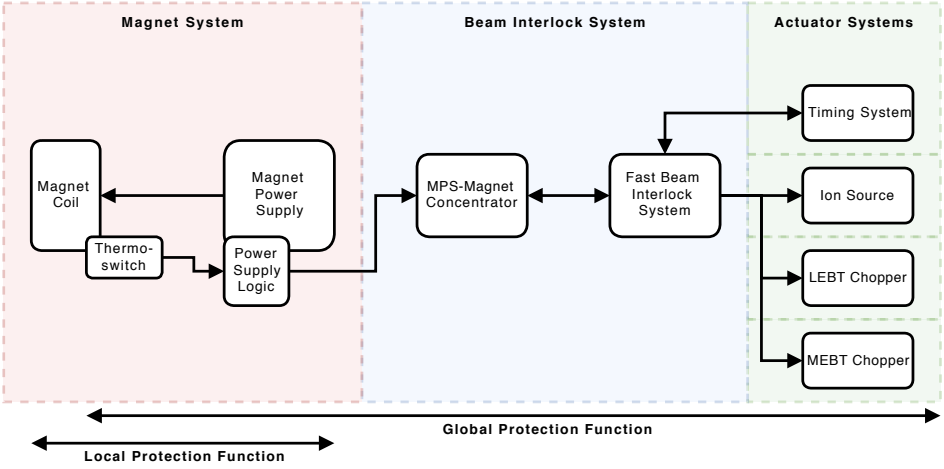


Fig. 2: Example of a global and local protection function as defined at ESS

4.5 Main function of the ESS MP-SoS

The main function of the MP-SoS is the following:

– MP-SoS has to achieve or maintain the protected state of the machine.

If all systems are properly operating and the beam is well controlled, the machine is in a protected state. However, if something that could lead to damage is detected, then the MP-SoS has to act accordingly to reach the protected state again. In order to achieve or maintain this state, the MP-SoS systems have to detect the hazard, interpret if it is acceptable and in case it is not acceptable, stop operation.

4.5.1 Global protection functions and beam related hazards

The global protection functions protect against hazards that are beam related. Some examples of beam related hazards are the following.

– Proton beam is out of the expected position/path and critical beam losses are generated (e.g., beam not well focused or steered or an element is inserted in the beam pipe unexpectedly).

– Beam power exceeds design specifications for a specific beam destination (BD) or insertable device (ID).

– One of the MP-related systems is not operative (e.g., RF transmitter is off and then beam would not be properly accelerated).

For these cases and others, the way to achieve or maintain the protected state is to stop beam operation or to not allow beam operation.

4.5.2 Beam related damage protection

Protection against beam-induced damage can be provided in the following ways.

– Evaluate the status of critical components or systems through the status signals from their local protection systems (LPS) or their dedicated MP systems. Critical components or systems are the ones that might cause beam-induced damage in case of failure (e.g., magnets or cavities).

– Use information from the beam monitoring systems to prevent or mitigate beam-induced damage. For example, in the case that a steerer magnet is not properly operated, the beam can be deflected, but a beam position monitor can detect this and trigger a beam stop before beam losses occur. BCMs or beam loss monitors (BLMs) can be used to detect beam losses. Beam monitoring systems can trigger a beam stop in case critical beam parameters are being detected.

– Ensure that beam parameters are within the limits defined by the configured beam mode (BM) and that the beam reaches the configured BD.

– Define procedures (e.g., for ramp up in beam power), restrictions in the control system (e.g., steerers maximum configurable current), or configuration control (e.g., firmware versions for critical electronic devices).

4.5.3 Interplay of MP systems

Figures 3 and 4 show the main types of systems that can generate the so-called beam interlocks and how they are connected to the BIS. The BIS consists of the MPSs and the fast beam interlock system (FBIS). The BIS is responsible for interpreting inputs and deciding if, based on the BM and BD, the beam has to be stopped or not. Finally, the FBIS requests the different actuators to trigger a beam stop. Since there are different ways to trigger the actuators, the status of the actuators is used to escalate in case of a failed trigger.

Systems that connect to the FBIS might generate ready signals (where the FBIS does not require an external reset) or beam permits (where the FBIS requires external reset). The decision of generating either a beam permit or a ready signal for each system might change through different phases of the project. These signals are grouped here under the umbrella name interlock.

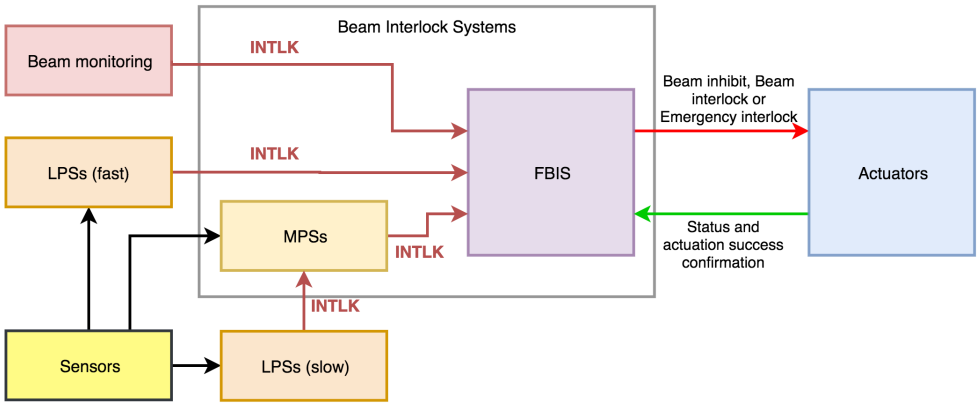


Fig. 3: Interaction of systems involved in achieving or maintaining the protected state of the machine

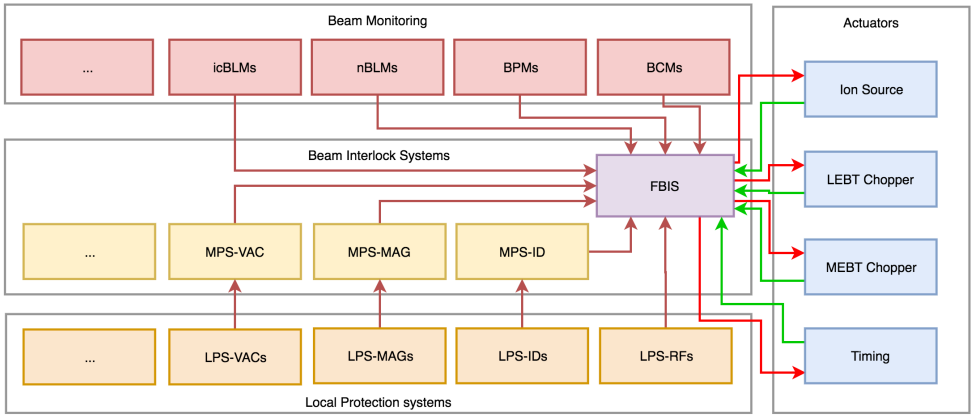


Fig. 4: Examples of systems involved in achieving or maintaining the protected state of the machine and their interaction.

5 A functional protection method and its lifecycle

5.1 The need for a functional protection method

In the field of safety engineering, formal processes for ensuring the correct level of safety are required and there are several existing standards tailored for the operation of large facilities, see Refs.[12−15]. Relevant safety systems and their product lifecycle claim compliance with a safety standard and culminate in a certification process, which is necessary to receive the green light for commencing operation. When protecting hardware systems and equipment from damage, there is no legal obligation to follow any international or national standard. Still, some facilities apply a formalized process for ensuring a protected system has been implemented, see Ref.[16] and Ref.[17]. In parallel with pushing the technology of modern research facilities forward, the demands on reliability and availability of the equipment increase as well.

This also goes for ESS. Machine protection at ESS has been identified as an important driver for successfully reaching the availability goals of the facility. The MP strategy is to identify and

analyse systems and devices that play a role in this goal and adapt their functional behaviour accordingly. Hence it was required to develop a robust method that allows management and analysis of identified risks in the context of availability driven machine protection.

Due to these demanding requirements, it is suitable to put in place a method that efficiently deals with a vast number of interacting components and systems, but at the same time incorporates flexibility and resilience in order to meet the challenges in changing project scopes, which are typical for the design and construction of complex research facilities. Traditional safety methods follow a classic waterfall progression where each specific item is dealt with before moving to the next, see Ref. [7]. While this makes sure that the method is applied in such a way that the risks are accurately handled, it would be difficult to enforce alongside the actual development of the systems and components themselves. Machine protection, as opposed to safety, needs to be intertwined with the systems themselves and often occurs on the system level rather than as a standalone.

The functional protection method, described here, has been developed at ESS and is an abductive, holistic approach that deals with the various hazards and protection-relevant subsystems, and derives a set of protection functions that treat risks in an appropriate and reliable way. The method of defining protection functions can be applied at any stage during the facility lifecycle, being it the design, commissioning, or operational phases.

5.2 What is functional protection?

Functional protection is a technical risk management method suitable for application to a SoS or other complex system. The method was developed at ESS together with the team of functional safety experts from Zurich University of Applied Sciences (ZHAW), Winterthur, Switzerland and can be integrated into the design, early commissioning, and operations phases of accelerator driven facilities to enhance their reliability and availability, see Ref.[18]. The risk management method for MP-SoS follows the applicable ISO standards 31000 and 16085. These standards outline the process of establishing a risk management context followed by identifying, analysing, evaluating, and treating the risks in a traceable and transparent way. Additionally, the IEC 61508 [12] and 61511 [13] standards for functional safety of electric, electronic, and programmable electronic (E/E/PE) systems are guiding the lifecycle approach as well as the classification of protective functions. This places ESS risk management within a robust framework that allows for efficient and purposeful analyses and accurate treatment of the technical risks associated with running a complex high power accelerator facility. The method is currently applied at ESS and builds on close collaboration between system owners and risk analysts, see Ref. [11].

5.3 The functional protection lifecycle

The lifecycle is shown in Fig. 7. It covers the concept and scope definition, analysis, as well as the integration, specification, design, implementation, installation, testing, operating, and adjusting phases.

Three interconnected teams are identified as vital for the functional protection lifecycle progress and execution of MP-SoS related work, see Fig. 8.

– Protection analysis team (PAT).

– Integrated protection team (IPT).

– Implementation and design team (IDT).

The three groups have continuous interaction and follow ups during the lifecycle of the system. At ESS the PAT and IPT are fixed but the IDT varies depending on the analysed system (work package) [19]. The teams are involved in all lifecycle steps but are the mainly responsible for certain parts in the lifecycle, which is described in more detail in the following.

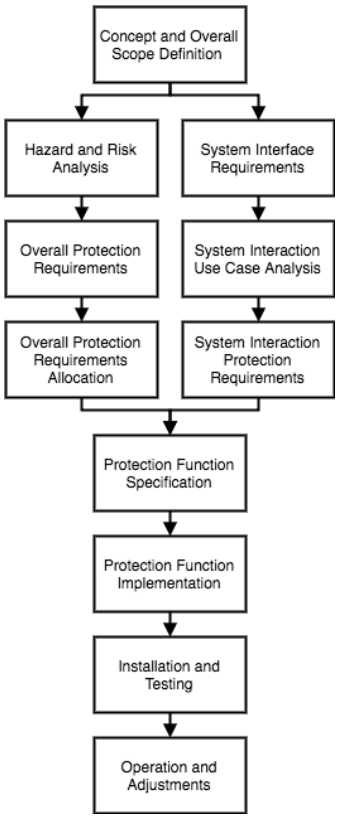


Fig. 5: The functional protection lifecycle. All steps of the process are iterative and might need to be revisited multiple times.

5.3.1 Concept and overall scope definition

The functional protection lifecycle starts by defining a scope and concept for a specific system that plays a role for machine protection.

This step defines the equipment to be protected (EUC), its environment, and what goals are to be met for its protection and availability. It is also useful to specify the types of damage events and hazards that are to be studied in the next step of the lifecycle, in order to create a clear boundary for the analysis. Like this one defines the systems and devices to be included in the analysis, their environment (where appropriate), the damage sources to be considered (e.g., electrical, mechanical, thermal, radiation), and the operational modes where the damage can occur.

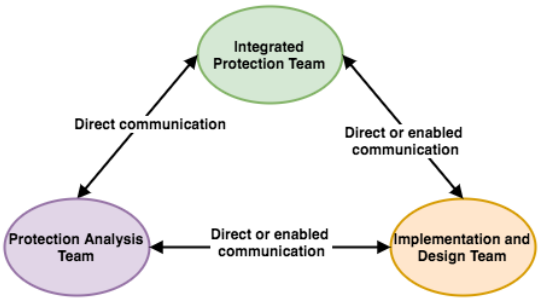


Fig. 6: The three groups and their interactions as identified by the functional protection method

5.3.2 PAT and its role

The PAT is in charge of translating global ESS requirements into manageable requirements and functions to be implemented into the hardware and software of the facility. The PAT is responsible for the hazard and risk analysis, the overall protection requirements, and the overall protection requirements allocation.

As the 'start', or root cause, of a damage event is purely subjective [20], attempting to define this is completely left out of the analysis. Instead, the functional protection analysis technique has the goal of identifying how each device can be damaged, as opposed to how the device can damage something else.

It is important that the teams involved in functional protection are aware of and consistent in the concept and scope of the analysis and risk management. This is primarily done in the risk identification, where the systems and devices that are agreed to be involved are documented.

5.3.2.1 Risk management process

The risk management process for MP at ESS identifies and analyses the damage events that are to be prevented or mitigated by machine protection. However, it is still the responsibility of the system owners to design robust and reliable systems in line with the ESS requirements. Due to this some of the systems and equipment contain certain protection barriers of their own, categorized as other risk reduction measures in the MP risk management framework, in accordance with the IEC 61508 standard. The remainder of the protection functionality is what is required to be carried out by MP-specific PFs, see Ref. [21]. The risk management process applies the same approach, coordinated centrally by the MP personnel, to be taken towards all of the systems and equipment present at ESS. This unifies the analysis efforts and allows for a natural overview of the MP-relevant systems and their role in fulfilling the overall ESS machine protection goals.

5.3.2.2 Damage event and risk ranking

Damage events are ranked according to their consequence for ESS operation. The consequence is defined through two risk matrices, as seen in Figs. 9 and 10. Risk matrix 1 defines a consequence category based on the cost and facility downtime associated with the damage event. Risk matrix 2 then takes this consequence category and associates a tolerable occurrence magnitude (TOM) with this event, where TOM0 is most relaxed and TOM3 has the highest demands [22].

		Downtime				
		< 1 h	1 h – 1 d	1 d – 14 d	14 d – 3 m	> 10 m
Cost	< 0.1 M€	Minor	Moderate	Significant	Significant	Severe
	0.1 M€ – 1 M€	Moderate	Moderate	Significant	Significant	Severe
	1 M€ – 5 M€	Significant	Significant	Significant	Severe	Severe
	> 5 M€	Severe	Severe	Severe	Severe	Severe

Fig. 7: Risk matrix 1, defining the risk category based on downtime and cost of a damage event



Fig. 8: Risk matrix 2, defining the tolerable occurrence magnitude for a damage event of a specified risk category.

Each level corresponds to an order of magnitude for tolerable mean time between occurrences of the damage event. In the case of ESS, TOM0 is once every 5 years, TOM1 every 50 years, TOM2 every 500 years, and TOM3 every 5000 years. The executive aim of the following risk management process can then be viewed as 'moving' damage events that are outside of the TOM into the tolerable region by defining PFs that mitigate the risks.

Identifying and analysing each damage event through the two risk matrices allows for the addressing of each hazard that could lead to this damage event. The hazards are defined top-down, meaning that the hazards that directly lead to a damage event, need to be defined first. Based on the expected occurrence (EO) of a hazard, each hazard frequency is estimated qualitatively by the PAT and IDT. These estimations are named EO0 (normal operations), EO1 (lifetime of facility), and EO2 (unexpected), and are specifically used to avoid unverified expert opinion on the probabilities of hazards, but at the same applying useful pragmatism. Normal operations are events within the specified and expected conditions and restrictions of the system, including all operating conditions; lifetime of facility refers to events that are outside the specified conditions and restrictions, but can be expected during the facility's lifetime, also including frequent disruptions that are overcome by the control systems; and unanticipated implies events outside the specified conditions and restrictions which are not expected to occur during the facility's lifetime.

5.3.2.3 Analysis process

Each hazard then receives one corresponding overall protection function (OPF), which is a generic and technology-free set of objectives compiled and associated with each system, 'negating' the associated hazard. The OPF is in itself only an intermediate step in defining the PF, but is useful in framing the protection efforts and should be checked for its objective in the concept and overall scope definition. The OPF also receives a functional integrity magnitude (FIM), which is the difference of the damage event's TOM and the EO level. That is, FIM=TOM−EO. As an example, the OPF for a TOM3 damage event and an EO1 hazard receives a FIM=3−1=2. The EO level thus corresponds to a reduction in the (occurrence) order of magnitude that needs to be treated, where 0 is no reduction, 1 is a reduction of one order of magnitude (10), and 2 is a reduction of two orders of magnitude (100). This way, the OPFs can be traced back to the corresponding damage event to verify that the TOM is fulfilled by the OPFs, before proceeding to the final step below.

The OPFs are then subjected to audit by the PAT together with the system experts, in order to derive technology-specific PFs, containing the sensors that monitor the hazard, the logic elements that take the decision on what action is required, and the actuators that carry out this action. The hierarchical analysis flow is seen in Fig. 11, starting at the system level. In addition, the PFs include a timing requirement for how quickly the PF needs to be carried out, and a PIL that gives requirements on the quality of the PF [22], [12].

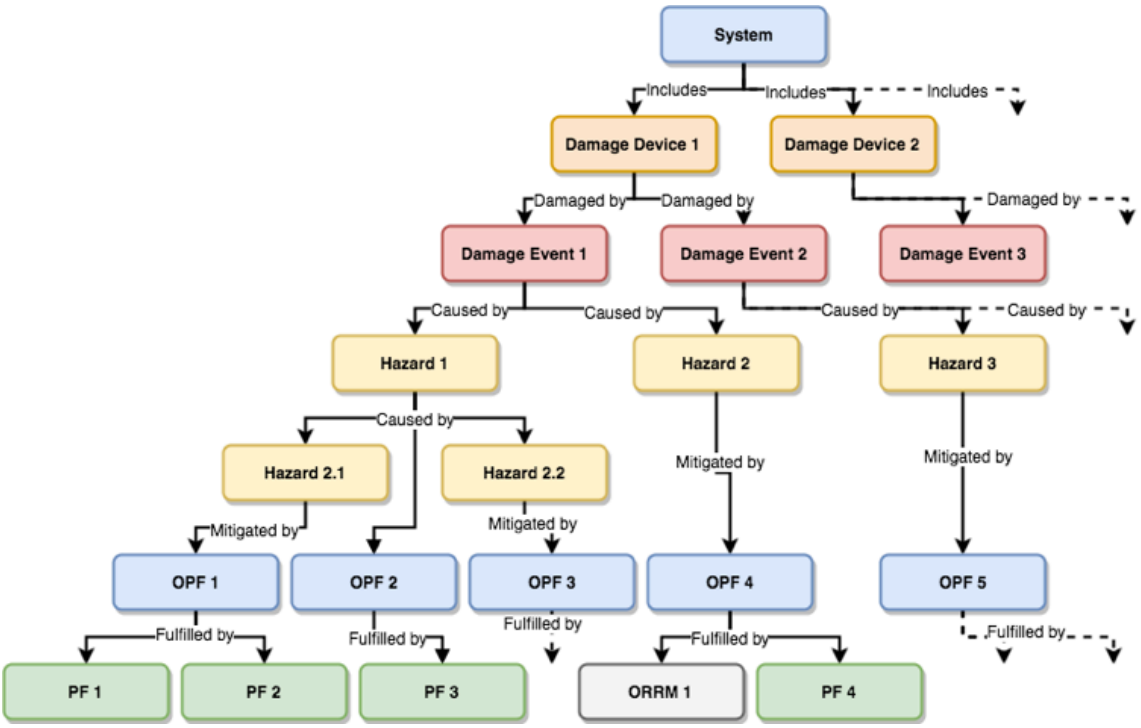


Fig. 9: The Functional protection abductive risk management process, starting at the system level and ending at the level of PFs.

For all of the identified PFs, the MP-related sensor signals are sent to a system specific logic element. This logic element performs the sensor data analysis and further distributes the signal to the ESS fast beam interlock system (FBIS) when a beam stop is required to prevent or mitigate a damage event. The FBIS receives inputs from the entire ESS facility and makes the final decision on when to stop the proton beam. If a beam stop is required, the signal is sent to the actuator systems. There are four systems at ESS that are used to stop beam operation. They are the ion source, the low energy beam transfer (LEBT) chopper, the MEBT chopper and the timing system.

All of the information and risk management process steps are required to be traceable and readily available for all relevant parties. This allows for a continuous workflow where needed, and all associated parts can follow and, where necessary, contribute to the analysis process. Once a set of PFs has finished its internal iterations and suitability checks, it is documented and uploaded to the official ESS document management system for approval by the ESS machine protection committee (MPC).

5.3.3 IPT and its role

The IPT analyses interfaces between protection-related systems and the rest of the facility. The analysis and integration work is mainly done in parallel, but it is beneficial to have access to a draft of the overall protection requirements and an idea of the architecture of the involved system (functional architecture) before starting the integration work. The IPT defines system interface requirements, coordinates the system interaction use case analysis and defines system interaction protection and behavioural requirements.

To obtain the system interaction model, behavioural requirements and to allocate PFs, system interaction use case workshops are held. The feasibility of different system architectures and PF implementations are discussed and simulated by going through foreseen operational sequences, use cases. The different architectures and use cases are documented using enterprise architect. Before the

work of the IPT starts, a representative for each protection relevant system has to be identified. The representative could be the designer, owner, or contact person for that system.

The representatives are then gathered together with the IPT during a series of workshops. During the workshops, the architecture, functionality, and possible interfaces of each system are discussed. Based on the concept and overall scope definition, certain PFs have to be performed. Depending on the level of progress of the development, the design and interfaces are more or less constrained. Typically, a more developed system is less flexible. This could cause difficulties if certain functions or hardware have not already been foreseen. When everybody has a clear view of the involved systems architecture, the allocation of PFs is discussed. Questions such as, 'Which sensors are available?', 'Where should the logic be implemented?', and 'How is the information transmitted?' are asked and answered.

A certain architecture and interface set is suggested and fixed for the coming system interaction use cases. The purpose of the use cases is to examine and document the intended interaction between the involved systems from a machine protection perspective. The goal is to derive appropriate and robust solutions for the signal types and interfaces. The use cases are one way to specify interfaces and requirements (especially with respect to PFs), however they are not the unique source. The parallel work of PAT also results in a set of PFs and requirements.

The use cases are set up to follow scenarios which will appear during regular machine operation such as a change of BD or a rearm after an interlock. Besides the 'normal flow', which documents the operational scenarios as they are expected to happen, so-called 'alternative flows' are examined. Alternative flows document how the systems react if the scenario deviates from the normal flow. Typically, this is due to a fault or failure of a system or component.

Each action or information exchange failure would cause a different sequence of events. It is not feasible to go through every single alternative flow due to time constraints. Instead, a set of alternative flows is selected based on the events that are most likely to fail or events that seem to have the largest damage potential. The selection of alternative flows is done by estimations of the involved experts. If unacceptable behaviour of the systems is observed due to a failure in the normal flow, it has to be adjusted. The normal flow should be adjusted to increase the robustness and decrease the vulnerability by changing the normal operational sequence.

The system interaction protection requirements will, similar to the functional protection analysis technique, propose a set of PFs. The requirements derived by the IPT are behavioural requirements. They are defined on the system level and based on the interfaces and interactions between protection-related and other systems. This analysis path completes the picture of the protection requirements in a way that is not possible through the damage-based analysis alone.

One or more use cases are selected and played through with the selected architecture. During the use case studies, it may be identified that systems perform well as standalone systems but that the signal exchange with other systems is flawed or causes a different action than expected. This analysis may also derive additional interfaces that are then added to the system interface requirements.

During each of the lifecycle phases previously defined properties might need to be adjusted. This involves the architecture, the allocation of PFs, system behaviour and interaction.

5.3.4 PF specification

The two tracks that merge into the PF specification in Fig. 7 apply two different approaches to protection. While the leftmost track uses a top-down, abductive analysis that yields requirements for the quality of the PFs, the rightmost track applies inductive reasoning to verify that possible failure modes, including undesired control actions, are handled correctly. This reasoning viewpoint is seen

graphically in Fig. A, where the abductive analysis is ‘top-down’ oriented and the inductive analysis is ‘bottom-up’.

5.3.5 IDT and its role

When the work of the two teams has reached a satisfying maturity, the IDT starts the implementation work. Finally, the IDT undertake the implementation of the requirements and serve as system experts in the discussions on possible and relevant designs and functionalities. The IDT plays an important role in the design and analysis phase by contributing with experience and judging the feasibility of different designs.

The installation of the protection-related systems is detailed in the system requirement specification for each system. Before installation of protection-relevant equipment, separate tests (commonly referred to as factory acceptance tests) are performed in the laboratories or factories where the equipment for the functions is assembled. These tests need to verify the functionality and the timing requirements of the systems, as well as their correct interface to the input and output systems. Once the complete functions are installed, a complete test of the functionality (site acceptance test) is carried out to verify that the requirements are met.

It is not possible to practically verify every aspect of a control or protection system before the facility is taken into operation, and therefore the first years of the facility operation (the initial operation) need to be dedicated to continuous follow ups and adjustments of the equipment and its behaviour. The theoretical numbers based on analysis assumptions do not always agree with what is observed in the first stages of the facility’s operational lifetime, but become more relevant as the facility and its organization and procedures mature into a steady state. This also displays the importance of an up-to-date risk record to be traceable and easily available.

6 Summary and outlook

As discussed in this paper, modern particle accelerators are increasingly complex and require robust strategies for them to be operated without damage and with the desired availability. Due to their complexity and vast number of systems and sub-systems, it is argued that these facilities cannot be analysed and treated accurately by applying traditional reliability approaches alone. Instead they need a system of systems mentality as well as recognition of the emergent and often incomplete descriptions of its complex properties.

The functional protection method, presented in chapter 5 aims at being the link between complex systems and acceptable damage risk. Following functional safety standards for protection purposes and connecting the lifecycle steps to proven-in-use risk management and analysis methods would give partial confidence in this link.

However, it is important to keep in mind that the machine protection objectives can only be achieved if a large number of systems, developed by different groups and divisions, interact in a well-orchestrated way.

Acknowledgement

We wish to thank our colleagues from CERN, SNS, ZHAW, and ESS for enlightening comments on the present topic.

References

[1] R. Schmidt, Proceedings of the 2014 Joint International Accelerator School: Beam loss and Accelerator Protection, Newport Beach, United States, 5–14 November 2014, edited by R. Schmidt, CERN 2016-002, https://doi.org/10.5170/CERN-2016-002.

[2] R. Schmidt, in Proceedings of the CAS-CERN Accelerator School: Advanced Accelerator Physics, Trondheim, Norway, 18-29 August 2013, edited by W. Herr, CERN-2014-009 (CERN, Geneva, 2014), pp. 221-244, https://doi.org/10.5170/CERN-2014-009.221.

[3] W. Blokland and C. Peters, A new differential and errant beam current monitor for the SNS accelerator, 2nd International Bean Instrumentation Conf., Oxford, UK, 16−19 September 2013.

[4] N.V. Mokhov and W. Chou, 7th ICFA workshop on High intensity high brightness hadron beams, USA, 1999.

[5] R. Schmidt *et al*., *New J. Phys.* **8** (2006) 290, https://doi.org/10.1088/1367-2630/8/11/290.

[6] T.Friedrich, C. Hilbes and A. Nordt, Systems of systems engineering for particle accelerator based research facilities: A case study on engineering machine protection, 2017, https://doi.org/10.1109/SYSCON.2017.7934806.

[7] A.P. Sage and C.D. Cuppan, *Inf. Knowl. Syst. Manag*., **2** (2001) 325.

[8] D. DeLaurentis and R.K. Callaway, *Rev. Policy Res*. **21** (2004) 829, https://doi.org/10.1111/j.1541-1338.2004.00111.x.

[9] M.W. Maier, *Syst. Eng*., **1** (1998) 267, https://doi.org/10.1002/(SICI)1520-6858(1998)1:4<267::AID-SYS3>3.0.CO;2-D.

[10] E. Bargalló, "ESS reliability and availability requirements," ESS Document (ESS-0008886), 2015.

[11] R. Andersson, S. Kövecses, E. Bargalló and A. Nordt, Challenges in technical risk management for high-power accelerators, ICANSXXII, 2017.

[12] IEC 61508:2010: Functional safety of electrical/electronic/programmable electronic safety-related systems. 2010.

[13] IEC 61511:2003: Functional safety - Safety instrumented systems for the process industry sector. 2004.

[14] IEC 61513:2011: Nuclear power plants - Instrumentation and control important to safety. 2011.

[15] IEC 62061:2015: Safety of machinery - Functional safety of electrical, electronic and programmable electronic control systems, 2015.

[16] M. Kwiatkowski, Methods for the application of programmable logic devices in electronic protection systems for high energy particle accelerators, 2013.

[17] C. Sibley, Machine protection strategies for high power accelerators, Proc. 2003 Bipolar/BiCMOS Circuits Technol. Meet. (IEEE Cat No03CH37440) 2003, https://doi.org/10.1109/PAC.2003.1288989..

[18] R. Andersson, University of Oslo, Oslo, 2017, https://www.duo.uio.no/handle/10852/590699.

[19] S. Kövecses, R. Andersson, A. Nordt, E. Bargalló and M. Rejzek, Applying the functional system interaction process at ESS, ICALEPS, 2017.

[20] N. Leveson, An STPA primer, Version 1, 2013.

[21] C. Hilbes and A. Nordt, Machine protection - systems requirements and architectural framework, ESS Internal Document (ESS-0057251), 2015.

[22] R. Andersson, E. Bargalló, C. Hilbes and A. Nordt, Machine protection risk management process, ESS Internal Document (ESS-0095000), 2017.

Proceedings of the CAS–CERN Accelerator School: Beam Injection, Extraction and Transfer, Erice, Italy, 10-19 March 2017, edited by B. Holzer, CERN Yellow Reports: School Proceedings, Vol. 5/2018, CERN-2018-008-SP (CERN, Geneva, 2018)

Exotic Injection and Extraction Methods

B. Goddard
CERN, Geneva, Switzerland

Abstract
This paper is intended to give a flavour of the creativity that can be applied to injection and extraction, in terms of concepts, systems, challenges, and solutions that are somewhat unconventional. A dazzling array of techniques and technologies were already described in great detail in the other papers; 'conventional' is, therefore, a very relative or subjective term. In this paper, some of the more unusual or creative methods for filling or emptying accelerators are presented and discussed.

Keywords
CAS; injection; extraction; kickers; septa; losses, emittance; laser; muons.

1 Introduction

The domain of beam transfer lends itself to the development of creative solutions to the problems and constraints that are particular to the processes of injecting and extracting particles from accelerators. An array of fantastic beasts were already described in the CAS course—some of the topics already seen include:

- electrostatic spiral inflectors for cyclotrons;
- golf-club longitudinal injection for small-aperture storage rings;
- multipole kicker for light-source top-up injection;
- corkscrew gantry for medical hadron beam delivery;
- slip-stacking and phase-space painting;
- magnetic splitting.

Given the richness already present in the mainstream techniques, in widespread use and based on standard systems, it is justifiable to wonder why even more exotic methods are required. Some of the rationale is to make possible what would otherwise be inaccessible, where no conventional solution exists. In addition, in the real, imperfect world, many limitations are present, which arise from the laws of physics, engineering limits, safety constraints, economic imperatives, etc. Specific factors that come into this category of limitation are beam stability, lifetime of unstable particles, physical space available, maximum electric or magnetic field, beam loss, and radiation dose.

Sometimes the progress is driven by the imagination of the accelerator builders—the new ideas are so elegant that they allow accelerators to move into new regimes of accelerator performance.

In this paper, the exotic technologies and methods described are a massless septum for extraction or for injection failure mitigation, an open C-core extraction kicker, injection into a muon decay ring, extraction with bent crystals, and laser stripping H^- injection. There are many, many other fascinating topics in the field, illustrating the interplay between hardware, technology, and accelerator physics—students are encouraged to explore the available literature, some of which is listed in the references section.

2 Massless septum

To an accelerator physicist charged with designing an injection or extraction system, a massless septum sounds very exciting. Ideally, as for any septum, a narrow septum width is desirable. The rather large

2519-8041 – © CERN, 2018. Published by CERN under the Creative Common Attribution CC BY 4.0 Licence.

https://doi.org/10.23730/CYRSP-2018-005.553

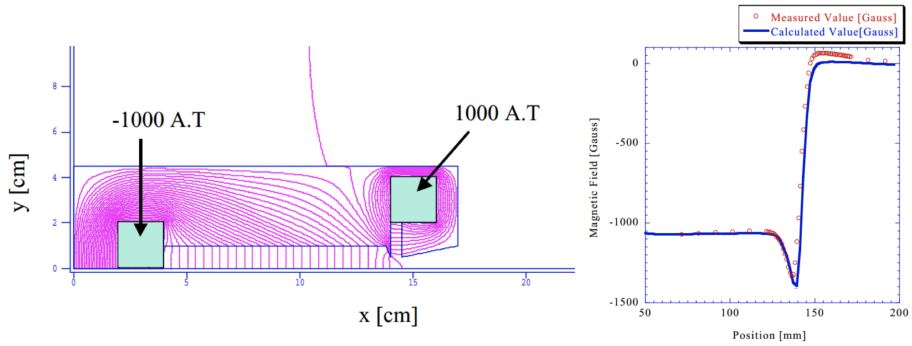


Fig. 1: Cross-section and field in septum region for a basic massless septum design

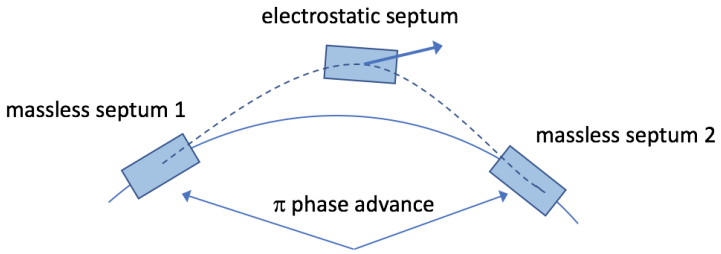


Fig. 2: Massless septum deployment for multiturn extraction loss reduction

practically attainable septum thickness constrains its use to a subset of applications, but the device remains a fascinating technological concept. The relatively thick 'septum' width is of the order of the gap height (Fig. 1)—this pushes its useful application domain to higher beam energies or electron beams, where small gaps are possible for the tiny physical emittances, or to low beam energies, where large bending angles can be achieved to clear the septum width.

The most attractive potential applications are those difficult systems for injection and extraction processes where beam interception at the septum blade is problematic, for example to reduce beam loss in slow or continuous extraction over many turns, or to prevent damage from mis-steered beams in injection or extraction systems of high-energy machines.

The massless septum has been deployed in several accelerators. For the 150 MeV Kyoto FFAG [1], a massless septum was used to extract a high-power CW beam with low losses, a concept which was has also been studied for other similar high-power machines. For the Kyoto FFAG, the massless septum was deployed as a closed bump around the electrostatic septum (ES), which allowed a large reduction in the density of particles impacting the ES. The FFAG is a hybrid of cyclotron and synchrotron, with quasi-continuous high current and adjacent turns that are very close together—requiring a thin septum to minimize extraction losses. A pair of massless septa can reduce the beam losses on the septum, as illustrated schematically in Fig. 2.

A version of this concept has also been explored in more detail in potential application to ES beam-loss reduction of a conventional synchrotron slow extraction [2]. The beam distribution is 'stretched' by a massless septum at $\pi/2$ phase advance upstream of the ES to reduce density. The septum width (where the field increases from zero to its nominal value) is crucial in the achievable loss reduction (Fig. 3). Only a single massless septum is required, but the disadvantage with this arrangement is that a large physical anode–cathode gap is needed for the ES to accommodate the normal and stretched part of the spiral step, which would limit the attainable ES field.

This limitation could be improved by shifting the ES position outwards to cut the stretched part of the distribution at the other end, with the perturbation closed by a second massless septum a further $\pi/2$

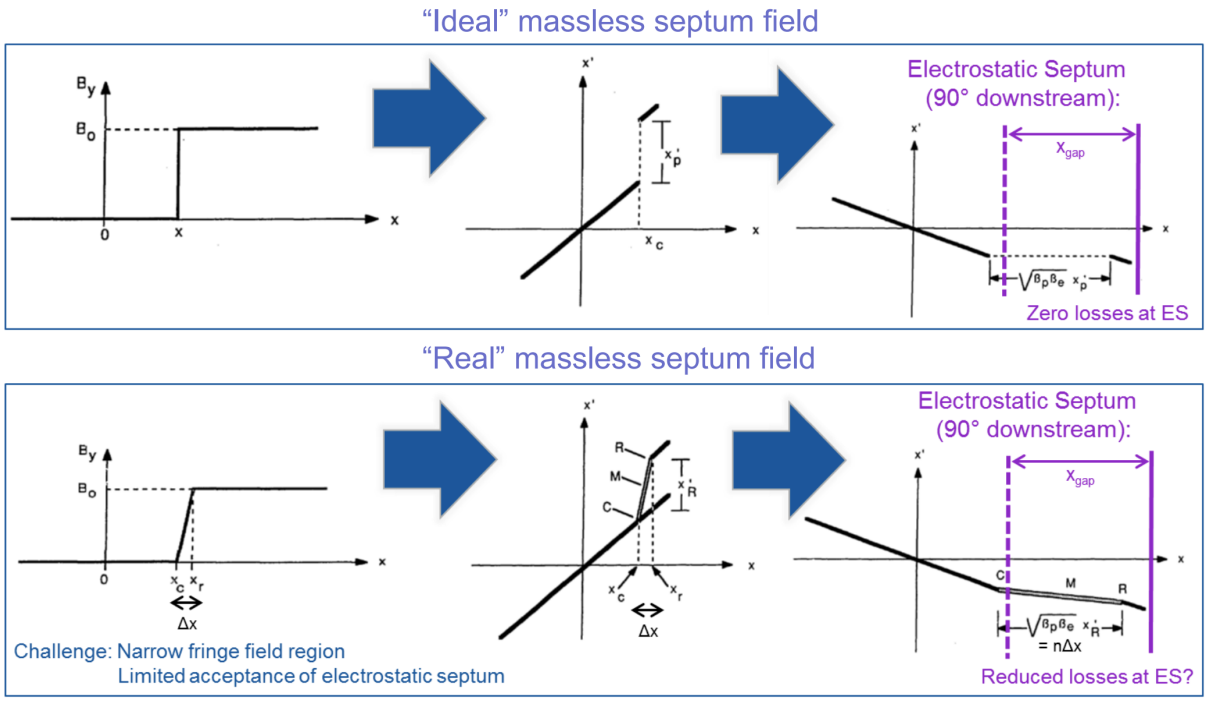


Fig. 3: Phase-space representation of massless septum used to reduce particle density at the electrostatic septum. Top: the case with a perfect zero-width septum. Bottom: a more realistic field distribution.

downstream. This maintains the separatrix size inside the ES gap, but requires a larger circulating beam aperture between the massless septa to accommodate the stretched separatrix. This scheme is illustrated in Fig. 4.

The massless septum concept has also been investigated for low-energy beam transport, for example, in the accelerator-driven neutron source FRANZ [3]. The low-energy proton beam at 120 keV, 50 mA, needs to be transported with very low losses. A novel massless septum system has been designed after the chopper for the beam separation to the dump (Fig. 5), which consists of a C-magnet with optimized pole shapes, which will extract the beam with minimal losses, and a magnetic tube to shield the beam from the fringe field of the dipole.

3 Open C-core kicker

It can be difficult to define extraction kicker magnets for high-energy beams, owing to conflicting constraints. One the one hand, high kick strength is always preferable, together with fast rise time, and high beam energy means a lot of $\int B.\mathrm{d}l$, which pushes the system towards small gaps in the non-kicked plane, high current, and long magnetic length. On the other hand, beam coupling impedance is often a performance concern, and this pushes the system towards larger gaps, shorter installed length, and shielding of ferrites, which can increase the gap and rise time.

Crucially, for a proton or ion accelerator, the beam at injection energy is generally significantly physically larger than at extraction, owing to the conservation of the normalized transverse emittance. A conventional full-aperture extraction kicker must be large enough to accept the injected beam. For large energy swings, this can mean a lot of 'wasted' aperture, as the physical beam size scales with $(\beta\gamma)^{0.5}$, plus some extra millimetres of aperture for injection oscillations. An interesting kicker concept is based on this fact, since the kick strength is only needed at the extraction energy, where the beam is small. The concept uses an open C-type kicker with a small vertical gap (Fig. 6), and a fast bumper system to move the beam into the kicker gap shortly before extraction, where it would remain for only a few tens

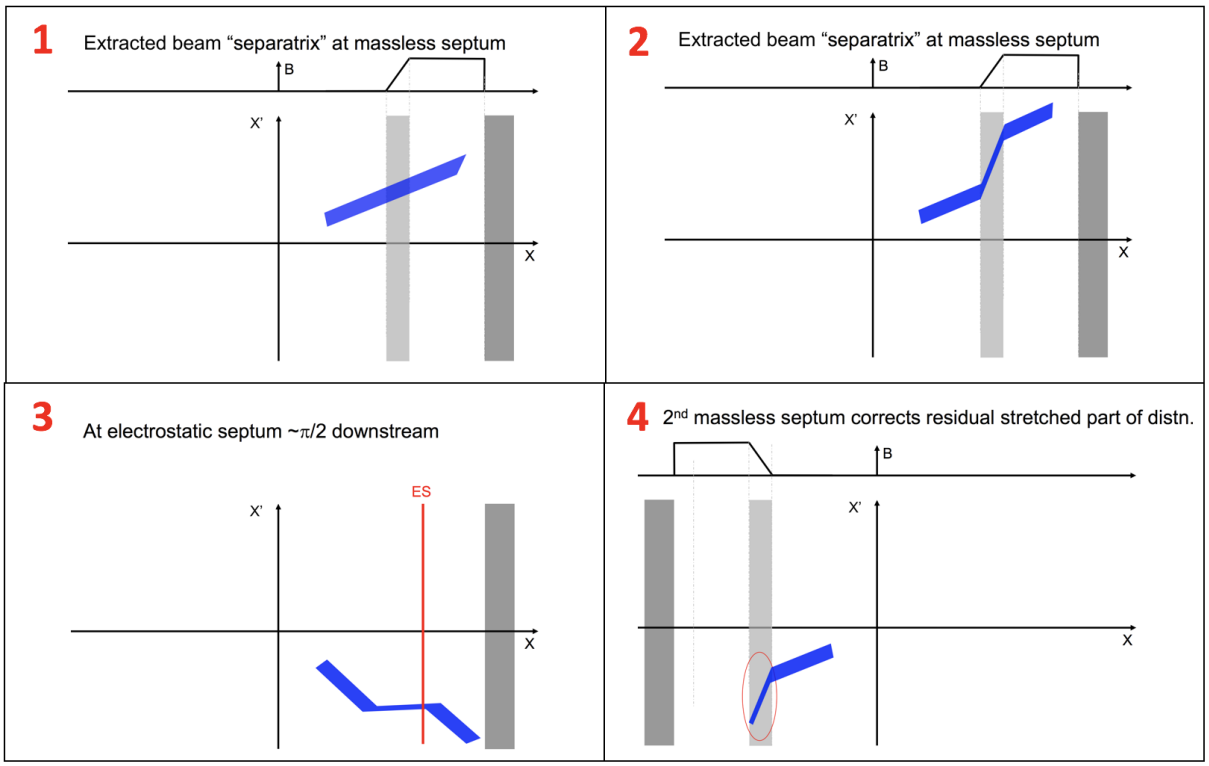


Fig. 4: Pair of massless septa used to reduce particle density at electrostatic septum, with the second septum refolding the separatrix back.

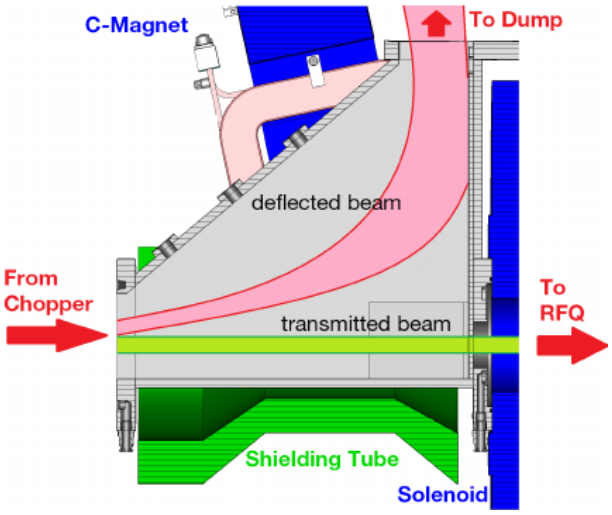


Fig. 5: Massless septum for 120 MeV proton beam separation for FRANZ chopper system

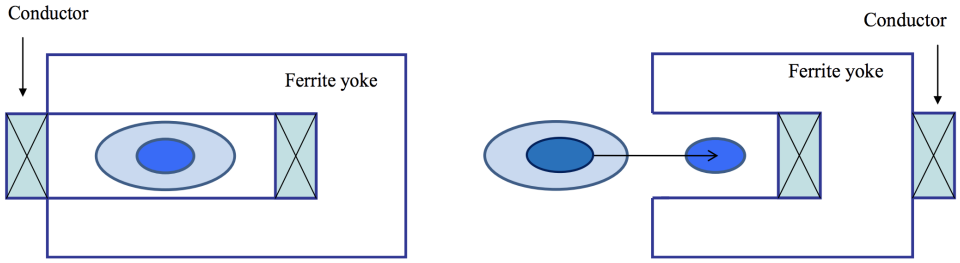


Fig. 6: Full-aperture (left) and open C-core (right) extraction kicker cross-section, where a fast bump is used to move the beam into the gap. The vertical gap height is determined by the injected beam size for the full-aperture system, and by the extracted beam size for the open C-core system.

of milliseconds.

Having the beams present in the kicker gap for a short time only, and then at high energy, would be a big advantage with respect to beam stability arising from the kicker beam coupling impedance. Similarly, kicker heating from the real part of the beam coupling impedance will be much less of an issue if the high coupling impedance seen by the beam is only seen for a very short time, resulting in a very small averaged power deposition. In addition, since a smaller vertical gap is required (for a given current), a shorter installed kicker length will be required, which helps reduce the overall beam coupling impedance. Overall, the kicker system can be much more compact, at the cost of needing an extra closed orbit bumper system. This concept has actually been used for the AGS beam extraction system at BNL [4].

4 Stochastic muon decay ring injection

Intense circulating muon beams are of great interest as a controlled source of neutrinos for, e.g., light sterile neutrino searches or precise cross-section measurements. Muons are unstable leptons with negative or positive charge (antimuons), 200 times the electron mass and a lifetime at rest of 2.2 µs. The muon decays into an electron, an electron anti-neutrino, and a muon neutrino:

$$\begin{aligned} \mu^- &\rightarrow \mathrm{e}^- + \overline{\nu}_\mathrm{e} + \nu_\mu \,, \\ \mu^+ &\rightarrow \mathrm{e}^+ + \nu_\mathrm{e} + \overline{\nu}_\mu \,. \end{aligned} \tag{1}$$

One concept to store muons proposes a 400 m circumference racetrack decay ring filled with muons [5], to produce intense neutrino beams. The muon lifetime is about 60 turns in the ring. The ring would be oriented such that the neutrinos produced from the decay straight are directed towards a detector, Fig. 7. The beam from the decay straight is preferred for the experiment, as it is uncontaminated by muon anti-neutrinos (for the case of a μ^- ring).

The muons are produced from decay of ∼5 GeV/c pions (themselves composed of an up and down quark–antiquark pair with a lifetime at rest of only 26 ns). The pions are produced from directing a high-power 10–100 GeV proton beam on a target. The injection system uses the fact that the 5 GeV/c pions decay to muons with very different momenta—a wide distribution centred on 3.8 GeV/c and ranging from about half the pion energy to almost the full energy. The ring optics and injection layout merges pions and already-circulating muons in a special dipole, to 'defeat' Liouville (a victory that is the ambition of every injection system designer). The injection straight needs to be long, to allow as many pions as possible to decay, with optics that are simultaneously matched for the two very different particle momenta, and the ring needs a huge momentum acceptance to be able to store the resulting muons.

This is shown schematically in Fig. 8. Note that only about 50% of the pions decay in the injection straight and, as they will not be transmitted through the acceptance of the downstream arc, an extraction

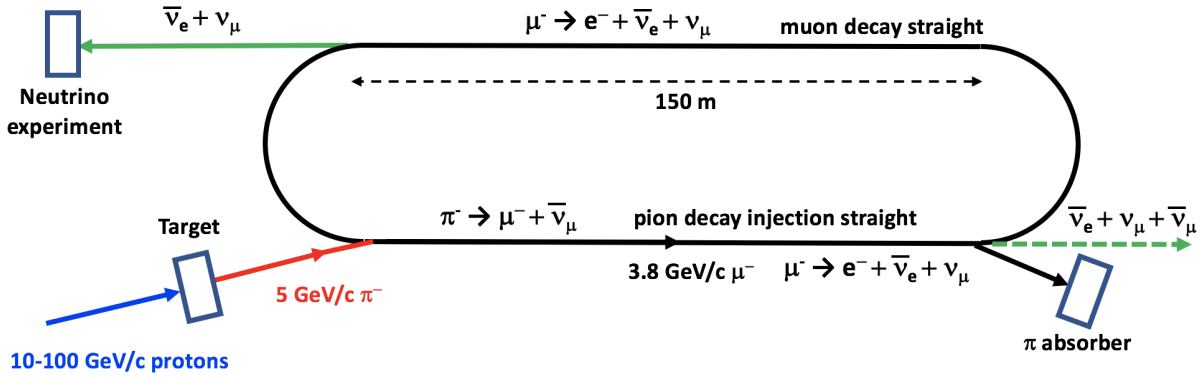


Fig. 7: Concept of muon storage ring to produce high-intensity neutrino beam

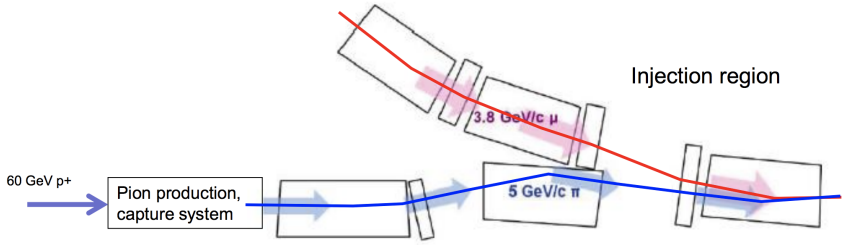


Fig. 8: Stochastic injection of pions in muon storage ring. Pions and muons of very different momenta are merged in the injection dipole.

(separation) dipole and absorber are needed at the end of the straight, essentially a reverse of the injection system. The ring acceptance needs to be about $\pm 10\%$, which is extremely large compared with most accelerators ($\sim \pm 0.1\%$ is more typical).

5 Extraction with bent crystals

In addition to the usual magnetic or electrostatic elements, charged particles can also be deflected by a bent crystal, see e.g., Ref. [6]. The potential wells between the atomic planes confine particles with small enough transverse momentum (Fig. 9); if the crystal is bent with an angle small enough to maintain the confinement, the particles can be bent though the same angle, in a very short longitudinal distance. These confined particles are said to be *channelled*; other distinct effects have been identified in the coherent and incoherent scattering behaviour, as elucidated in Ref. [7]. Channelling gives the biggest deflection, while volume reflection results in the highest efficiency.

Bent crystals are already deployed in high-energy proton accelerators, including heavy-ion collimators for the LHC [8]. One other very interesting potential application is slow extraction of multi-teraelectronvolt protons [9]. Producing a slow-extracted beam from the LHC at 7 TeV or from FCC at 50 TeV will be very challenging, as it is more and more difficult to design a conventional slow extraction system as the beam energy reaches the teraelectronvolt regime, with superconducting rings. A large machine aperture is needed for 1/3 integer resonance, superconducting magnets are very sensitive to beam losses, and the technological limit of $\sim$10 MV/m for electrostatic septa coupled with the extreme beam rigidity means that the length of extraction systems simply becomes unfeasibly large. For example, for FCC the $B\rho$ at 50 TeV is a huge 166,000 Tm. Around 500 m of ES field length would be needed to produce a deflection of the order of 0.1 mrad, which would be almost impossible to accommodate in a

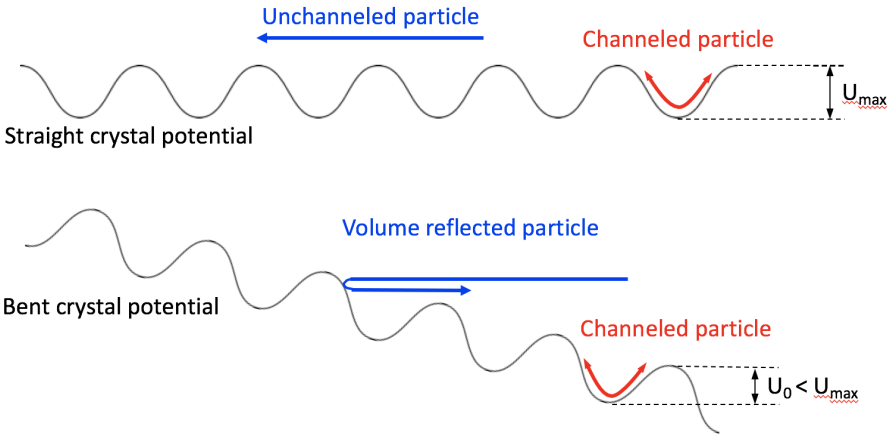


Fig. 9: Potential distribution for straight (top) and bent (bottom) crystal. The channelling probability for an impacting proton depends critically on the transverse momentum, which must be less than the potential well height. Bending the crystal introduces an asymmetry into the potential well (bottom), which reduces the channelling probability.

lattice, or to align in practice, given that the septum is a thin element of <0.1 mm thickness.

A bent crystal with a bending radius R channels a beam of energy pv if $pv/R \simeq 1$ GeV/cm. A condition can be derived for the maximum angle θ_c that can be accepted:

$$\theta_c \leq \sqrt{\frac{2U_{\max}}{pv}} \,. \tag{2}$$

This is simply astounding—taking 7 TeV protons, this gives $R \simeq 70$ m. For comparison, the 8 T LHC dipoles bend the 7 TeV protons with $R \approx 3000$ m. So the bent crystal is as effective as a 330 T magnet. A 2 cm length of bent crystal can deflect 7 TeV protons by 0.3 mrad, which is more than enough to act as the first element in an extraction system. Bent crystals, therefore, may have the potential to replace huge 5–500 m long ES devices needed at 100 GeV to 50 TeV energies, with extremely compact millimetre- or centimetre-long devices.

In addition to the possible application for collimation and extraction elements, crystals have also been proposed [10] as a way to reduce beam loss on a thin ES for conventional slow extraction. If the crystal alignment and position can be maintained through the spill in such a way as to ensure channelling, the low-loss coherent deflection of protons impacting the crystal can be used to shadow the ES and reduce by a large factor the number of particles impacting the ES (Fig. 10).

6 Laser stripping for H^- injection

The H^- ion is a fantastic ion for accelerators. It is charged, so can be accelerated and deflected; it is easy to produce and stable (the e^- binding energies are 0.75 and 13.6 eV); and the weakly-bound e^- can easily be removed with thin foil, allowing the astute designer to 'defeat Liouville' by injecting on top of an existing distribution. However, the interaction processes in the physical stripping foil means that beam loss, foil heating, or damage (Fig. 11), and unwanted emittance growth can all occur. These problems are especially acute in high-energy, high-power machines, such as SNS [11].

Before entering into the details of the laser stripping process, an aside into the dynamics of the H^- ion is needed. An H^- ion passing through a transverse magnetic field B with velocity βc experiences an electric field E in its own rest frame:

$$E = \beta\gamma c \times B \,. \tag{3}$$

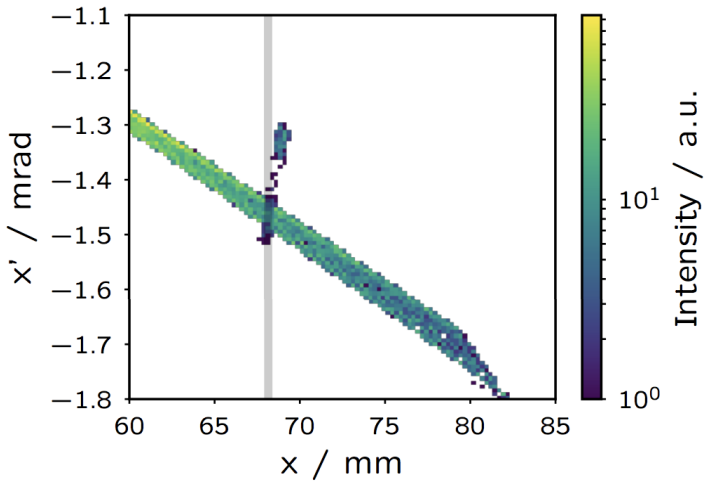


Fig. 10: Transverse phase-space distribution of 1/3 integer slow extraction separatrix at ES location, after coherent channelling by an upstream bent crystal. A reduction in particle density at the ES (vertical line) is clearly visible.

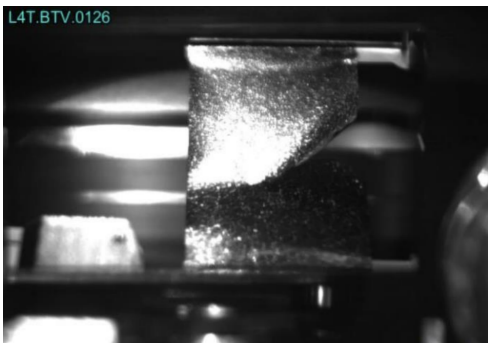


Fig. 11: Damaged carbon stripping foil of the CERN PS booster 160 MeV H^- injection system

This electric field can be strong enough to remove the loosely bound outer electron, which only has a 0.75 eV binding energy. This field stripping (or Lorentz stripping) can lead to uncontrolled beam loss. The effect worsens with increasing beam energy, compounded by the (usual) increase in magnetic field for accelerator elements at higher energy. The fractional beam-loss rate per metre of field can be derived as:

$$\frac{\mathrm{d}F}{\mathrm{d}S} = \frac{B}{A_1} \exp\left(-\frac{A_2}{\beta\gamma cB}\right), \tag{4}$$

where A_1 and A_2 are constants with numerical values 2.47×10^{-6} V s/m and 4.49×10^9 V/m.

Normally, radio-protection considerations limit beam loss to something like 0.11 W/m of beam-loss power, which places a practical upper bound on the maximum allowable magnetic field, which depends very strongly on beam energy. This becomes an issue for injection and transfer line (dipoles, quadrupoles, correctors) at H^- beam energies at or above the 1 GeV level. But this unique effect of Lorentz stripping provides a great opportunity for the creative thinking that has gone into the laser stripping concepts. U The basis for laser stripping is the observation that excited neutral H^0 is easy to field strip in a magnetic field, with the lifetime as a function of field depending on the principle quantum number n of the excited state (Fig. 12). For instance, at 1 GeV, both the ground-state H^- and H^0 in the $n = 3$ excited state are easy to Lorentz strip with a 1–2 T magnetic field. The lifetime is between 10^{-10} and 10^{-9} s, which with $v \simeq c$ means that the electron is removed within a few tens of centimetres of field.

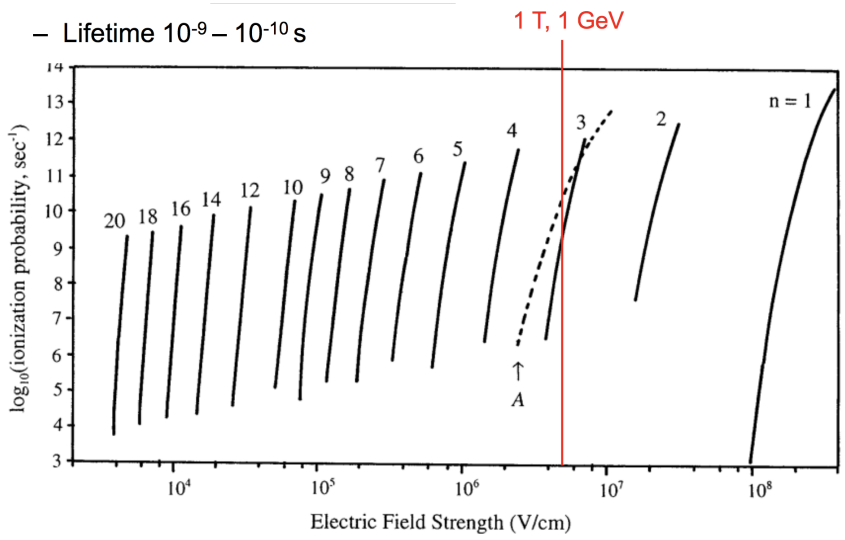


Fig. 12: Ionization probability vs. electric field for different H^0 principle quantum numbers. The electric field associated with a 1 GeV proton in a 1 T field is shown as the red vertical line.

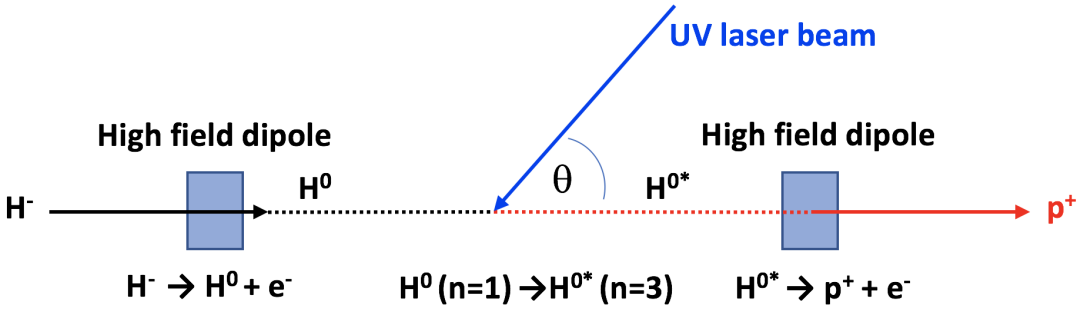


Fig. 13: Three-stage laser stripping scheme, relying on two field-stripping stages with an intermediate resonant excitation of the H^0.

A basic laser stripping scheme [12] is illustrated in Fig. 13. In fact, the laser does *not* strip either electron directly. There are three distinct steps.

1. Removal of loosely bound e^- from H^- with strong magnet ($\sim$2 T). This is in the vertical plane to avoid increasing spread in horizontal angles x'.
2. Resonant excitation of second e^- to high energy level: $n = 1$ to $n = 3$ (or higher) transition, using a high-power tuned laser.
3. Removal of excited electron with another strong magnet ($\sim$2 T).

The injection needs to be accommodated in the real machine layout, and also provide the normal advantages of the H^- injection in terms of phase-space painting over many turns. A schematic layout of an injection system designed for 4 GeV H^- injection with laser stripping is shown in Fig. 14 [13].

The key to the laser stripping is resonant laser excitation of the neutral H^0 ion. Here the hydrogen energy level diagram, Fig. 15, is revealing. For 1 GeV H^-, at least the $n = 3$ level needs to be reached, which requires absorption of a 12.1 eV photon. This is very far into the ultraviolet, at 103 nm, and difficult to produce with the required high laser intensity. But here nature comes to the rescue with a rare example of a free lunch—the relativistic Doppler shift modifies the frequency f_s of the photon in the

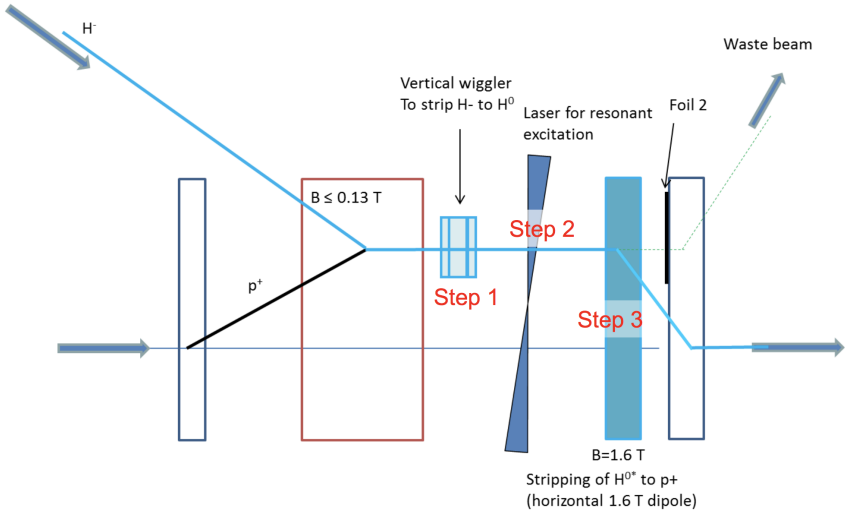


Fig. 14: Injection system designed for 4 GeV H^- injection with laser stripping

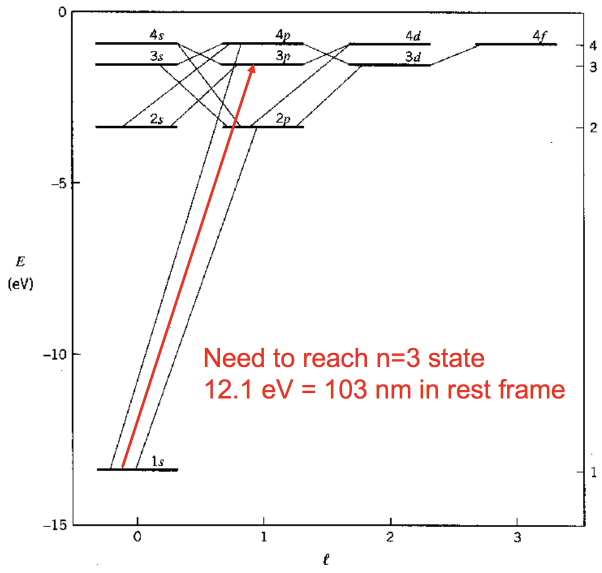


Fig. 15: Neutral hydrogen energy level diagram, including fine-structure splitting

frame of the ion, according to the relativistic parameters and the incident angle θ_0, according to

$$f_s = f_0\gamma(1 + \beta\cos(\theta_0)) \,. \tag{5}$$

For angles approaching head-on, and for high beam energies, this enhancement factor approaches 2γ, so for a 1 GeV proton the resonance to excite the $n = 3$ state can be achieved with 355 nm photons interacting with an angle of about 40° to the protons. This opens another very attractive feature, which is the ability to tune the effective resonant laser frequency by adjusting the interaction angle, allowing the use of standard laser wavelengths for high-power systems, e.g., 532 or 355 nm.

The devil, as always, is in the details, and one of the most difficult problems is the laser power required to saturate the resonance transition. There are several intricate details worthy of explanation, which beautifully illustrate the inventiveness and ingenuity of those who have evolved this concept.

The first problem is to excite all the neutral ions to the higher quantum state. For a two-state quantum system, this can be neatly achieved by adiabatic rapid passage, where the exciting frequency is

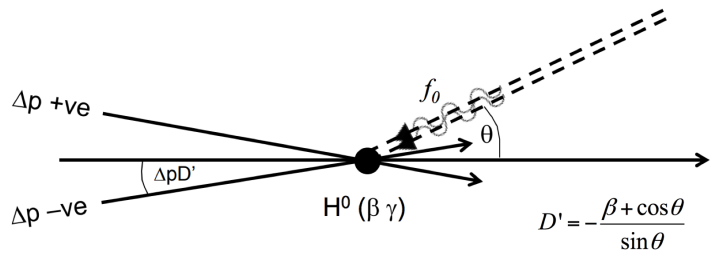


Fig. 16: Dispersion tailoring to cancel Doppler shift at interaction point for laser stripping resonant excitation step, by correlating particle momentum with interaction angle.

swept across the transition frequency, which, for a suitable choice of parameters, ensures that 100% of the population finishes in the upper state. Sweeping the laser frequency is deeply unattractive, but again nature provides a way out, in the angular dependence of the Doppler shift given in Eq. 5. The correct conditions can be arranged by making the laser beam slightly divergent at the interaction region, as the change in θ is seen by the beam as a change in frequency. Only a very small beam divergence, of the order of microradians, is needed to to cover the narrow resonance linewidth.

A more difficult problem is presented by the spread in particle momenta and angles, which means that each individual particle has a different resonant frequency. This is effectively a Doppler spread, as encountered in conventional spectroscopy. The spread in particle angles from the beam emittance and lattice is typically 10^{-4} to 10^{-5} with small α, large β and normalized emittance of the order of micrometres. The beam momentum spread $\Delta p/p_0$ is typically 10^{-3} to 10^{-4}, and these effects change the resonant frequency by the same amount. These are much larger than the laser frequency spread or the resonance linewidth; increasing the laser frequency spread would require a large increase in the overall required laser power, which is technically very difficult.

To overcome the Doppler spread, it is possible to further increase laser divergence at the interaction point to the order of a milliradian, such that the resulting frequency spread then covers the Doppler spread from Δp. This solves the technical issue associated with spreading the laser frequency, but only a small fraction of the available laser photons are actually in resonance with the transition, which again requires more laser power.

A more elegant way to overcome Doppler broadening is to arrange a correlation between the individual particle momentum offset and the interaction angle θ. This can be achieved by dispersion tailoring: arranging the dispersion angle D' at the IP such that a change in angle of the individual particle generated by its momentum offset matches the difference in frequency of the resonance arising from this offset. This is illustrated schematically in Fig. 16. It is simple to show that the required D' is given by

$$D' = -\frac{\beta + \cos(\theta)}{\sin(\theta)} \,. \tag{6}$$

The final part of the laser stripping process is the removal of the excited second electron. In the SNS experiment, this is done by stripping in the fringe field of the chicane dipole. The lifetime of the excited ion depends on both the energy level and the magnetic field. Since the stripped particle is then deflected in the remainder of the magnetic field, stripping at a different S position gives a different deflection angle, and hence an emittance increase. This can be acceptable for beams with large emittance, but could be an issue for very bright beams. The stripping dipole field should be as 'hard-edged' as possible, to minimize the stripping length, but another complication comes from the different stripping rates associated with the fine and hyperfine structure in the excited ion energy levels. This effect is illustrated in Fig. 17 [14], where the fraction of unstripped beam remaining for the different excited states is plotted as a function of position in the stripping dipole field. The advantage of a high-lying excited state is clearly apparent.

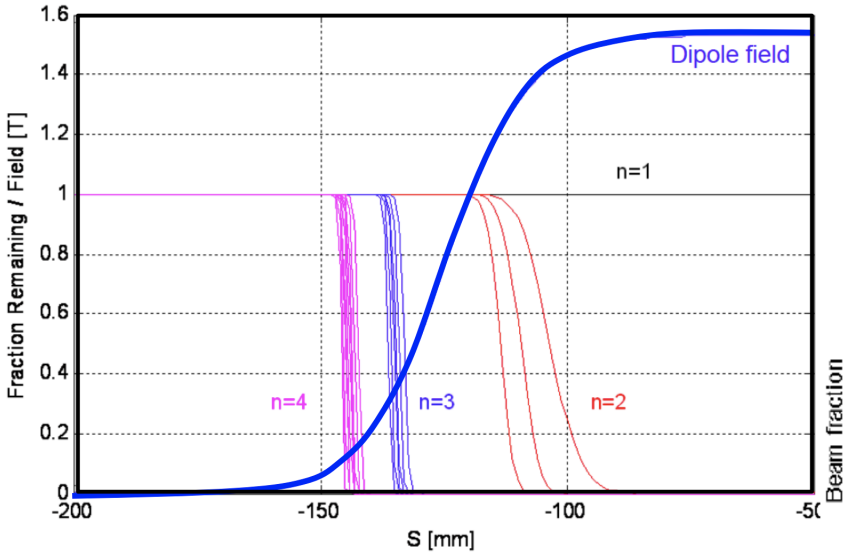


Fig. 17: Stripping of excited 4 GeV H^0 in a 1.5 T dipole fringe field. The multiple curves for each principle excited state are due to the fine and hyperfine structures.

The main performance limitation is the laser power to reach a high efficiency, given by:

$$P_{\text{peak}} = \frac{\ln(1/\delta)\hbar^2\epsilon_0 c^2 \kappa\omega_0\tau_1 \sin\theta h}{2\mu_{1n}^2\gamma(1+\beta\cos\theta)^2}\,, \tag{7}$$

where $\delta \ll 1$ is the ratio of unexcited to excited atoms, $2h$ the vertical beam size, ω the laser frequency in the laboratory frame, $\kappa = 6\Delta\lambda/\lambda$ the full relative frequency change, τ_1 the laser micropulse duration and μ_{1n} the Einstein dipole transition coefficient. For the SNS injection at 1 GeV, this turns out to be about 10 MW peak power, which is obviously not feasible in any quasi-CW way. Many ingenious proposals are combined to reduce the average laser power to a technically feasible level:

- dispersion tailoring to reduce Doppler spread: 10 MW $\rightarrow$ 2 MW;
- reduce vertical beam size and α_x: 2 MW $\rightarrow$ 1 MW;
- match laser macro-pulse to H^- beam duty factor: 1 MW $\rightarrow$ 20 kW;
- mode-lock laser pulses to 30 ps at 402 MHz: 20 kW $\rightarrow$ 240 W;
- $\times$50 recycling of laser power in Fabry–Pérot cavity: 240 W $\rightarrow$ 5 W.

7 Conclusion

The examples presented in this paper illustrate that beam transfer offers some unique opportunities for merging different domains of physics, engineering, and technology. The next breakthrough may come from a totally a different domain, as evidenced by the overlap already with solid-state physics and quantum optics.

There is always scope for creativity and elegance in the design of these systems. It is crucial to stay in touch with other disciplines, and to use imagination in approaching new (and indeed old) problems.

8 Acknowledgement

Thanks are due to the many colleagues who contributing knowingly or unknowingly to the topics described herein, and for their great creativity in tackling difficult problems in beam transfer.

References

[1] Y. Yonemura *et al.*, Beam extraction of the PoP FFAG with a massless septum, Proc. PAC, 2003.

[2] M. Barnes and B. Goddard, Considerations on a new fast extraction kicker concept for SPS, sLHC Project Note 0018, (CERN, Geneva, 2010).

[3] C. Wiessner *et al.*, Massless beam separation system for intense ion beams, Proc. PAC, 2015.

[4] W.T. Weng *et al.*, The AGS fast kicker magnet systems, Part. Accel. Conf., Washington, DC, 1981.

[5] D. Neuffer, *IEEE Trans. Nucl. Sci.* **NS-28**, (1981) 2034, https://doi.org/10.1109/TNS.1981.4331584.

[6] E.N. Tsyganov, Preprint TM-682, TM-684, Fermilab, Batavia (1976).

[7] W. Scandale *et al.*, High-efficiency volume reflection of an ultrarelativistic proton beam with a bent silicon crystal, PRL 98, 154801 (2007), https://doi.org/10.1103/PhysRevLett.98.154801.

[8] D. Mirachi, Crystal collimation for LHC, CERN-ACC-2015-0143, (CERN, Geneva, 2015).

[9] W. Scandale *et al.*, *Nucl. Instrum. Methods Phys. Res. B* **355** (2015) 390, https://doi.org/10.1016/j.nimb.2015.01.043.

[10] F. Velotti, Reduction of resonant slow extraction losses with shadowing of septum wires by a bent crystal, Proc. IPAC, 2017, p. MOPIK050.

[11] J. Wei *et al.*, Evolution of the spallation neutron source ring lattice, Proc. 20th ICFA Advanced Beam Dynamics Workshop, Chicago IL, 2002, https://doi.org/10.1063/1.1522612.

[12] V. Danilov *et al.*, *Phys. Rev. ST Accel. Beams* **6** (2003) 053501, https://doi.org/10.1103/PhysRevSTAB.6.053501.

[13] W. Bartmann *et al.*, A doublet-based injection-extraction straight for PS2, Proc. IPAC'10, Kyoto, 2010, p. 3945.

[14] B. Goddard *et al.*, Laser stripping for the PS2 charge-exchange injection system, Proc. IPAC'09, Vancouver, 2009, p. TU6RFP028.

Participants

ALEXANDRE, P.	Synchrotron Soleil, Gif sur Yvette , FR
ATANASOV, M.G.	CERN, Geneva, CH
AVVAKUMOVA, I.	JINR, Dubna Moscow region, RU
AYALA CINTAS, N.	ALBA Synchrotron Light Source, Cerdanyola del Valles, ESP
BAUCHE,J.	CERN, Geneva, CH
BELVER-AGUILAR, C.	CERN, Geneva, CH
BERJILLOS BARRANCO, A.	CERN, Geneva, CH
BERKAEV, D.	Binp sb ras, Novosibirsk, RU
BERNHARD, J.	CERN, Geneva, CH
BLANCO GARCIA, O. R.	Lnf/infn, Frascati, IT
CHARITONIDIS, N.	CERN, Geneva, CH
CHETVERTKOVA, V.	Gsi, Darmstadt, DE
CHMIELINSKA, A.	CERN, Geneva, CH
COMBLIN,J.F.	CERN, Geneva, CH
CONLON, M.	European Spallation Source ERIC, Lund, SE
D'AGOSTINO, G.	INFN-Laboratori Nazionali del Sud, Catania, IT
DANZECA, S.	CERN, Geneva, CH
DE FRANCO, A.	EBG MedAustron GmbH, Wiener Neustadt, AT
DELRIEUX, M.	CERN, Geneva, CH
DI GIOVANNI, G. P.	CERN, Geneva, CH
DI GIULIO, C.	Infn-lnf, Roma, IT
DRESSLER, O.	Helmholtz-Zentrum Berlin, Berlin, DE
ESPOSITO, L. S.	Adam sa, Versoix, CH
EVANS, N.	Oak Ridge National Laboratory, Oak Ridge Tn, US
FOGGETTA, L. G.	INFN-Laboratori Nazionali di Frascati, frascati, IT
FORMENTO CAVAIER, R.	Centre National de la Recherche Scientifique (FR), Nantes, FR
FORTE, V.	CERN, Geneva, CH
FRANZINI, G.	Infn-lnf, Roma, IT
FUCHSBERGER, K.	CERN, Geneva, CH
FURUTANI, K.	Mayo Clinic Rochester, Rochester, US
GARATTINI, M.	CERN, Geneva, CH
GEITHNER, W.	Gsi Helmholtzzentrum für Schwerionenforschung, Darmstadt, DE
GUIDOBONI, G.	CERN, Geneva, CH
HEMELSOET, G. H.	CERN, Geneva, CH
HESSLER, C.	CERN, Geneva, CH
HOLMA, J.	CERN, Geneva, CH
LITVINOV, S.	Gsi, Darmstadt, DE
MATLOCHA, T.	NPI of the CAS, Rez near Prague, CZ
MYALSKI, S.	EBG MedAustron GmbH, Wiener Neustadt, AT
OGUR, S.	CERN, Geneva, CH
PRYANICHNIKOV, A.	Physical Technical Center of P.N. Lebedev Physical Institute of the Russian Academy of Sciences, Moscow, RU
REGGIANI, D.	PSI, Villigen PSI, CH
RINGVALL MOBERG, A.	CERN, Geneva, CH

ROMAGNOLI, G.	CERN, Geneva, CH
RUAN, S.	Chinese Academy of Sciences, Lanzhou, CN
SALOU, P.	Pantechnik, Bayeux, FR
SANZ Ull, A.	CERN, Geneva, CH
SAPINSKI, M.	Gsi, Darmstadt, DE
SCHAUMANN, M.	CERN, Geneva, CH
SCHMIDT, J.	CERN, Geneva, CH
SHENG, L.	Chinese Academy of Sciences, Lanzhou, CN
SHREYBER-TECKER, I.	CERN, Geneva, CH
STEIN, O.	CERN, Geneva, CH
STOEL, L.	CERN, Geneva, CH
TOPALOUDIS, A.	CERN, Geneva, CH
VALUCH, D.	CERN, Geneva, CH
VEGA, L.	CERN, Geneva, CH
VENESS, R.	CERN, Geneva, CH
VEYRUNES, E.	CERN, Geneva, CH
VLACHODIMITROPOULOS, V.	CERN, Geneva, CH
WARREN, S.	CERN, Geneva, CH
WIESNER, C.	CERN, Geneva, CH
WINKELMANN, T.	Heidelberg Ionenstrahl-Therapiezentrum (HIT), Heidelberg, DE
WU, B.	Chinese Academy of Sciences, Lanzhou, CN
ZHANG, J. J.	Chinese Academy of Sciences, Lanzhou, CN
ZHURAVLEV, A	Binp, Novosibirsk, RU