

Stephen Myers
Herwig Schopper *Editors*

Particle Physics Reference Library

Volume 3: Accelerators and Colliders



Springer Open

Particle Physics Reference Library

Stephen Myers • Herwig Schopper
Editors

Particle Physics Reference Library

Volume 3: Accelerators and Colliders

 Springer Open

Editors

Stephen Myers
ADAM SA
Geneva, Switzerland

Herwig Schopper
CERN
Geneva, Switzerland



ISBN 978-3-030-34244-9
<https://doi.org/10.1007/978-3-030-34245-6>

ISBN 978-3-030-34245-6 (eBook)

This book is an open access publication.

© The Editor(s) (if applicable) and The Author(s) 2013, 2020

Open Access This book is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this book are included in the book's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

For many years the series Landolt-Börnstein—Group I Elementary Particles, Nuclei and Atoms, Volume 21A (Physics and Methods Theory and Experiments, 2008), Vol. 21B1 (Elementary Particles Detectors for Particles and Radiation. Part 1: Principles and Methods, 2011), Vol. 21B2 (Elementary Particles Detectors for Particles and Radiation. Part 2: Systems and Applications), and Vol. 21C (Elementary Particles Accelerators and Colliders, 2013), has served as a major reference work in the field of high-energy physics.

When, not long after the publication of the last volume, open access became a reality for HEP journals in 2014, discussions between Springer and CERN intensified to find a solution for the “Labö” which would make the content available in the same spirit to readers worldwide. This was helped by the fact that many researchers in the field expressed similar views and their readiness to contribute.

Eventually, in 2016, at the initiative of Springer, CERN and the original Labö volume editors agreed in tackling the issue by proposing to the contributing authors a new OA edition of their work. From these discussions a compromise emerged along the following lines: transfer as much as possible of the original material into open access; add some new material reflecting new developments and important discoveries, such as the Higgs boson; and adapt to the conditions due to the change from copyright to a CC BY 4.0 license.

Some authors were no longer available for making such changes, having either retired or, in some cases, deceased. In most such cases, it was possible to find colleagues willing to take care of the necessary revisions. A few manuscripts could not be updated and are therefore not included in the present edition.

We consider that this new edition essentially fulfills the main goal that motivated us in the first place—there are some gaps compared to the original edition, as explained, as there are some entirely new contributions. Many contributions have been only minimally revised in order to make the original status of the field available as historical testimony. Others are in the form of the original contribution being supplemented with a detailed appendix relating recent developments in the field. However, a substantial fraction of contributions has been thoroughly revisited by their authors resulting in true new editions of their original material.

We would like to express our appreciation and gratitude to the contributing authors, to the colleagues at CERN involved in the project, and to the publisher, who has helped making this very special endeavor possible.

Vienna, Austria
Geneva, Switzerland
Geneva, Switzerland
July 2019

Christian Fabjan
Stephen Myers
Herwig Schopper

Contents

1	Accelerators, Colliders and Their Application	1
	E. Wilson and B. J. Holzer	
2	Beam Dynamics	15
	E. Wilson and B. J. Holzer	
3	Non-linear Dynamics in Accelerators	51
	Werner Herr and Etienne Forest	
4	Impedance and Collective Effects	105
	E. Metral, G. Rumolo, and W. Herr	
5	Interactions of Beams with Surroundings	183
	M. Bruggen, H. Burkhardt, B. Goddard, F. Cerutti, and R. G. Alia	
6	Design and Principles of Synchrotrons and Circular Colliders	205
	B. J. Holzer, B. Goddard, Werner Herr, Bruno Muratori, L. Rivkin, M. E. Biagini, J. M. Jowett, K. Hanke, W. Fischer, F. Caspers, and D. Möhl	
7	Design and Principles of Linear Accelerators and Colliders	295
	J. Seeman, D. Schulte, J. P. Delahaye, M. Ross, S. Stapnes, A. Grudiev, A. Yamamoto, A. Latina, A. Seryi, R. Tomás García, S. Guiducci, Y. Papaphilippou, S. A. Bogacz, and G. A. Krafft	
8	Accelerator Engineering and Technology: Accelerator Technology	337
	F. Bordry, L. Bottura, A. Milanese, D. Tommasini, E. Jensen, Ph. Lebrun, L. Taviani, J. P. Burnet, M. Cerqueira Bastos, V. Baglin, J. M. Jimenez, R. Jones, T. Lefevre, H. Schmickler, M. J. Barnes, J. Borburgh, V. Mertens, R. W. Aßmann, S. Redaelli, and D. Missiaen	

9	Accelerator Operations	519
	M. Lamont, J. Wenninger, R. Steinhagen, R. Tomás García, R. Garoby, R. W. Assmann, O. Brüning, M. Hostettler, and H. Damerau	
10	The Largest Accelerators and Colliders of Their Time	585
	K. Hübner, S. Ivanov, R. Steerenberg, T. Roser, J. Seeman, K. Oide, Karl Hubert Mess, Peter Schmüser, R. Bailey, and J. Wenninger	
11	Application of Accelerators and Storage Rings	661
	M. Dohlus, J. Rossbach, K. H. W. Bethge, J. Meijer, U. Amaldi, G. Magrin, M. Lindroos, S. Molloy, G. Rees, M. Seidel, N. Angert, and O. Boine-Frankenheim	
12	Outlook for the Future	797
	C. Joshi, A. Caldwell, P. Muggli, S. D. Holmes, and V. D. Shiltsev	
13	Cosmic Particle Accelerators	827
	W. Hofmann and J. A. Hinton	

About the Editors



Stephan Myers was born in Belfast, Northern Ireland, and worked at CERN, Geneva, from 1972 until 2015. He was responsible for the commissioning and the energy upgrade of the CERN Large Electron-Positron Collider (LEP). In 2008, he was nominated CERN Director of Accelerators and Technology until January 2014; during this time, he directed the repair of the CERN Large Hadron Collider (LHC) after the serious accident and steered the operation of the collider in 2010, 2011, and 2012. On July 4, 2012, the data from the collider allowed the discovery of a “Higgs” boson for which Peter Higgs and Francois Englert received the Nobel Physics Prize in 2013.

He is currently executive chair of a Geneva-based company (ADAM SA), which is developing a linear accelerator for proton therapy of cancer, and non-executive Director of the parent company Advanced Oncotherapy (AVO).



Herwig Franz Schopper joined as a research associate at CERN since 1966 and returned in 1970 as leader of the Nuclear Physics Division and went on to become a member of the directorate responsible for the coordination of CERN’s experimental program. He was chairman of the ISR Committee at CERN from 1973 to 1976 and was elected as member of the Scientific Policy Committee in 1979. Following Léon Van Hove and John Adams’ years as Director-General for research and executive Director-General, Schopper became the sole Director-General of CERN in 1981.

Schopper's years as CERN's Director-General saw the construction and installation of the Large Electron-Positron Collider (LEP) and the first tests of four detectors for the LEP experiments. Several facilities (including ISR, BEBC, and EHS) had to be closed to free up resources for LEP.

Chapter 1

Accelerators, Colliders and Their Application



E. Wilson and B. J. Holzer

1.1 Why Build Accelerators?

Accelerators are modern, high precision tools with applications in a broad spectrum that ranges from material treatment, isotope production for nuclear physics and medicine, probe analysis in industry and research, to the production of high energy particle beams in physics and astronomy. At present about 35,000 accelerators exist world-wide, the majority of them being used for industrial and medical applications. Originally however the design of accelerators arose from the request in basic physics research, namely to study the basic constituents of matter.

The first accelerators were inspired by the early experiments in nuclear physics. In the early years of the twentieth century Rutherford discovered that by using alpha particles from radioactive disintegration and detecting the pattern of particles scattered by atoms one might deduce that the nucleus was a tiny but massive central element in the atom. Alpha particles from disintegration can only be of energies of 10 MeV; comparable with the nuclear binding forces. Higher energies were needed and a more reliable and steady supply to ease the tedium of counting occasional flashes of light on the scintillation screen that was Rutherford's detector. De Broglie had shown that there was an inverse relationship between the momentum, and hence the energy of a particle and the wavelength of its representation in quantum mechanics.

$$\lambda = \frac{h}{p}$$

E. Wilson · B. J. Holzer (✉)
CERN (European Organization for Nuclear Research), Meyrin, Genève, Switzerland
e-mail: Bernhard.Holzer@cern.ch

© The Author(s) 2020
S. Myers, H. Schopper (eds.), *Particle Physics Reference Library*,
https://doi.org/10.1007/978-3-030-34245-6_1

where h represents Planck's constant and p the particles' momentum which relates to its energy via the well-known equation of special relativity, $E^2 = p^2c^2 + m^2c^4$. The limit to the scale of detail that experiments can reveal is set by the length of the wave which is scattered: rather as the wave breaking in the beach can only be deflected by islands larger than itself. It was argued correctly that higher energy particle, having the property of shorter wavelengths could better reveal the structure of the nuclei that Rutherford has detected.

Such arguments led to the invention of the first accelerators and have sustained the development of particle accelerators of higher and higher energy over the best part of the last 100 years. At first, physicists used accelerators to probe the structure of the nucleus, but went on to use higher energy accelerators to search for structure in the "fundamental" particles—protons, neutrons and electrons they discovered. Inevitably higher energies implied larger accelerators, for it was quickly discovered that the best way to accelerate repetitively was to keep particle in a circular path whose radius was itself proportional to energy, limited by the strength of the magnetic field one might use to do the bending.

As energies were raised physicists found new and interesting particles to fit into the pattern of those that their theories might predict. Einstein's

$$E = mc^2$$

tells us that only high energies will create the more massive particles. The latest and largest accelerator, LHC, flagship of the whole community, was designed to search for the Higgs Boson and the successful discovery of this missing puzzle piece in 2013 allowed us to complete the Standard Model of Particle Physics.

As we write, this machine is carrying on the search for physics beyond the standard model, seeking to disclose the nature of dark matter and dark energy.

As more powerful accelerators have been developed for high energy particle physics, advances in the field have been exploited in a whole range of smaller accelerators for other applications. From the time of the first cyclotrons they have been used for producing isotopes and for treating cancer. The development of compact high-frequency linac structures triggered the manufacture of hundreds of small electron linacs producing X-rays for cancer treatment in hospitals around the developed and, latterly, the developing world. Electron rings of a few GeV, specially designed to produce beams of synchrotron radiation have become popular. Each facility serves scores of experiments to investigate the structure of complex molecules—particularly the proteins of today's biomedical studies. Proton accelerators of about 1 GeV produce pulsed beams of neutrons by spallation which are used principally to study the structure of materials. In addition thousands of lower energy accelerators are used in industry for sterilisation and ion implantation in the fabrication of sophisticated CPU chips for computers.

1.2 Types and Evolution of Accelerators

The development of accelerators to ever higher energy is marked by a number of milestones. Each of these marks the invention of a new type of accelerator or the invention of a new principle of transverse or longitudinal focusing which enables a higher energy to be reached for a lower unit cost. The best way to describe this evolution and introduce the different types of accelerator is to follow the road charted by these milestones. Each is described in one of the sections which follow.

1.2.1 Early Accelerators

The nineteenth century had produced a number of electrostatic high-voltage generators. They were unpredictable in performance and electrical breakdown became a serious problem above a few tens of kV. Early accelerators were simply two electrodes enclosed in an evacuated tube with external connections to such high voltage source. A proton or electron source close to one electrode at a potential of V (or $-V$ for electrons) provided the particles which were then accelerated towards the second electrode at earth potential. They emerged or were observed through a small hole in the earthed electrode. The energy acquired by each particle with charge, e Coulombs, was just e_*V Joules or, in the units commonly used for accelerated beams, V electron-Volts. An electron Volt is then just 1.6×10^{-19} Joules. If the particle is a fully stripped ion of an atom with atomic number A and charge Z then the energy is ZV/A electron Volts per nucleon.

The first high-voltage generator to approach 1 MeV was built by Cockcroft and Walton [1–3] in the 1930s to accelerate particles for their fission experiments. Their combination of diodes and capacitors, also known as rectifier circuit, is still used today to apply high voltage to the ion or proton source at the beginning of many linacs and synchrotrons although these are gradually being replaced by radio frequency quadrupoles.

The early 1930s also saw the invention by R.J. Van der Graaf [4] of an electrostatic generator which used a moving belt to carry charge into the high voltage terminal until it reaches a potential of several MV (Fig. 1.1). Van der Graaf accelerators have proved a useful source of low energy particles to this day but are inevitably limited by problems of voltage breakdown. Voltages up to 27 MV have been reached, putting the device in a discharge suppressing gas atmosphere (e.g. SF_6). Although it is possible in theory to chain together several electrostatic accelerators, each with its cathode connected to the anode of the next, each stage increases the potential between the ends of the device and between the ends and ground and eventually electrical breakdown discharges the high voltage terminals.

Fig. 1.1 Van der Graaf accelerator

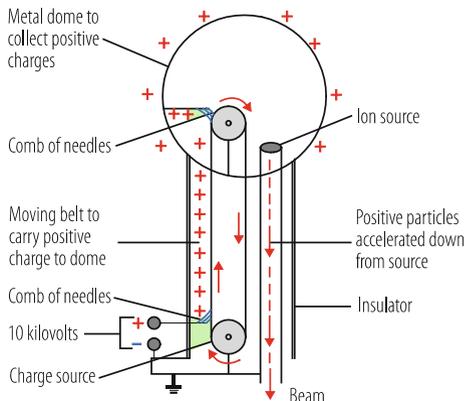
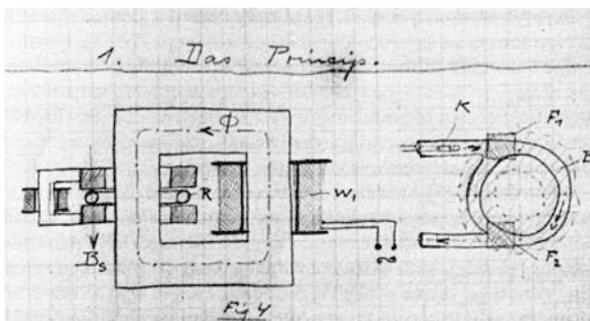


Fig. 1.2 Wideröe's sketch of the ray transformer



1.2.2 The Ray Transformer

The earliest idea of how to overcome the limitations of electrostatic acceleration involved using the time varying property of magnetic fields and came from the inventive mind of Rolf Wideröe.

Beginning his studies at Karlsruhe Technical University in 1923, he wondered if electrons in an evacuated ring would flow in the same way as the electrons in copper if they replaced the secondary winding of a transformer. His notebooks of that time contain sketches of a device he called a “ray transformer”; the first circular accelerator and the precursor of the “betatron” [5].

These sketches show a beam tube, in the form of an annulus, R , placed in the gap between the parallel poles or faces of a small electromagnet (on the left in Fig. 1.2). This magnet is in the form of a “C” and the field between the poles, B_z , guides particles in a circular orbit in the mid plane between the poles. A circular hole is cut in each pole through which the yoke of the transformer passes linking the beam tube. The primary winding of the transformer, labelled W_1 , is powered with alternative voltage from the mains. The beam tube is placed where one would normally expect the secondary winding. The beam within it carries the induced secondary current.

Unlike almost all accelerators that followed, the ray transformer relied entirely upon the inductive effect of a varying magnetic field. It is the rate of change of flux, ϕ , in the yoke which establishes an accelerating potential difference around the beam's path. The windings, that of the C-magnet and of the primary of the transformer, W_1 , give independent control of the guide field and accelerating flux.

Wideröe calculated that electrons circulating in a ring of only 10 or 20 cm diameter could reach several MeV within one quarter wave of the AC excitation of the transformer. He had to use Einstein's newly discovered theory of special relativity to correctly describe the motion of particles close to the speed of light. He also found an important principle which ensures that the beam radius does not change as it accelerates. To ensure constant radius during acceleration the total flux linking the beam including that generated by both sets of the coils, B_a , must be twice that generated by the left hand coil pair which produces the field keeping the beam in a circular orbit, B_g .

$$\dot{B}_a = 2\dot{B}_g$$

Unfortunately, Wideröe was dissuaded from building the ray transformer by difficulties with surface fields and by his professor, who wrongly assumed the beam would be lost because of gas scattering. However, his Ray Transformer and the 2 to 1 ratio, now known as the Wideröe principle, were important discoveries which were put into practice 15 years later when D.W. Kerst and R. Serber [6] built a series of betatrons.

Wideröe went on to develop a second basic acceleration method to overcome the electrostatic limitation: the drift tube linac.

1.2.3 Repetitive Acceleration

There are two broad classes of accelerator characterized by the way they achieve repetitive acceleration and which overcome the insulation problems of the electrostatic machines. The simplest concept is that of the linear accelerator. Particles pass through cavities excited by radio frequency generators. They arrive on the threshold of each cavity with the energy they have already received and gain a further increment in energy from the electric field in the cavity which points in their direction of motion. Each cavity performs the function of the gap between the anode and cathode of an electrostatic accelerator but, unlike the electrostatic case, the increments of energy may be added together without developing a huge voltage to ground in any part of the apparatus. Of course, there is a limit to the voltage (energy increment) each cavity can apply and the length of the device becomes very long for energies above 1 GeV. Nevertheless, a linac has become the only way of accelerating highly relativistic electrons which radiate a large fraction of their energy when bent into a circular path.

As alternative concept, circular machines, like cyclotrons and synchrotrons use the same set of accelerating cavities over and over again as the particles make complete turns around the accelerator, being guided and focused by the magnet structure of the ring which is thus constraining their orbit. On each turn an increment of energy is added and, once accelerated, particles may be allowed to circulate indefinitely at their top energy. Two circulating beams of say protons and antiprotons or electrons and positrons can be sustained in the same ring and, colliding at experiments around the circumference, create new particles up to a mass (centre of mass energy as it is called) that is the sum of the two energies. Colliders are today the preferred configuration for a high-energy machine. Earlier, new particles were sought in the debris from a particles collision with a nucleon in a fixed target but such collisions are limited to a smaller centre of mass energy—which rises only as the square root of the accelerated beam energy.

1.2.4 Linear Accelerators

Although disappointed by the rejection of his ray-transformer as a subject for his PhD, Wideröe was led to the idea of a linear accelerator by a paper by G. Ising [7] who tried to overcome the voltage breakdown problem of a single stage of acceleration by placing a series of hollow cylindrical electrodes one after another in a straight line to form what today we would call a ‘drift tube linac’ or linear accelerator. Wideröe realised that an oscillating potential applied to one drift tube flanked by two others which are earthed, accelerates at both gaps provided the oscillator’s phase changes by 180° during the flight time between gaps.

In 1927 he built a three-tube model which accelerated sodium ions. At the wavelengths that radio transmitters generated at that time a particle travelling near the velocity of light would travel hundreds of meters in the time it would take for the r.f. to swing by half a sine wave. This would make the length of a drift tube impractically large. Sodium ions, being rather heavy compared with protons or electrons, travelled much slower than the velocity of light and this helped keep the apparatus down to table-top proportions. Although he realised that one might extend such a series of tubes indefinitely he did not take the idea any further as he was due to start his professional employment designing high voltage circuit breakers. Between 1931 and 1934, D. Sloan and E.O. Lawrence at Berkeley took up Wideröe’s idea and constructed linacs with as many as 30 drift tubes to accelerate mercury ions but, these were never used for research.

Much later, in the mid-1940s, and when suitable high-power high-frequency oscillators had become available to meet the needs of war-time radar, L.W. Alvarez (1946) started to build the first serious proton linac at the Radiation Laboratory of the University of California. Figure 1.3 shows an Alvarez linac. A series of drift tubes are mounted within a copper-lined cylinder excited by a radio transmitter. As in Wideröe’s linac, particles gain energy from the accelerating potential differences between the ends of the drift tube, but now the phase shift between drift tube

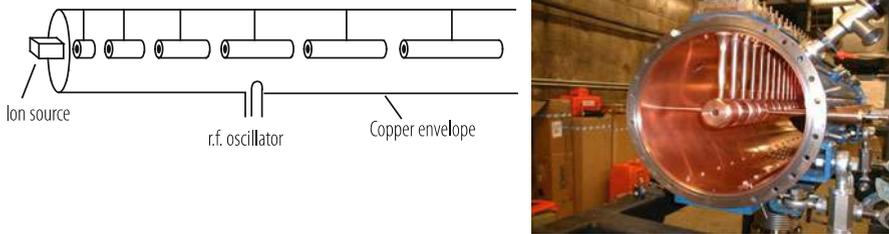


Fig. 1.3 left: The concept of the drift tube linac (from [8]); right: CERN's Linac 1

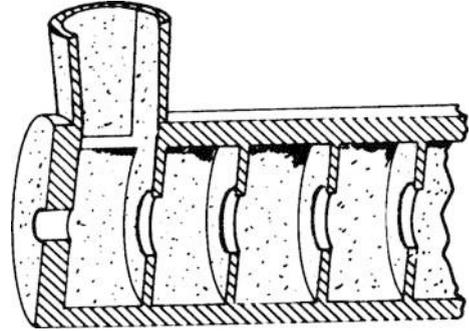
gaps is 360° . Each gap appears to the particle to be an identical field gradient which accelerates particles from left to right. The particles are protected from the decelerating phase while inside the metallic drift tubes. Although the particle gains energy steadily as it passes each gap, the total voltage between parts of the assembly and ground does not become larger along the length of the device as it would for an electrostatic machine.

The distances between gaps, or the lengths of the tubes, increase as the particle is accelerated since it travels an ever increasing distance during one swing of the radio frequency oscillation. At low energy, we would expect this distance to increase with the velocity or the root of the kinetic energy but when the energy is large we find the length of the drift tubes and their spacing no longer increases—a practical demonstration of special relativity. The Alvarez structure is still widely used, especially for non-relativistic proton and ion beams.

It was well known at the time that waves might be propagated along a much simpler smooth waveguide and that some of the modes have an accelerating electric field in the direction of propagation. Closer examination however shows that the stumbling block is that the phase velocity of these modes in a wave guide is always greater than that of light and hence the particle sees a field which sometimes accelerates and then decelerates as the wave overtakes the particle. It was later found that the phase velocity could be reduced by a series of iris diaphragms in the pipe. Such a structure (Fig. 1.4) is very popular in electron linacs and also in storage rings in which the particle is close to the velocity of light and cavities need not be tuned to follow the acceleration cycle.

These diaphragm-loaded linac structures have been commonly used as injectors for circular accelerators to accelerate electrons and protons to energies in the range 10 to 1000 MeV. As compact high frequency structures they have also been widely used to accelerate electrons to, typically 10 MeV, as a source of X-rays for cancer therapy. An early and very successful adventure in the electron linac development was the “two-mile long” Stanford Linear Accelerator at SLAC in California which has been the work horse for a number of ground breaking fixed target experiments and circulating beam storage ring projects at 20 to 50 GeV. With the help of two semi-circular arcs it was used to bring beams of electrons and positrons into head-on collision in the Stanford Linear Collider Project. This project, is forerunner for

Fig. 1.4 Iris loaded structure (from [9]). The ‘chimney’ is the input waveguide



today’s projected Linear Colliders in which linear accelerators accelerate positrons and electrons to energies approaching 1 TeV to collide them head on in a bid to overcome the very considerable energy lost by an electron to synchrotron radiation in circular lepton rings at high energy.

1.2.5 Cyclotrons

Unlike a linac, whose length must be extended to reach a higher energy, the cyclotron, as it is called, is a relatively compact accelerator in which the energy is only limited by the diameter and field strength of the magnet. The cyclotron idea first occurred to E.O. Lawrence who, reading through Wideröe’s thesis, ruminated on the possibility of using a magnetic field to recirculate the beam through two of drift tubes. The cyclotron idea was published in 1930 [10] and another colleague, M.S. Livingston, who was also later to contribute much to the field, was given the job of making a working model as his doctoral thesis.

In Fig. 1.5 we see the two ‘Dee’s’ which comprise the positive and negative electrodes of the accelerating system between the poles of the magnet. These are like two halves of a closed cylinder divided along its diameter. A radio-frequency generator excites them with an alternating field of constant frequency. The potential difference between the ‘Dee’s’ accelerates the ions as they pass the gap between the two halves of the structure. The fundamental trick is that the field oscillates at the particle’s circulation frequency and hence the sign of the potential difference at each gap is always in the accelerating direction.

As long as cyclotrons accelerate ions to modest energies, classical rather than relativistic mechanics still applies. In Fig. 1.6 we see the balance between centripetal acceleration of motion in a circle and the force exerted by the vertical magnetic field,

$$evB = \frac{mv^2}{\rho}, \text{ if } v \ll c, \quad (1.1)$$

Fig. 1.5 The principle of the cyclotron

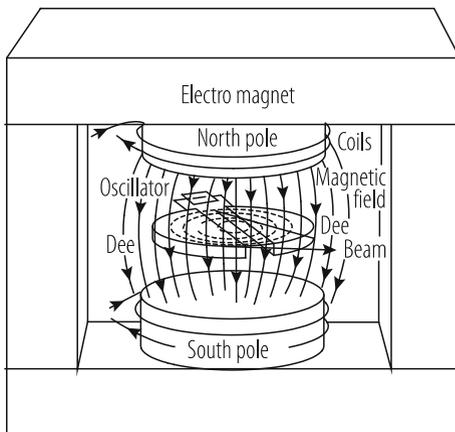
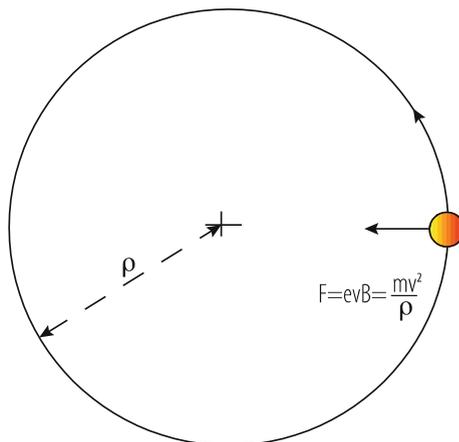


Fig. 1.6 Balance of forces in a cyclotron



and, rearranging, we can define the magnetic rigidity—the reluctance of the beam to be bent in a curve:

$$BQ = \frac{mv}{e}, \text{ if } v \ll c. \tag{1.2}$$

In the relativistic regime if we replace the classical momentum, mv , by the relativistic momentum, $p = \gamma mv$, with γ being the Lorentz factor, we obtain the equation, valid in the relativistic regime:

$$BQ = \frac{p}{e}. \tag{1.3}$$

By good fortune the radius of the orbit in a cyclotron is proportional to the velocity and the frequency of revolution this being the inverse of the time of

revolution—just the length of one turn divided by the particle’s velocity

$$f = \frac{v}{2\pi\rho} = \frac{v}{2\pi} \cdot \frac{eB}{mv}. \quad (1.4)$$

has a numerator and denominator which are both proportional to v . This frequency remains constant as the particle is accelerated in the low energy, classical, regime. Thus, the circulating particles stay in synchronisation with the oscillating RF field and a continuous stream of ions injected in the centre will follow a spiral path to reach their highest energy at the rim of the poles.

Unfortunately, the synchronism between r.f. voltage and revolution frequency breaks down as the particles velocity begins to approach that of light and the relativistic mass in the above equation is no longer constant. This happens over 30 MeV for protons and at double this energy for deuterons. Electrons are much too light and relativistic to be accelerated in a cyclotron to any significant energy. For them other acceleration concepts are more adequate, like the disk loaded travelling wave linac or the betatron that both were described before.

The possible remedy of making the field stronger at the edge of the poles would have preserved synchronism and continuous beams but, as we shall see, was in conflict with the need to have a negative radial gradient to the field to provide vertical weak focusing. As a consequence a more powerful concept had to be developed to achieve highest particle beam energies: The synchrotron.

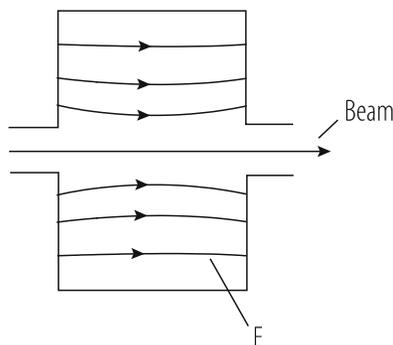
1.2.6 *The Synchrotron*

Meanwhile, in the 1940s, still higher energies were needed to pursue the aims of physics and the stage was set for the discovery of the synchrotron principle which opened the way to the series of circular accelerators and storage rings which have served particles physics up to the present day. It was Australian physicist Mark Oliphant who synthesized three old ideas into a new concept—the synchrotron. The ideas were: accelerating between the gaps of resonators, varying the frequency, and pulsing the magnet. In 1943 he described his invention in a memo to the UK Atomic Energy Directorate (see [11]).

Particles should be constrained to move in a circle of constant radius thus enabling the use of an annular ring of magnetic field . . . which would be varied in such a way that the radius of curvature remains constant as the particles gain energy through successive accelerations by an alternating electric field applied between coaxial hollow electrodes.

Unlike the cyclotron, the synchrotron accelerates the beam as a series of discrete pulses or “bunches” as they are called. Each short pulse is injected at low field and then the field rises in proportion to the momentum of particles as they are accelerated. This ensures that the radius of the orbit remains constant. In contrast to cyclotrons and betatrons, the synchrotron needs no massive poles to support a

Fig. 1.7 A simple accelerating cavity



magnetic field within the beam's circular orbit. The guide field is instead provided by a slender ring of individual magnets. The fact that the machine is pulsed and the frequency must be controlled to track the increasing speed of particles is a complication, but it solves the difficulty that isochronous cyclotron builders had encountered in accelerating relativistic particles.

Instead of the Dees of a cyclotron acceleration is provided in a synchrotron by fields within a hollow cylindrical resonator or "pillbox" cavity, Fig. 1.7, excited by a radio transmitter. A particle passes from left to right as it completes each turn of the synchrotron receiving another increment in energy at each revolution.

The early synchrotrons, like the cyclotron before them, relied on a slight negative radial gradient in the vertical magnet field to produce field lines which belly outwards from the magnet gap. A small radial field component deflects any particles which head off towards the poles back to the median plane. Unfortunately, this field shape has the opposite (defocusing) effect horizontally but, up to a certain, rather weak, gradient strength focusing is assured by a slight imbalance between the central force and the centrifugal acceleration. The gradient cannot be too large—hence the term "weak focusing".

Oliphant was the first to start building a proton synchrotron (at Birmingham University) but he was overtaken by Stan Livingston's 3 GeV Cosmotron at Brookhaven National Laboratory and later by the 6 GeV Bevatron at Berkeley.

Due to the weak focusing forces in these first synchrotrons, the particles' excursions, both horizontally and vertically are large and the magnet pole width and gap correspondingly so. Strong focusing changed this. It was invented at the Cosmotron, which was actually the first proton synchrotron to operate, whose weak focusing 'C' shaped magnet was open to the outside. The top energy of the Cosmotron was limited by the extra fall-off in field caused by the effect of saturation. Stan Livingston and E.D. Courant wanted to compensate this by re-installing some of the C magnets with their return yokes towards the outside. They were afraid of the variations in gradient around the ring but were surprised to calculate that the focusing seemed to improve as the strength of the alternating component of the gradient increased. Courant, Livingston, and H.S. Snyder [12, 13] were able to explain this retrospectively with an optical analogy of alternating focusing by equal



Fig. 1.8 The CERN 25 GeV proton synchrotron

convex and concave lenses which will transport rays which pass through the centres of defocusing lenses.

Alternating gradient or strong focusing greatly reduces the beam's excursions and so the cross section of the magnet gap by more than an order of magnitude. Its discovery enabled Brookhaven and CERN to build the next generation of proton synchrotrons, AGS and PS, to reach 30 GeV—five times the energy of the Bevatron—yet use beam pipes of only a few centimetres height and width.

This was to lead to huge economies in the cost per unit length of the magnet system. Figure 1.8 shows how this was applied to the first of the two synchrotrons, AGS and PS that used this focusing system. From then on all synchrotrons and, later, storage ring colliders use this scheme. The history of synchrotrons has been always to seek methods of improving focusing and economizing on magnet aperture. The only other step function in their development to higher energies has been the use of superconducting magnets whose higher fields reduce the circumference of the machine by a factor between 3 and 5.

1.2.7 Phase Stability

When the first synchrotrons were built it was by no means obvious that the circulating beam and the accelerating voltage would remain in step. There were those who thought that any slight mistiming of the sine wave of accelerating voltage in the cavity might build up over many turns until particles would begin to arrive

within the negative, decelerating, phase of the sine wave and be left behind. Even if one succeeded in achieving synchronism for the ideal, *synchronous particle*, others of slightly different energy would not have the same velocity and take a different time to circulate around the machine. Would not these particles gradually get out of step until they were lost? After all, particles had to make many hundred thousand turns before reaching full energy and while transverse focusing was understood there was no apparent focusing available in the longitudinal direction. Fortunately the comforting principle of phase stability, which prevents this happening, was soon to be independently discovered by V. I. Veksler in Moscow in 1944 [14] and McMillan in Berkeley in 1945 [15], opening the way to the construction of the first synchrotrons. We shall return to this later.

When it came to the next generation of synchrotrons, interest focused on colliding two opposing beams of particles. It had been known for some time that the energy available in the centre of mass from a collision of particles, one in the beam with energy E and the other of mass m_0 in a fixed target, only increased with the square root of the accelerators energy, $\sqrt{m_0 E}$. Two particles of the same mass and energy E colliding head on made available all their energy in the centre of mass, $2E$. The difficulty was making the two bunches of particles of sufficient density to have a significant probability of collision or, in technical jargon a high enough luminosity. Once this problem was solved a series of colliders: ISR, Sp \bar{p} S, LEP, Tevatron, HERA and finally LHC followed. Some of these (ISR, HERA and LHC) were two separate rings which intersected to collide particles at several points around the circumference. Others (Sp \bar{p} S and LEP) collided protons with antiprotons and electrons with their anti-particles: positrons. These exploited the fact that beams of particles and antiparticles will circulate on identical trajectories, but in opposite directions, in a single ring of bending and focusing magnets.

At present several studies are ongoing, to pave the way to even higher energies, mainly increasing the size of the machine and using super conducting magnets with higher critical field, to gain more bending and focusing fields in the lattice. One example, the Future Circular Collider study, FCC, under the guidance of CERN, is studying a 100 km proton storage ring to achieve centre of mass energies of up to 100 TeV. The R & D effort of accelerators of this dimension and complexity, in any case, has to be done by a truly international, in other words worldwide effort.

References

1. J.D. Cockcroft, E.T.S. Walton: Proc. Roy. Soc. A 129 (1930) 477-489.
2. J.D. Cockcroft, E.T.S. Walton: Proc. Roy. Soc. A136 (1932) 619-630.
3. J.D. Cockcroft, E.T.S. Walton: Proc. Roy. Soc. A 137 (1932) 229-242.
4. R.J. Van der Graaf: Phys. Rev. 38 (1931) 1919.
5. R. Wideröe.: ETH Library, Zürich, Hs 903 (1923-28) 633-638.
6. D.W. Kerst, R. Serber: Phys. Rev. 60 (1941) 53-58.
7. G. Ising: Arkiv för matematik o. fysik 18 (1924) 1-4.
8. J.J. Livingood: *Principles of cyclic particle accelerators*, van Nostrand (1961).

9. P. Lapostolle, A. Septier: *Linear Accelerators*, North Holland (1971).
10. E.O. Lawrence, N.E. Edelfsen: *Science* 72 (1930) 376-7.
11. M. Oliphant: *The genesis of the Nuffield Cyclotron and the Proton Synchroton*, Publ. Department of Physics, University of Birmingham.
12. E.D. Courant, M.S. Livingston, H.S. Snyder: *Phys. Rev.* 88 (1952) 1190-1196.
13. E.D. Courant, H.S. Snyder: *Annals of Physics* 3 (1958) 1-48.
14. V.I. Veksler: *Comptes rendues (Doklady) de l'Academie des Sciences de l'URSS* 43 (1944) 329-341.
15. E.M. MacMillan: *Phys. Rev.* 68 (1945) 143.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 2

Beam Dynamics



E. Wilson and B. J. Holzer

2.1 Linear Transverse Beam Dynamics

Now let us look in detail at the motion of particles in the transverse coordinates of the coordinate system defined in Fig. 2.1.

2.1.1 Co-ordinate System

The guide field of a synchrotron is usually vertically directed, causing the particle to follow a curved path in the horizontal plane (Fig. 2.1). The force guiding the particle in a circle is horizontal and is given by:

$$\mathbf{F} = e \cdot \mathbf{v} \times \mathbf{B}, \quad (2.1)$$

where:

\mathbf{v} is the velocity of the charged particle in the direction tangential to its path,
 \mathbf{B} is the magnetic guide field.

The guide field inside a dipole magnet is uniform and the ideal motion of the particle is simply a circle of (local) radius of curvature, $\rho(s)$. The trajectory of an ideal particle (ideal in energy and without any amplitude) that is defined by the arrangement of the dipole magnets is called *design orbit*. The machine is usually designed with this orbit at the centre of its vacuum chamber. Now there is no such

E. Wilson · B. J. Holzer (✉)
CERN (European Organization for Nuclear Research), Meyrin, Geneva, Switzerland
e-mail: Bernhard.holzer@cern.ch

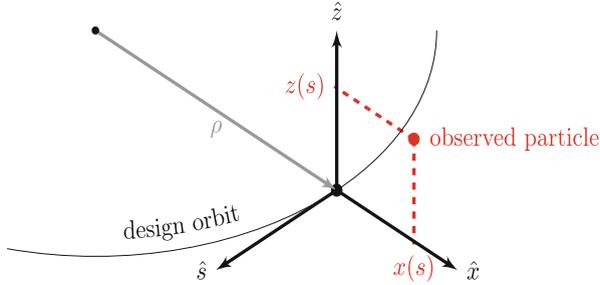


Fig. 2.1 Charged particle orbit in magnetic field

thing as an ideal particle. Still, we shall suppose that it is possible to find an orbit or curved path for the non-ideal particle which closes on itself around the synchrotron, which we call the *closed* or *equilibrium orbit* and it should be close enough to the ideal design orbit.

2.1.2 Displacement and Divergence

A beam of particles enters the machine as a bundle of trajectories spread about the ideal orbit. At any instant a particle may be displaced horizontally by x and vertically by z from the ideal position and may also have divergence angles horizontally and vertically:

$$x' = dx/ds, \quad \text{and} \quad z' = dz/ds. \quad (2.2)$$

The divergence would cause particles to leave the vacuum pipe except for the carefully shaped field which restores them back towards the beam centre so that they oscillate about the ideal orbit. The design of the restoring fields determines the transverse excursions of the beam and the size of the cross section of the magnets and is therefore of crucial importance to the cost of a project.

2.1.3 Bending Magnets and Magnetic Rigidity

The design of a synchrotron; the diameter of the ring and its sheer size and cost for a given energy is driven by the fact that bending particle trajectories depends on a magnetic rigidity. The rigidity increases with momentum and is a function of the bending field which, for room temperature magnets, saturates at about 2 T.

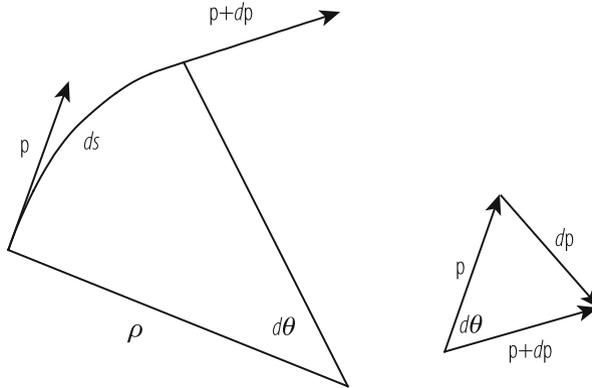


Fig. 2.2 Vector diagram showing differential changes in momentum for a particle trajectory

We will now briefly derive an expression for the magnetic rigidity of a relativistic. A particle has a relativistic momentum vector \mathbf{p} and travels perpendicular to a field \mathbf{B} which is into the plane of the diagram (Fig. 2.2).

We write the Lorentz force on the particle on its circular path as

$$\mathbf{F}_{Lorentz} = e * (\mathbf{v} \times \mathbf{B})$$

Assuming an idealized homogeneous dipole magnet along the particle orbit, having pure vertical field lines, the condition for a perfect circular orbit is defined as equality between this Lorentz force and the centrifugal force.

$$\mathbf{F}_{centrifugal} = \frac{\gamma m v^2}{\rho}$$

This yields the following condition for the idealized ring:

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

where we are referring to protons and have accordingly set $q = e$. We conclude that the beam rigidity $B\rho$, given by the magnetic field and the size of the machine, defines the momentum of a particle that can be carried in the storage ring, or in other words, it ultimately defines, for a given particle energy, the magnetic field of the dipole magnets and the size of the storage ring.

We really should use the units Newton-second for p and express e in Coulombs to give $(B\rho)$ in Tesla-metres. However, in charged particle dynamics we often talk in a careless way about the ‘momentum’ pc . This actually has the dimensions of an energy and is expressed in units of GeV.

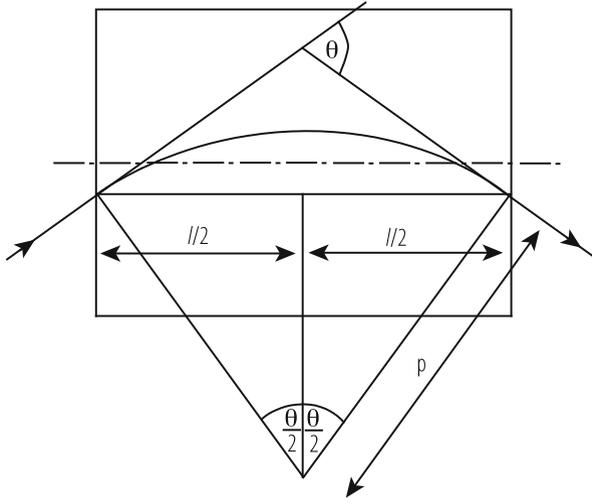


Fig. 2.3 Geometry of a particle trajectory in a bending magnet of length ℓ and deflecting angle θ

A useful rule of thumb formula based on these units is:

$$B\rho \text{ [T} \cdot \text{m]} = 3.3356 \rho c \text{ [GeV]}. \quad (2.3)$$

2.1.4 Particle Trajectory in a Dipole Bending Magnet

The trajectory of a particle in a bending magnet or dipole of length ℓ is shown in Fig. 2.3. Usually the magnet is placed symmetrically about the arc of the particle's path. One may see from the geometry that:

$$\sin(\theta/2) = \frac{\ell}{2\rho} = \frac{\ell B}{2(B\rho)}, \quad (2.4)$$

and, if $\theta \ll \pi/2$

$$\theta \approx \frac{\ell B}{(B\rho)}. \quad (2.5)$$

So the bending angle provided by a dipole magnet is given by the ration of its integrated field strength and the beam rigidity.

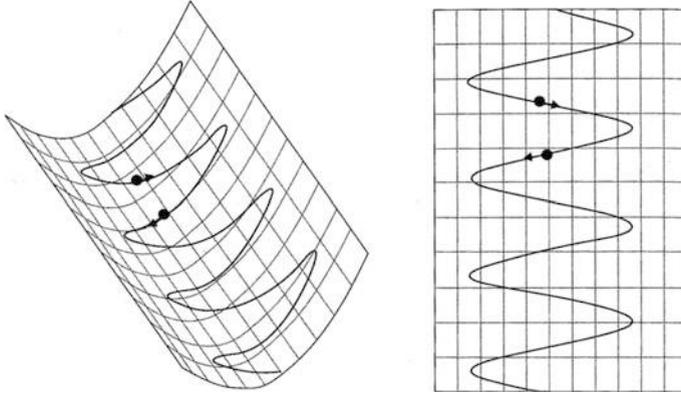


Fig. 2.4 Two views of a sphere rolling down a gutter as it is focused by the walls

2.1.5 Weak Focusing

We have mentioned that cyclotrons and early synchrotrons relied on weak focusing to constrain circulating particles within the vacuum chamber. In order to provide this, the vertical guide field has a slight negative gradient in the radial direction around the rim of the accelerator. The field lines belly out from the outer gap of the magnet. It can be shown, by applying $\nabla \times \overline{\mathbf{B}} = \mathbf{0}$, that there will be horizontal field components in this region. These produce vertically directed forces on errant particles causing them to oscillate about the median plane in a potential well (Fig. 2.4).

The motion is analogous to a small sphere rolling down a slightly inclined gutter with constant speed. Figure 2.4 shows two views of this motion and from the bottom view we recognise the motion as a sine wave. Note too that the sphere makes four complete oscillations along the gutter. In the language of accelerators, its motion has a wave number or “tune”, $Q = 4$.

To complete the analogy of a weak focusing synchrotron we imagine that we bend the gutter into a circle rather like the brim of a hat. We provide the necessary instrumentation to measure the displacement of the sphere from the centre of the gutter each time it passes a given mark on the brim and we also have a means to measure its transverse velocity. With the aid of a computer, we might convert this information into the divergence angle, which is used as vertical axis in Fig. 2.5:

$$x' = \frac{dx}{ds} = \frac{v_{\perp}}{v_{\parallel}}. \quad (2.6)$$

We can make a ring shaped gutter out of a slightly different length of gutter than is shown so that Q is not an integer. We can plot a point for each arrival of

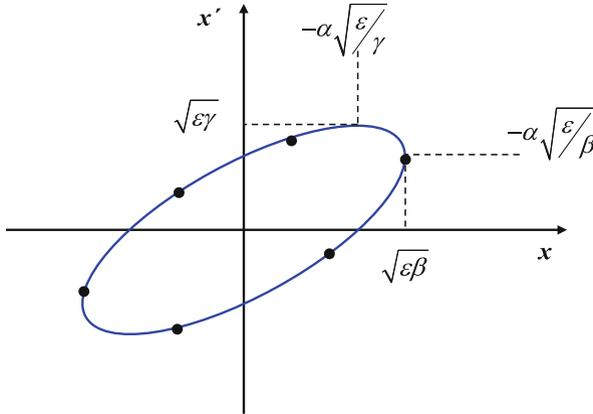


Fig. 2.5 The elliptical locus of a particle's history in phase space as it circulates in a synchrotron

the sphere in a diagram of x' against x which we call a 'phase space diagram' of transverse motion. The sphere has a large transverse velocity as it crosses the axis and has almost zero transverse velocity as it reaches its maximum displacement.

If we plot these 'observations' they will be an ellipse (Fig. 2.5) and the phase of the oscillator will advance by Q evolutions each time the particle returns. Of course, only the fractional part of Q may be deduced from our observations since our measurements do not reveal what happens round the rest of the hat's brim.

Now let us use the analogy to define some of the transverse dynamical quantities of a particle beam. The area of the ellipse is a measure of how much the particle departs from the ideal trajectory, represented in the diagram by the origin.

$$\text{Area} = \pi \varepsilon \text{ [mm} \cdot \text{rad]}. \quad (2.7)$$

In accelerator notation we use ε , the product of the semi-axes of the ellipse as a measure of the area called the emittance. The emittance is usually quoted in units of π mm·mradians. Thus if the semi-axes are 1 mm and 1 mrad the emittance will be 1 mm·mradian but the area will be π mm·mradian. The maximum excursion in displacement, the major axis, of the ellipse is defined as:

$$\hat{x} = \sqrt{\varepsilon\beta}, \quad (2.8)$$

At locations where the beta function reaches an extremum, i.e. $\alpha = 0$, we obtain hence

$$\hat{x}' = \sqrt{\varepsilon/\beta}. \quad (2.9)$$

We shall see that β (later to be called the envelope or betatron function) is a property of the gutter, not the beam. In the synchrotron it varies around the ring and is the envelope function we have plotted in Fig. 2.10 and again in Fig. 2.11. By analogy, the “brim of the hat” which represents the alternating gradient focusing system shown in this figure will vary its width and curvature around the crown and β will follow this variation in some way.

2.1.6 Alternating Gradient Focusing

In Chap. 1 we described a major break-through in the design of synchrotrons: the discovery of alternating gradient (AG) focusing (see [1] for an excellent summary of the dynamics of AG focussing). This allowed designers to use much stronger focusing systems with considerable savings in the space needed for the beam cross section.

The principle is shown in Fig. 2.6 which depicts an optical system in which each lens is concave in one plane while convex in the other and they alternate. It is possible, even with lenses of equal strength, to find a ray which is always on axis at the D lenses in the horizontal plane and therefore only sees the F lenses. To appear like Fig. 2.6 the spacing of the lenses would have to be $2f$. If the ray is also central in the lenses which are vertically defocusing, the same condition will apply simultaneously in the vertical plane. At least one particular particle or trajectory corresponding to this ray will never be defocused and be contained indefinitely.

The alternating gradient idea will work even when the rays in the D lenses do not pass exactly at their centre and the lenses are not spaced by precisely $2f$. In fact it is sufficient for the lens strengths and spacing to be chosen to ensure that the particle trajectories *tend* to be closer to the axis in D lenses than in F lenses as shown in Fig. 2.10.

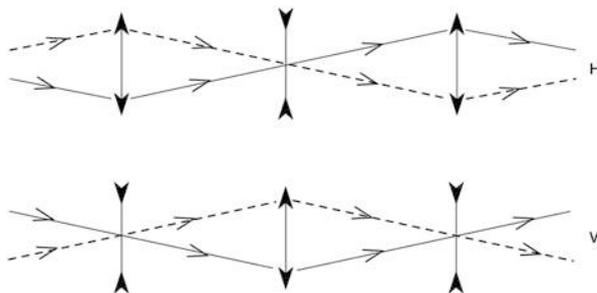


Fig. 2.6 Optical analogy with an alternating pattern of lenses

2.1.7 Quadrupole Magnets

The first alternating gradient synchrotrons used alternating magnetic lenses formed by bending magnets having the same vertical guide field but a radial gradient of alternating sign. In a modern synchrotron the functions of guiding and focussing the beam are separated. The dipole magnets which do the guiding have no gradient. The principal focusing elements are quite a different kind of magnet with four poles which produce gradient but no bending. The poles of these quadrupole magnets are truncated rectangular hyperbolae and alternate in polarity around the aperture circle which just touches the poles.

Figure 2.7 shows a particle's view of the fields and forces in the aperture of a quadrupole as it passes through normal to the plane of the paper. The field shape is such that it is zero on the axis of the device but its strength rises linearly with distance from the axis. This can be seen from a superficial examination of Fig. 2.7 if we remember that the product of field and length of any field line joining the poles is a constant. Symmetry tells us that the field is vertical in the median plane (and purely horizontal in the vertical plane of asymmetry). The field must be downwards on the left of the axis if it is upwards on the right.

The horizontal focusing force, $-evB_z$, has an inward direction on both sides and, like the restoring force on a weight suspended from a spring, rises linearly with displacement, x . The strength of the quadrupole is characterised by its gradient dB_z/dx normalised with respect to magnetic rigidity:

$$k = \frac{1}{(B\rho)} \frac{dB_z}{dx}. \tag{2.10}$$

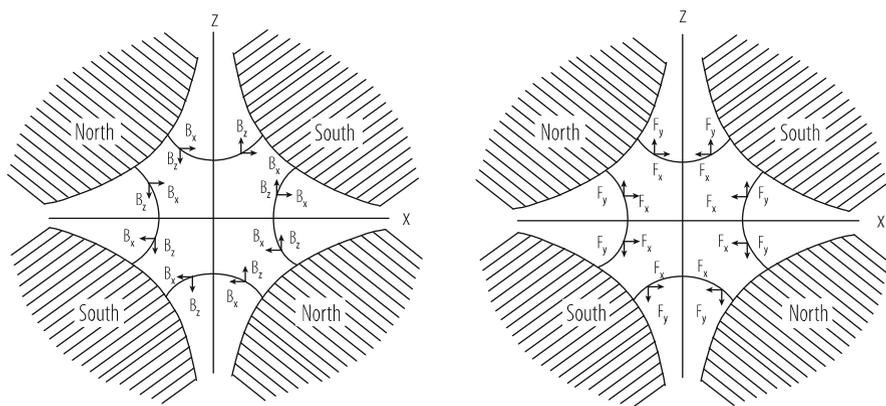


Fig. 2.7 Components of field and force in a magnetic quadrupole. Positive ions approach the reader on paths parallel to the s axis (orthogonal to x and z) [2]

The angular deflection given to a particle passing through a short quadrupole of length, ℓ and strength k , at a displacement x is therefore:

$$\Delta x' = \theta = \ell B' / (B\rho) = \ell B' x / (B\rho) = \ell k x. \quad (2.11)$$

The use of x' to indicate the divergence angle of a trajectory is defined in Fig. 2.5. Compare this with a converging lens in optics:

$$\Delta x' = -x/f \quad (2.12)$$

and we see that the focal length of a horizontally focusing quadrupole is

$$f = -1 / (k\ell) \quad (2.13)$$

The particular quadrupole shown in Fig. 2.7 would focus positive particles coming out of the paper or negative particles going into the paper in the horizontal plane. A closer examination reveals that such a quadrupole deflects particles with a vertical displacement away from the axis—vertical displacements are defocused. This can be seen if Fig. 2.7 is rotated through 90° .

2.1.8 The Equation of Motion

Earlier we derived an expression for the change in divergence of a particle passing through the quadrupole. A horizontally focusing quadrupole (which is at the same time vertically defocusing) has a negative k .

We first look at the vertical plane. The angular deflection given to a particle passing through a short quadrupole of length ds and strength k at a displacement z is therefore:

$$dz' = -kz ds. \quad (2.14)$$

From this we can deduce a differential equation for the motion

$$z'' + k(s)z = 0. \quad (2.15)$$

Here we would like to make a clear statement: While inside a lattice element, say a quadrupole lens, the normalised gradient k is constant and we get an equation that we know from Hook's law in classical mechanics, (see Eq. 2.12), the situation now is more general. We allow $k(s)$ to change, while our particles are running through the accelerator. The corresponding equation (2.15) is called Hill's Equation, a second order linear equation with a periodic coefficient, $k(s)$ which describes the distribution of focusing strength around the ring. The above form of Hill's equation

applies to the motion in the vertical plane while in the horizontal plane the effect of the dipole magnets has to be included:

$$x'' + \left[\frac{1}{\rho^2(s)} - k(s) \right] x = 0. \quad (2.16)$$

Here the sign in front of $k(s)$ is reversed so that the quadrupole focuses. The extra focusing term $1/\rho^2$ due to the curvature of the orbit can be significant in small rings. In the old constant gradient synchrotrons, this weak focusing term was the only form of horizontal focusing.

We see in Fig. 2.10, the pattern of one cell of a simple synchrotron lattice—a pattern which is repeated many times around the circumference as may be seen in Fig. 2.11 which shows—in addition to the focusing and defocusing lenses also the bending magnets—bending magnets. Within this pattern of dipole and quadrupole focusing and defocusing (F and D), particles make betatron oscillations within the envelopes described by β_x and β_z , or more precisely, the square roots of these quantities (here we use the variable y to represent either the horizontal or the vertical coordinate, x or z)

$$y = \sqrt{\varepsilon\beta(s)} \sin(\phi(s) + \phi_0). \quad (2.17)$$

If one tries to verify that this is the solution of Hills Equation an important and necessary condition emerges:

$$\phi' = 1/\beta \quad (2.18)$$

From which we see that $2\pi\beta$ is the local wavelength of the transverse oscillations.

2.1.9 Matrix Description

Usually in alternating gradient (AG) machines, the ring is a repetitive pattern of focusing fields that we call the “lattice”. Each lattice element may be expressed by a matrix and whole sections of the ring which transport the beam from place to place may be represented as the product matrix of the single element matrices involved, which makes the description of particle trajectories very simple and very elegant at the same time. Any linear differential equation, like Hill’s Equation, has solutions which can be traced from one point, s_1 , to another, s_2 , by a 2×2 matrix, the transport matrix:

$$\begin{pmatrix} y(s_2) \\ y'(s_2) \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} y(s_1) \\ y'(s_1) \end{pmatrix} = M_{21} \begin{pmatrix} y(s_1) \\ y'(s_1) \end{pmatrix}. \quad (2.19)$$

The transport matrix M_{21} has a rather simple form for each focusing quadrupole that the particle encounters and for the drift length between quadrupoles and it is easy to compute the four elements numerically once we define the length and focusing strength. We can trace particles by simply forming the product of these elementary matrices. But there is also a general relation between the elements a , b , c , d and the amplitude and phase of transverse motion between any two points. Each term in M_{21} can be expressed as a function of $\beta(s)$ and $\phi(s)$. The functions of $\beta(s)$ and $\phi(s)$ may be calculated by comparing the numerical result of multiplying the individual matrices for quadrupoles and drift lengths with what we know must be the general form of each element.

As a first step, we derive the general form of a periodic transport matrix.

To simplify the notation we drop the explicit dependence of β and ϕ on s from the expressions—we will just have to remember that they vary with s . We also introduce a new quantity:

$$w = \sqrt{\beta}. \quad (2.20)$$

just to avoid too many terms in what follows.

In this new notation we can write the solution of the Hill Equation:

$$y = \varepsilon^{1/2} w \cos(\varphi + \phi_0). \quad (2.21)$$

Taking the derivative and substituting $\varphi' = 1/\beta = 1/w^2$ we have:

$$y' = \varepsilon^{1/2} w' \cos(\varphi + \phi_0) - \frac{\varepsilon^{1/2}}{w} \sin(\varphi + \phi_0). \quad (2.22)$$

Next we substitute these explicit expressions for y and y' in both sides of the matrix equation. We do this first with the initial condition $\varphi_0 = 0$, this is the so-called ‘cosine’ solution, and then we do it again for the ‘sine’ solution with $\varphi_0 = \pi/2$. This is exactly equivalent to tracing the paraxial and central rays through an optical lens. We write $\phi_2 - \phi_1 = \phi$ for each case. Each of the two solutions give us two equations for y and y' and thus we obtain four simultaneous equations which can be solved for a , b , c , d in terms of w , w' , and φ . The result is the most general form of the transport matrix between the positions s_1 and s_2 :

$$M_{12} = \begin{pmatrix} \frac{w_2}{w_1} \cos \varphi - w_2 w_1' \sin \varphi & w_1 w_2 \sin \varphi \\ -\frac{1+w_1 w_1' w_2 w_2'}{w_1 w_2} \sin \varphi - \left(\frac{w_1'}{w_2} - \frac{w_2}{w_1} \right) \cos \varphi & \frac{w_1}{w_2} \cos \varphi + w_1 w_2' \sin \varphi \end{pmatrix}. \quad (2.23)$$

This rather formidable looking expression simplifies a lot, if we refer to a full circle, in other words, if we restrict M to apply between two identical points in successive turns or cells of a periodic structure. Then $w_2 = w_1$, $w_2' = w_1'$,

and φ to become μ , the phase advance per cell. The matrix for one period is now:

$$M = \begin{pmatrix} \cos \mu - ww' \sin \mu & w^2 \sin \mu \\ -\frac{1+w^2 w'^2}{w^2} \sin \mu & \cos \mu + ww' \sin \mu \end{pmatrix}. \quad (2.24)$$

Next we invent some new functions of β :

$$\begin{aligned} \beta &= w^2, \\ \alpha &= -ww' = -\frac{\beta'}{2}, \\ \gamma &= \frac{1+(ww')^2}{w^2} = \frac{1+\alpha^2}{\beta}. \end{aligned} \quad (2.25)$$

These functions (which are not the same as the parameters used in special relativity!) are a complete and compact description of the dynamics. The matrix now becomes even simpler:

$$M = \begin{pmatrix} \cos \mu + \alpha \sin \mu & \beta \sin \mu \\ -\gamma \sin \mu & \cos \mu - \alpha \sin \mu \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \quad (2.26)$$

This is the Twiss matrix. It is the basic matrix for periodic lattices and should be memorized.

We can imagine that if we can only find an independent way of computing the numerical values of the four elements we can solve and find:

$$\begin{aligned} \cos \mu &= (\text{Tr } \mathbf{M}) / 2 = (a + d) / 2, \\ \beta &= b / \sin \mu > 0, \\ \alpha &= (a - b) / (2 \sin \mu), \\ \gamma &= -c / \sin \mu. \end{aligned} \quad (2.27)$$

These Twiss parameters, μ , β , α , and γ , are therefore rigorously determined by the overall effect of the focusing properties of the lattice elements. Still, they vary around the ring and apply to the point chosen in the period as a starting and finishing point. We shall see that each individual component, quadrupole, dipole, or drift space in the ring has its own matrix and this provides the independent method of calculation. We must first choose the starting point, the location, s , where we wish to know β and the other Twiss parameters. By starting there and multiplying the element matrices together for one turn we are able to find a , b , c , d numerically for that location. We can then apply the above four equations to find the Twiss matrix. If the machine has a natural symmetry in which there are a number of identical periods, it is sufficient to do the multiplication up to the corresponding point in the next period. The values of α , β , and γ would be the same if we went on for the whole ring. By choosing different starting points we can trace $\beta(s)$ and $\alpha(s)$. We now give the matrices for the three basic lattice elements.

2.1.10 Transport Matrices for Lattice Components

An empty space or drift length is the simplest of the lattice component matrices. Figure 2.8(a) shows the analogy between a particle trajectory and a diverging ray in optics. The angle of the ray and the divergence of the trajectory are related:

$$\theta = \tan^{-1}(x'). \quad (2.28)$$

The effect of a drift length in phase space is a simple horizontal translation from (x, x') to $(x + \ell x', x')$ and can therefore be written as a matrix:

$$\begin{pmatrix} x_2 \\ x'_2 \end{pmatrix} = \begin{pmatrix} 1 & \ell \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x'_1 \end{pmatrix}. \quad (2.29)$$

The next case is that of a thin quadrupole magnet of infinitely small length but finite integrated gradient:

$$\ell k = \frac{1}{(B\rho)} \frac{\partial B_z}{\partial x}. \quad (2.30)$$

The optical analogy of a thin quadrupole with a converging lens is illustrated in Fig. 2.8(b). A ray, diverging from the focal point arrives at the lens at a displacement, x , and is turned parallel by a deflection:

$$\theta \approx \frac{1}{f} x. \quad (2.31)$$

This deflection will be the same for any ray at displacement x irrespective of its divergence. This behaviour can be expressed by a simple matrix, the thin lens matrix:

$$\begin{pmatrix} x_2 \\ x'_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x'_1 \end{pmatrix}. \quad (2.32)$$

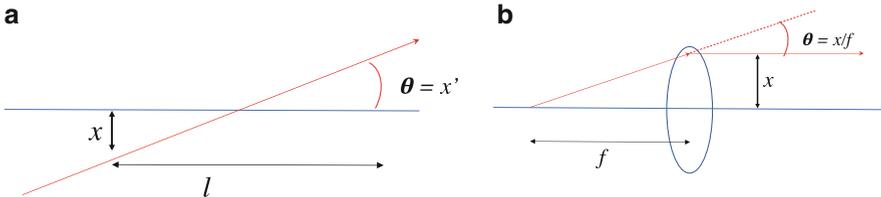


Fig. 2.8 The effect of a drift—(a), left side—and a focusing quadrupole lens—(b) right side—on a particle trajectory. The mathematical expressions are given in Eqs. (2.29) and (2.32)

A particle arriving at a quadrupole lens at a displacement x obeys Hill's equation

$$x'' + kx = 0. \quad (2.33)$$

Hence the small deflection θ is just:

$$\Delta x' = -kx\ell. \quad (2.34)$$

Comparing quadrupoles with optical lenses we remember that $\ell k = 1/f$ and is the power of the lens and that the matrix, for a thin lens, can be written:

$$\begin{pmatrix} 1 & 0 \\ -k\ell & 1 \end{pmatrix}. \quad (2.35)$$

Under the influence of these focusing and defocusing fields, a particle trajectory will finally look like a more or less zig-zag shaped curve; which for the example of eight regular cells it is shown in Fig. 2.9.

Quadrupoles are sometimes not short compared to their focal length. One must therefore use the matrices for a long quadrupole when one comes to compute the final machine:

$$\begin{aligned} M_F &= \begin{pmatrix} \cos \ell\sqrt{k} & \frac{1}{\sqrt{k}} \sin \ell\sqrt{k} \\ -\sqrt{k} \sin \ell\sqrt{k} & \cos \ell\sqrt{k} \end{pmatrix}, \text{ and} \\ M_D &= \begin{pmatrix} \cosh \ell\sqrt{k} & \frac{1}{\sqrt{k}} \sinh \ell\sqrt{k} \\ -\sqrt{k} \sinh \ell\sqrt{k} & \cosh \ell\sqrt{k} \end{pmatrix}. \end{aligned} \quad (2.36)$$

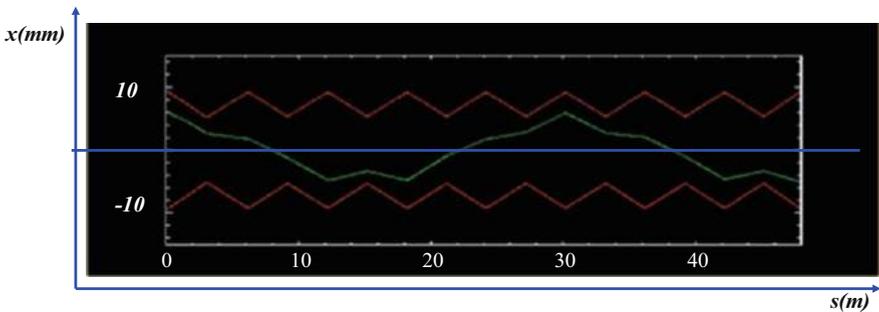


Fig. 2.9 A single particle trajectory in a ring: At each part of the lattice the amplitude and angle, (x, x') of the particle are described by a matrix transformation, according to Eq. (2.32). The blue line corresponds to an ideal particle, with $x = x' = 0$ and so refers to the ideal orbit

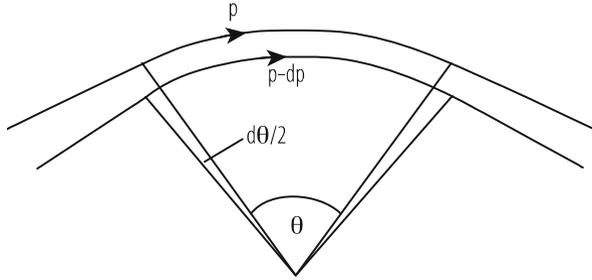


Fig. 2.10 The focusing effect of trajectory length in a pure sector dipole magnet

We can compare this with the solutions of Hill's equations within F and D quadrupoles:

$$\begin{aligned} z &= \cos \sqrt{k} \ell z_0 + \frac{1}{\sqrt{k}} \sin \sqrt{k} \ell z'_0, \\ x &= \cosh \sqrt{k} \ell x_0 + \frac{1}{\sqrt{k}} \sinh \sqrt{k} \ell x'_0. \end{aligned} \quad (2.37)$$

We have so far ignored the bending that takes place in dipole magnets and these may be thought of as drift lengths in a first approximation. An exact calculation should include the focusing effect of their ends. A pure sector magnet, whose ends are normal to the beam will give more deflection to a ray which passes at a displacement x away from the centre of curvature (Fig. 2.10). This particle will have a longer trajectory in the magnet. The effect is exactly like a lens which focuses horizontally but not vertically. The matrices for a sector magnet are:

$$\begin{aligned} M_H &= \begin{pmatrix} \cos \theta & \rho \sin \theta \\ -(1/\rho) \sin \theta & \cos \theta \end{pmatrix}, \\ M_V &= \begin{pmatrix} 1 & \rho \theta \\ 0 & 1 \end{pmatrix}. \end{aligned} \quad (2.38)$$

Some bending magnets are not sector magnets as in Fig. 2.9, but have end faces which are parallel. It is easier to stack laminations this way than on a curve. The entry and exit angles are therefore, $\theta/2$, and the horizontal focusing effect is reduced but there is an additional focusing effect for a particle whose trajectory is displaced vertically. In the computer model one may convert a pure sector magnet into a parallel faced magnet by simply adding two thin lenses at each face. They are horizontally defocusing and vertically focusing and their strength is:

$$k \ell = -\frac{\tan(\theta/2)}{\rho}. \quad (2.39)$$

Unlike early lattice designers we have computers to help when we come to multiply these elements together to form the matrix for a ring or a period of the lattice [3–5]. A lattice program such as MAD [6] does all the matrix multiplication to obtain (a, b, c, d) from each specified point, s , and back again. It prints out β and φ and other lattice variables in each plane, and we can plot the result to find the beam envelope around the machine. This is the way machines are designed. Lengths, gradients, and numbers of FODO normal periods are varied to match the desired beam sizes and Q values.

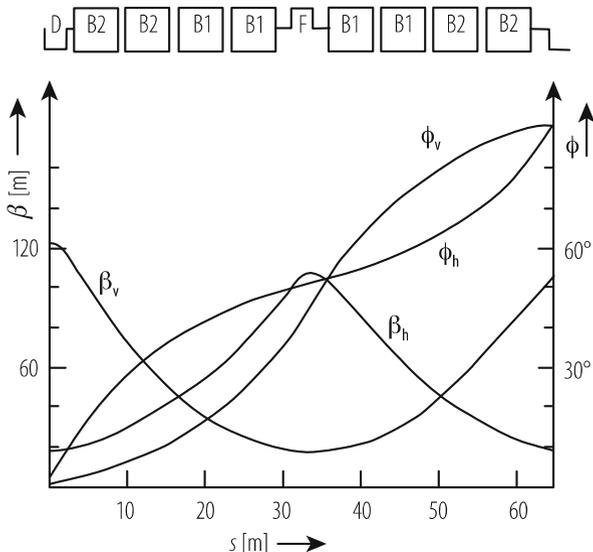
In Fig. 2.8 we saw the trajectory of a particle, oscillating in a pattern of alternating focussing and defocusing quadrupoles (FODO). The trajectories in general all lie within an envelope which has the general features of the optical model in Fig. 2.6. If we were to repeat the observation of the displacement and divergence of a particle on successive turns we would find the elliptical locus of its motion (Fig. 2.5). The aspect ratio of this ellipse would depend upon where in the ring we choose to make the observation. The ellipse would be squat near D lenses and elongated near F 's. The figure would appear just the same if we were to plot it between what are F quadrupoles in the vertical plane. Of course, the whole pattern of quadruples and the envelope is shifted by the distance between adjacent quadrupoles because F -quadrupoles in one plane are D in the other (et vice versa).

2.1.11 The Betatron Envelopes

To recapitulate, a modern synchrotron consists of pure bending magnets and quadrupole magnets or lenses which provide focusing. These are interspersed among the bending magnets of the ring in a pattern called the lattice. By suitable choice of strength and spacing of the lenses the envelope function $\beta(s)$ can be made periodic in such a way that it is large at all F quadrupoles and small at all D 's. Symmetry will ensure this is true also in the vertical plane. Particles oscillating within this envelope will always tend to be further off axis in F quadrupoles than in D quadrupoles and there will therefore be a net focusing action. We have already seen that β is the aspect ratio of the phase space ellipse (see also [7, 8]).

In Fig. 2.11 we see an example of such a magnet pattern which is one cell, or about 1% of the circumference, of the 400 GeV SPS at CERN. Although the SPS is now considered a rather old fashioned machine its simplicity leads us to use it as an example. The focusing structure is FODO and in this pattern half of the quadrupoles (F) focus, while the other half, defocus (D) the beam. Bending magnets, which in a first approximation do no focussing are represented together with other non-focussing elements by the letter “O”. The envelope of these oscillations follows a function $\beta(s)$ which has waists near each defocusing magnet and has a maximum at the centres of F quadrupoles. Since F quadrupoles in the horizontal plane are D quadrupoles vertically, and vice versa, the two functions $\beta_h(s)$ and $\beta_v(s)$ are one half-cell out of register in the two transverse planes. The function β has the dimensions of length but the units bear no relation at this stage to physical beam

Fig. 2.11 One cell of the CERN SPS representing 1/108 of the circumference. The pattern of dipole (B) magnets and quadrupole (*F* and *D*) lenses is shown above



size. The reader should be persuaded that particles do not follow the $\beta(s)$ curves but oscillate within them in a form of modified sinusoidal motion whose phase advance is described by $\phi(s)$. The phase change per cell in the example shown is close to $\pi/2$ but the rate of phase advance is modulated throughout the cell.

2.2 Coupling

Until now we have considered the motion in the vertical and horizontal direction to be orthogonal and independent. This is the ideal case. Now we look at what happens when there is a skew quadrupole or solenoidal field in the machine which couples horizontal motion into vertical and vice versa. This is rather a special case affecting mainly electron synchrotrons and the reader may choose to skip to Sect. 2.3 and leave coupling to a second reading.

In a fully coupled machine the betatron oscillations in the two transverse directions are like two harmonic oscillators which transfer energy from one to the other with a frequency which is just the difference between their Q 's. They act like coupled pendula. In this way all the horizontal “emittance” can add to the vertical emittance and the beam exceeds the available vertical aperture.

The phenomenon is particularly important in electron rings. The electron beam would damp to zero emittance were it not for quantum emission in the horizontal plane exciting betatron oscillations. There is no comparable excitation in the vertical plane and only coupling of the horizontal oscillations into the vertical plane gives

the beam any vertical dimensions. Vertical emittance, and the magnet gap needed to accept it, is directly proportional to coupling.

2.2.1 Coupling Fields

There are two principal configurations of field which excite coupling. The first we shall consider is a skew quadrupole, i.e. a quadrupole whose poles lie symmetrically in the horizontal and vertical planes (Fig. 2.12).

A particle with horizontal position x , experiences not a B_z as would be the case in a normal quadrupole and which would change its x' , but a B_x which together with the paraxial velocity deflects vertically in the direction of $\mathbf{v} \times \mathbf{B}$. Of course once the particle has acquired a vertical displacement z after a number of turns it experiences a vertical field, for in a skew quadrupole the field is:

$$\begin{aligned} B_x / (B\rho) &= kx, \\ B_z / (B\rho) &= -kz. \end{aligned} \tag{2.40}$$

Thus a horizontal displacement couples into the vertical plane leading to a vertical divergence and displacement. The vertical displacement goes on to couple back into the horizontal plane modifying the horizontal displacement and divergence—and so it proceeds transferring transverse momentum back and forth from one plane to the other.

A solenoid is the other field configuration that can couple the two planes but this kind of coupling is less important in synchrotrons and we leave it to the reader to consult a more exhaustive treatment of coupling in [9].

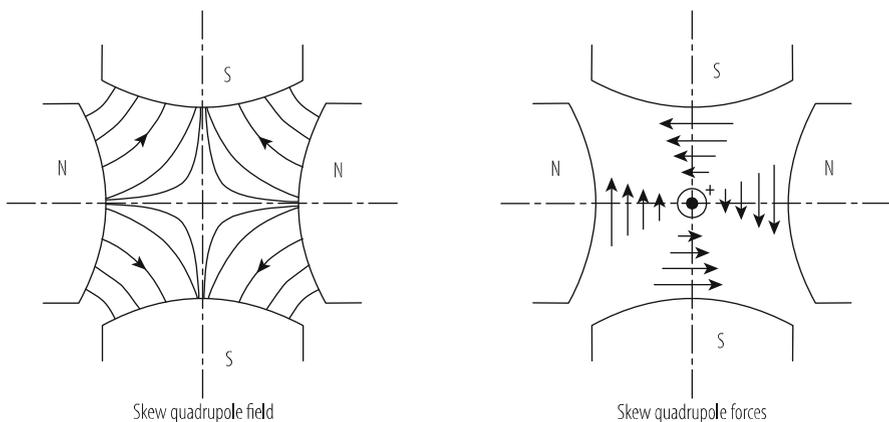


Fig. 2.12 The magnetic field and force in a skew quadrupole

2.2.2 Qualitative Treatment of Coupling

In our treatment the theory is deliberately simplified to reveal the physical mechanisms at work. We assume that the coupling is driven by a single skew quadrupole at the centre of one of the existing lattice machine quadrupoles where β is maximum and its derivative zero. We ignore the changes in betatron phase of one plane with respect to the other within a single turn.

The skew quadrupole gradient is normalized:

$$k = \frac{1}{(B\rho)} \left(\frac{\partial B_x}{\partial x} \right)_{z=0}, \tag{2.41}$$

$l = \text{length of the quadrupole.}$

Figure 2.13 on the left shows the betatron motion in the horizontal plane. We have normalized the elliptical phase space trajectory into a circle at the location of the skew quadrupole by multiplying the divergence by β_x . On the right we have done the same for the vertical plane. The angular kick Δp , on passing the skew quadrupole is calculated from a similar diagram for the vertical plane and

$$\Delta p_x = \beta_x k l w \cos Q_V \theta, \tag{2.42}$$

where $w = \sqrt{\varepsilon_V \beta_z}$ is the radius of the circle for vertical motion, and $u = \sqrt{\varepsilon_H \beta_x}$ is the radius horizontally.

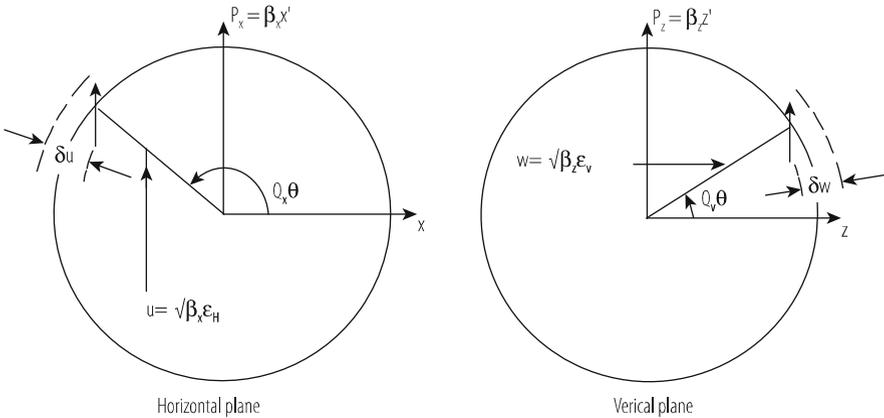


Fig. 2.13 Phase space diagram

The kick, projected as an amplitude increment becomes:

$$\delta u = w\beta_x k\ell \sin Q_H\theta \cos Q_V\theta. \quad (2.43)$$

When we use:

$$\sin A \cos B = \frac{1}{2} \sin(A - B) + \frac{1}{2} \sin(A + B)$$

and ignore the second, high frequency term; we obtain the coupled equations for a single passage:

$$\begin{aligned} \frac{\delta w}{w} &= -\sqrt{\frac{\varepsilon_H}{\varepsilon_V}} \sqrt{\frac{\beta_x\beta_z}{2}} k\ell \sin(Q_H - Q_V)\theta, \\ \frac{\delta u}{u} &= \sqrt{\frac{\varepsilon_V}{\varepsilon_H}} \sqrt{\frac{\beta_x\beta_z}{2}} k\ell \sin(Q_H - Q_V)\theta. \end{aligned} \quad (2.44)$$

These are incremental equations which we must sum over the n turns as the coupling enhances u at the expense of w .

Figure 2.14 shows diagrammatically the coupled motion. The vertical betatron amplitude decrease from $w + \Delta w$ to w in one quarter period of the slow oscillation which takes $1/4|Q_H - Q_V|$ turns. The mean value of the cosine is taken as $2/\pi$. We then arrive at the expressions for the maximum excursions in amplitude:

$$\begin{aligned} \frac{\Delta w}{w} &= \sqrt{\frac{\varepsilon_H}{\varepsilon_V}} \frac{\sqrt{\beta_x\beta_z}}{4\pi|Q_H - Q_V|} k\ell, \\ \frac{\Delta u}{u} &= \sqrt{\frac{\varepsilon_V}{\varepsilon_H}} \frac{\sqrt{\beta_x\beta_z}}{4\pi|Q_H - Q_V|} k\ell. \end{aligned} \quad (2.45)$$

We now move from the phase plane into real space. Some machines were designed to have a rectangular ‘‘vacuum chamber’’ which would accept particles which simultaneously have large horizontal and vertical ‘‘emittances’’. In this sense emittance is defined for a single particle

$$\varepsilon_H = \pi u^2/\beta_x, \quad \varepsilon_V = \pi w^2/\beta_y. \quad (2.46)$$

In the presence of coupling, the particle motion is a series of Lissajous figures filling the rectangular cross-section but always touching it somewhere on each turn (Fig. 2.15). It is inevitable therefore that if coupling increases either amplitude by $\Delta u/u$ or $\Delta w/w$, some fraction of particles will be lost.

A rigorous treatment of coupling is too lengthy to include here but the reader may consult [9–11] for a complete description. This lengthier treatment leads to a model in which the modes of betatron oscillations are no longer about the vertical and horizontal planes but about two orthogonal principal planes inclined with respect to the vertical and horizontal frame.

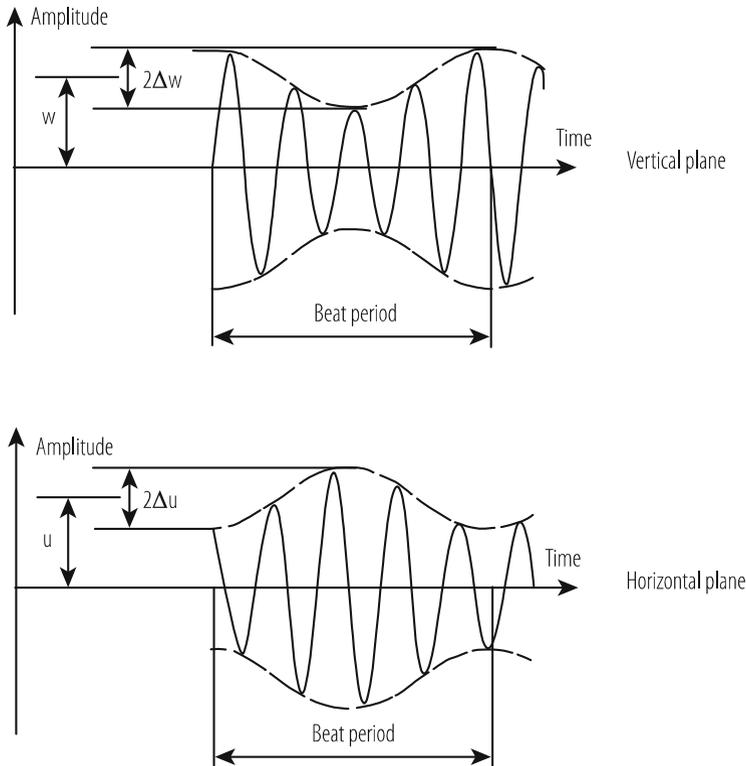


Fig. 2.14 Coupled betatron oscillations for $1/|Q_H - Q_V|$ turns

2.3 Liouville’s Theorem

Now let us return to the ‘mainstream’ of transverse dynamics. Liouville’s theorem is a conservation law that applies to the area occupied by a number of particles plotted in phase space.

We should think of a beam of particles as a cloud of points within a closed contour in a transverse phase space diagram (Fig. 2.16). Liouville’s theorem tells us that this area within the contour is conserved. The contour is usually, but not always, an ellipse. In Fig. 2.5 we came across such an elliptical contour—the locus of a particle’s motion plotted in phase space (x, x') and we call its area, the emittance. We could also think of it as a limiting contour enclosing all the particles in the beam which we would again call the emittance—not of the particle but of the beam as a particle ensemble.

We express beam emittance in units of π mm-milliradians. According to Liouville the emittance area will be conserved as the beam passes down a transport line or circulates in a synchrotron whatever magnetic focusing or bending operation we do on the beam—provided that only conservative forces are taken into account.

Fig. 2.15 Particle lost due to coupling

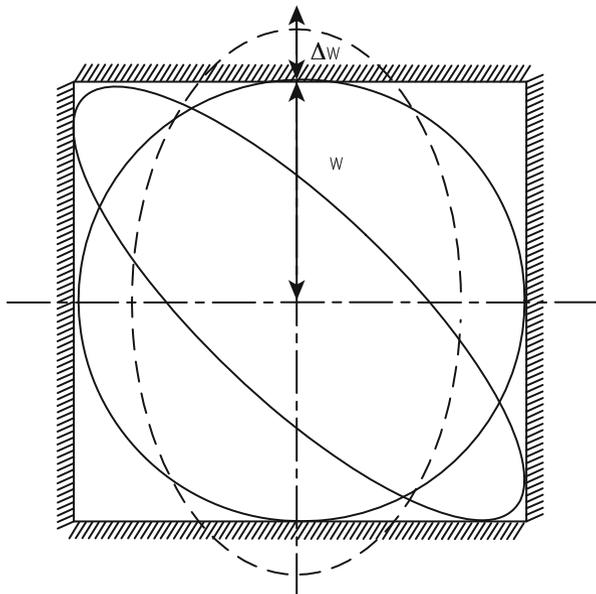
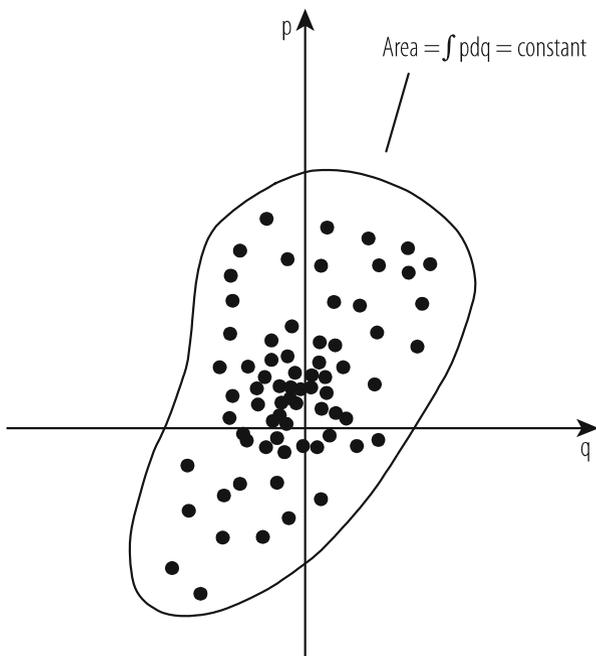


Fig. 2.16 Liouville's theorem applies to this contour



Even though the ellipse may appear to have many shapes around the accelerator its phase space area will not change (Fig. 2.17). The aspect ratio of the ellipse will change however. At a narrow waist, near a *D* quadrupole (a) in Fig. 2.17, its

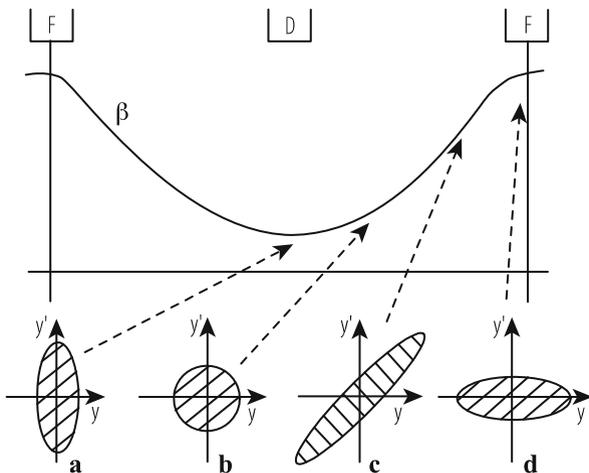


Fig. 2.17 How the conserved phase space appears at different points in a FODO cell

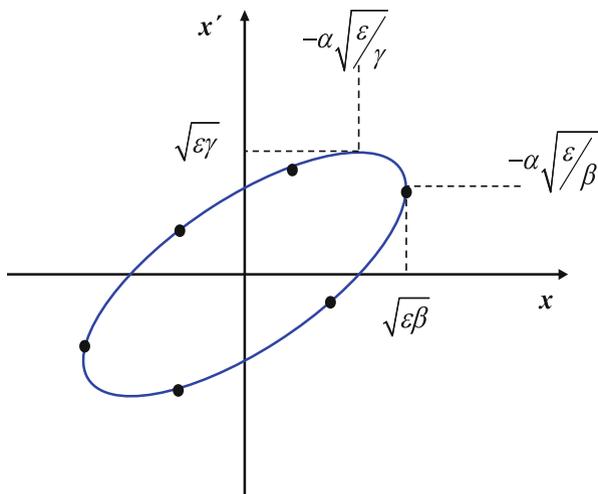


Fig. 2.18 The parameters of a phase-space ellipse containing an emittance ϵ at a certain point in the lattice. The shape and orientation of the ellipse are determined by the Twiss parameters at the given location

divergence will be large, while in an F quadrupole (d) where the betatron function is maximum, its divergence will be small. The beam is also seen at a broad waist or maximum in the beta function and a place where the beam is diverging.

In Fig. 2.18 we see how the various features of the ellipse are related to the Twiss parameters. The equation of the ellipse, often called the Courant and Snyder

invariant, has the form

$$\gamma(s)y^2 + 2\alpha(s)yy' + \beta(s)y'^2 = \varepsilon. \quad (2.47)$$

Here y is used to mean either of the transverse displacements, x or z . It is straight forward determine the relation between the shape and orientation of the (x, x') ellipse and the Twiss parameters α , β , γ as indicated in Fig. 2.18.

The invariance of the (x, x') space area, as we move to different points in the ring is an alternative statement of Liouville's theorem.

A word of caution—another, stricter, version of Liouville's theorem states that:

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

This rules out the application of Liouville's theorem to situations in which space charge forces within the beam play a role or when there is a velocity dependent effect such as when particles emit synchrotron light. However we may apply Liouville to proton beams which do not normally emit synchrotron light and to electrons travelling for a few turns in a synchrotron. This is usually too short a time for electrons to emit enough synchrotron light energy to affect their transverse motion.

Liouville's theorem does not apply as a proton beam is accelerated. Observations tell us this is not the case. The beam appears to shrink. This is because the co-ordinates we have used so far, y and y' , are not 'canonical' in the sense defined by Hamiltonian in his mechanics, which is part and parcel of Liouville's mathematical theory of dynamics. We should therefore express emittance in Hamilton's canonical phase space and relate this carefully to the co-ordinates, displacement, y , and divergence, y' , which we have been using so far. We can then define an emittance which is conserved even as we accelerate.

We shall have to be particularly careful not to confuse Twiss parameters, and the parameters of special relativity: In special relativity we use β as the ratio of the particles velocity and the speed of light and the Lorentz factor γ describes the total energy divided by the rest energy. The reader will have to examine the context to be sure. For those who have not met Hamiltonian mechanics, it is sufficient to know that the canonical co-ordinates of relativistic mechanics are:

$$p = \frac{m_0 \dot{y}}{\sqrt{1 - v^2/c^2}}, q = y. \quad (2.48)$$

Here q or y is a general transverse co-ordinate, p its conjugate momentum and we define β and γ when used in the context of special relativity to be:

$$\begin{aligned} \beta &= v/c, \\ \gamma &= 1/\sqrt{1 - \beta^2}, \\ m_0 &= \text{rest mass}, \\ c &= \text{velocity of light.} \end{aligned} \quad (2.49)$$

We may find the relationship between canonical momentum and divergence from the substitution:

$$p_y = m_0 \frac{dy}{dt} \gamma = m_0 \frac{ds}{dt} \frac{dy}{ds} \gamma = mc (\beta\gamma) y'. \quad (2.50)$$

Writing Liouville's Theorem expressed in canonical coordinates we can use the above expression to define a conserved quantity and relate it to the area in (y, y') space

$$\int p_y dy = m_0 c (\beta\gamma) \int y' dy = p_0 \int y' dy \quad (2.51)$$

where p_0 is the momentum in the direction of motion of the particle.

This invariant is the emittance, ε , of our transverse phase space multiplied by the relativistic $\beta\gamma$ which is proportional to momentum. Accelerator physicists often call this the invariant or 'normalised' emittance:

$$\varepsilon^* = \beta\gamma\varepsilon \quad [\pi \text{ mm} \cdot \text{mrad}] \quad (2.52)$$

This normalised emittance, ε_* , is conserved as acceleration proceeds in a synchrotron and the physical emittance within the right-hand side of the equation must fall inversely with momentum if the whole term is to be conserved. Close to the velocity of light this implies that it is inversely proportional to energy.

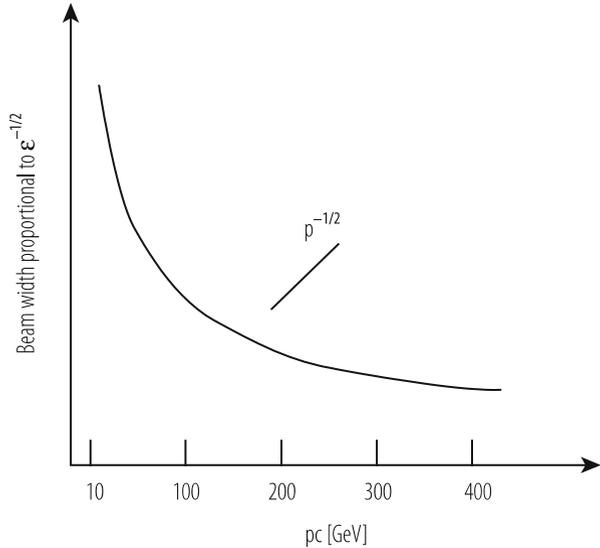
$$\text{Emittance} = \pi\varepsilon = \int y' dy = \pi\varepsilon^* / (\beta\gamma) \propto 1/p_0. \quad (2.53)$$

We therefore expect the beam dimensions to shrink as (Fig. 2.19) a phenomenon called 'adiabatic damping'.

2.3.1 Chains of Accelerators

As a consequence of the adiabatic shrinking, the beam emittance is largest at low energy, and so is the beam dimension. Proton accelerators need their full aperture at injection and it is then that their design is most critical. For this reason it is economic to split a single large ring into a chain of accelerators—the smaller radius rings having a large aperture while the higher energy rings with large radius can have smaller apertures. In these chains of proton accelerators, such as the CERN accelerator complex, Linac – Booster – PS – SPS, the invariant emittance, determined by the parameters of the beam as it leaves the ion source at the beginning of the linac, may be conserved to several hundred GeV. Of course one must guard against mismatches between machines or non-linear fields which dilate the emittance.

Fig. 2.19 Adiabatic shrinking of the beam as function of the beam momentum during acceleration



2.3.2 Exceptions to Liouville's Theorem

The invariance of normalised emittance of a proton beam and the shrinking of its physical emittance with energy is quite the opposite of what happens in an electron machine. Liouville's theorem only applies to conservative systems, where particles are guided by external fields and not to electron machines where particles emit some of their own energy. Electrons, being lighter than protons and hence more relativistic, emit quanta of radiation as they are accelerated. This quantised emission causes particles to jump around in momentum, leading to changes in the trajectories amplitude and angle. These changes couple into both planes of transverse phase space. At the same time, there is a steady tendency for particles near the edge of the emittance to lose transverse energy and fall back towards the centre. In an electron machine the emittance is determined not by the Liouville but by the equilibrium between these two effects. In fact, it grows with E^2 .

Consider a number of protons which have the maximum amplitude present in the beam. They follow trajectories at the perimeter of the ellipse but at any instant have a random distribution of initial phases ϕ_0 . If we were able to measure y and y' for each and plot them in phase space, they would lie around the ellipse of area $\pi \epsilon$ and their co-ordinates would lie in the range of

$$\begin{aligned} -\sqrt{\beta\epsilon} &\leq y \leq \sqrt{\beta\epsilon}, \\ -\sqrt{\epsilon\gamma} &\leq y' \leq \sqrt{\epsilon\gamma}. \end{aligned} \quad (2.54)$$

Particles in a beam are usually distributed in a population which appears Gaussian when projected on a vertical or horizontal plane. In a proton machine

the emittance boundary used to be conventionally chosen to be that of a proton with amplitude 2σ . This would include about 90% (strictly 87%) of a Gaussian beam where σ is the standard distribution. In an electron machine a 2σ boundary would be too close to the beam and an aperture stop placed at this distance would rather rapidly absorb most of the beam as particles redistribute themselves, moving temporarily into the tails due to quantum emission and damping. The safe physical boundary for electrons depends on the lifetime required but is in the region of 6σ to 10σ . The emittance which is normally quoted for an electron beam corresponds to an electron with the amplitude of σ in the Gaussian projection. We are then free to choose how many σ 's we must allow. There is consequently a factor 4 between emittance defined by electron and proton experts.

2.4 Momentum Dependent Transverse Motion

In the previous chapters, we have studied the motion of a particle as it swings from side to side about the ideal orbit around the synchrotron: the transverse motion. Nothing is perfect, however, and so we cannot assume that each and every proton in a large ensemble of up to 10^{11} particles will have exactly the ideal momentum. Instead we expect a certain momentum spread in the beam and therefore we have to study how transverse motion depends on small departures, $\Delta p/p_0$, from the synchronous momentum p_0 .

2.4.1 Dispersion

The central closed orbit of a synchrotron is matched to an ideal (synchronous) momentum p_0 . A particle of this momentum and of zero betatron amplitude will pass down the centre of each quadrupole, be bent by exactly 2π by the bending magnets in one turn of the ring and remain synchronous with the r.f. frequency. Its path is called the central (or synchronous) momentum closed orbit. In Fig. 2.8 this ideal orbit is the horizontal axis and we see particles executing betatron oscillations about it but these oscillations do not replicate every turn. The synchronous orbit, however, closes on itself so that x and x' remain zero.

We now consider a closed orbit which is distorted in the horizontal plane by non-ideal bends in the dipole. Figure 2.20 shows a particle with a lower momentum $\Delta p/p < 0$ and which is bent horizontally more in each dipole of a FODO lattice. We could argue that the total deflection, being more than 2π would cause it to spiral inwards and hit the vacuum chamber wall. On the other hand there is a closed orbit for this lower momentum in which the extra bending forces are compensated by extra focusing forces as the orbit is displaced inwards in the F quadrupoles and less so in the defocusing in the D 's (Fig. 2.20). We may describe the shape of this new closed orbit for a particle of unit $\Delta p/p$ by a *dispersion* function $D(s)$. The

Fig. 2.20 The extra inward force given to a low momentum particle by the dipoles is balanced by the focusing the quadrupoles and defines a dispersion function

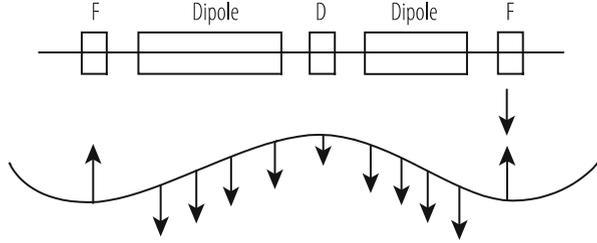
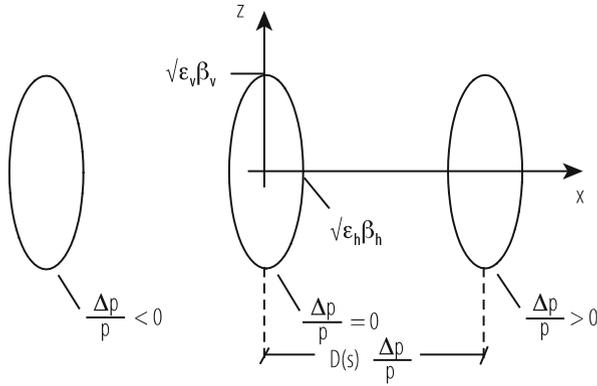


Fig. 2.21 The beam cross sections in real space for beams of three different momenta at a point where the dispersion function is large



displacement of the closed orbit is:

$$x(s) = D(s) \frac{\Delta p}{p_0}. \tag{2.55}$$

In Fig. 2.21 we see how the effect of dispersion for off momentum orbits adds to the betatron motion to widen the beam section. The betatron motion of each of the three particles: $\Delta p/p < 0$, $\Delta p/p = 0$, and $\Delta p/p > 0$, is within an ellipse in physical (x, z) space. The ellipses for each momentum are separated by a distance $D(s)\Delta p/p$. The semi-aperture required will be:

$$a_v = \sqrt{\beta_v \epsilon_v}, a_h = \sqrt{\beta_h \epsilon_h} + D(s) \frac{\Delta p}{p}. \tag{2.56}$$

2.4.2 Chromaticity

This effect is equivalent to the chromatic aberration in a lens. It is defined as a quantity Q' :

$$\Delta Q = Q' \frac{\Delta p}{p}. \tag{2.57}$$

The chromaticity [12] arises because the focusing strength of a quadrupole has $(B\rho)$ in the denominator and is therefore inversely proportional to momentum:

$$k = \frac{1}{(B\rho)} \frac{dB_z}{dx}. \quad (2.58)$$

A small spread in momentum in the beam, $\pm \Delta p/p$, causes a spread in focusing strength:

$$\frac{\Delta k}{k} = \mp \frac{\Delta p}{p}. \quad (2.59)$$

Integrated over all focusing (and defocusing) elements in the ring, we obtain a change in the tune of the machine

$$\Delta Q = \frac{1}{4\pi} \int \beta(s) \delta k(s) ds. \quad (2.60)$$

This enables us to calculate Q' :

$$\Delta Q = \frac{1}{4\pi} \int \beta(s) \delta k(s) ds = \left[\frac{-1}{4\pi} \int \beta(s) k(s) ds \right] \frac{\Delta p}{p}. \quad (2.61)$$

The quantity in square brackets is the chromaticity Q' . To be clear, this is called the natural chromaticity. For most alternating gradient machines, its value is about $-1.3Q$. Of course there are two Q values relating to horizontal and vertical oscillations and therefore two chromaticities. Chromaticity may be corrected with sextupole magnets (see Chap. 3, and Sects. 6.1 and 8.1).

2.5 Longitudinal Motion

2.5.1 Stability of the Lagging Particle

Suppose two particles are well below the velocity of light. A particle A , that arrives at the right moment to in the RF resonator and thus will obtain exactly the right acceleration voltage. We call this particle “*synchronous*” (see Fig. 2.22). A second particle, B , arrives late, and so receives an extra energy increment which will cause it to speed up and overtake the synchronous particle, A . In so doing, its energy defect, ΔE , grows and, provided the amplitude is not too large, its trajectory will follow an ellipse in phase space. This describes this motion up and down the r.f. wave (Fig. 2.22) and may remind some readers of the representation of a simple harmonic oscillator, or pendulum. When plotted in a phase space diagram of velocity versus longitudinal displacement we indeed obtain a shape that is elliptical. The trajectory

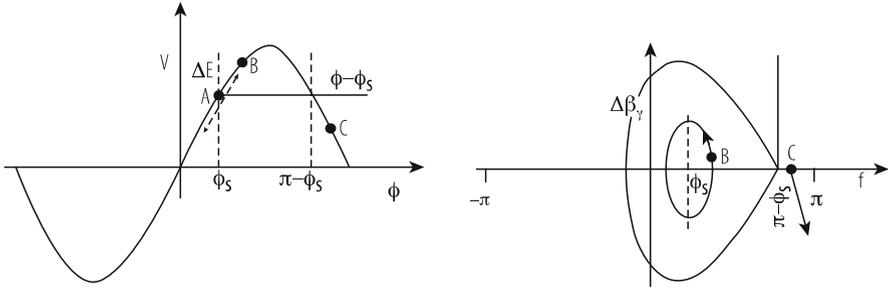


Fig. 2.22 The limiting trajectory for a particle in a ‘moving’ or accelerating bucket when the stable phase is not zero

of this longitudinal movement in phase space is closed and over many turns the average deviation from the synchronous energy is zero. This *phase stability* depends upon the fact that δE is positive when $\phi - \phi_s$ is small and positive [13, 14].

When a particle reaches the non-linear part of the r.f. wave and over the top of the wave, it will still be restored and oscillate about the stable phase provided it does not reach and pass the point where it receives less incremental energy than the synchronous particle. On this non-linear part of the curve the motion is no longer an ellipse but is distorted into a fish-shape but its trajectory is still closed and stable. However, if a particle, C , oscillates with such large amplitude that it falls below the synchronous voltage, an increase in ϕ will cause a negative ΔE which in turn causes ϕ to move further away from the synchronous particle. This particle is clearly unstable and will be continuously decelerated. There is a particle which, starting at $\phi = \pi - \phi_s$, would trace out a limiting fish-shaped trajectory which is the boundary or *separatrix* between stable and unstable motion. The region within this separatrix is called the r.f. bucket and is shown in the lower half of Fig. 2.22. Formulae for the calculation of the parameters of moving buckets are to be found in [15].

Let us look more carefully at the argument that a particle, arriving late because of its lower energy, would see a higher RF voltage from the rising waveform and, accelerated to a higher velocity, would catch up with the synchronous particle. Dispersion may make the situation more complicated. Giving the errant particle more energy will speed it up but may also send it on an orbit of larger radius.

The path length that the particle, B , must travel around the machine, or more correctly, the change in path length with momentum, must depend upon the dispersion function. The closed orbit will have a mean radius:

$$R = R_0 + \bar{D} \frac{\Delta p}{p}. \quad (2.62)$$

Close to the velocity of light where acceleration can increase momentum but not velocity, the longer path length will more than cancel the small effect of velocity and the particle, instead of catching up with its synchronous partner, will arrive

even later than it did on the previous turn. This seems to defeat the whole idea of phase stability. Depending on how the synchrotron is designed and which particles it accelerates, there can be a certain energy where our initial ideas of phase stability break down. This is called the transition energy. Fortunately there is also a way of ensuring stability above transition.

2.5.2 Transition Energy

The rigorous argument to resolve the question of velocity versus path length is to examine how the revolution time (or its reciprocal, the revolution frequency) varies as the particle is given extra acceleration. The revolution frequency is:

$$f = \frac{\beta c}{2\pi R}, (\beta = v/c). \quad (2.63)$$

This revolution frequency, f , depends on two momentum dependent variables, the relativistic $\beta=v/c$ and R , the mean radius. The penultimate equation gives the change in the radius. The momentum dependence of β is determined by:

$$p = \frac{E_0\beta}{\sqrt{1-\beta^2}}. \quad (2.64)$$

The rate of “catching up” depends upon a “slip factor”, η , which is defined as logarithmic differential of frequency as a function of momentum. The procedure of partial derivatives tells us there must be two terms. Hence:

$$\eta_{\text{rf}} = \frac{\Delta f/f}{\Delta p/p} = \frac{p}{\beta} \frac{d\beta}{dp} - \frac{p}{R} \frac{dR}{dp} = \frac{1}{\gamma^2} - \frac{\overline{D}}{R_0}. \quad (2.65)$$

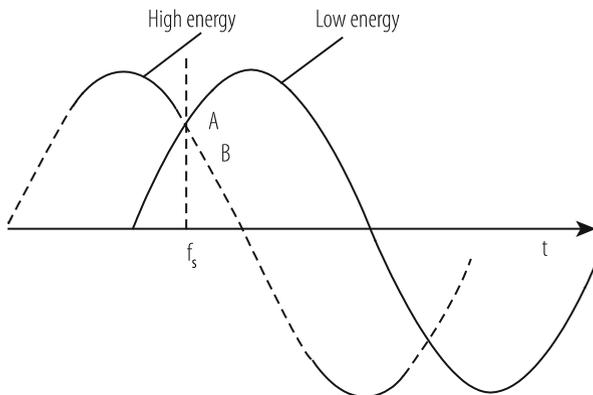
The first term on the right-hand side describes the increase in speed with p and the other (negative), how the path to be completed increases with p .

The second term is energy independent while the first term shrinks as acceleration proceeds. At low energy this is largest and η is >0 . But, since $\gamma = E/E_0$, the first term becomes smaller than the second at high energy so that η changes sign from positive to negative. During the acceleration process there is a certain energy, the transition energy, at which η is momentarily zero. At transition, the value of γ satisfies:

$$\frac{1}{\gamma_{\text{tr}}^2} = \frac{\overline{D}}{R}. \quad (2.66)$$

In proton synchrotron design this condition tends to be encountered mid-way through the acceleration cycle and can only be avoided with some ingenuity in the design of the lattice. This was a worry to the designers of the PS and AGS, the

Fig. 2.23 Shows how changing the phase of the RF voltage waveform can give the lagging particle, B, less energy rather than more and lead to stability above transition



first proton synchrotrons of high enough energy to encounter this problem during acceleration but it was then realised that one could, almost instantaneously, change the phase of the voltage wave in the RF cavities to be falling rather than rising at the moment of the synchronous particles arrival (see Fig. 2.23). With such a reversed slope, particles arriving late are given less than their ration of energy and take a inner circular path—a short cut—to arrive earlier next time.

Electron machines are fortunate in that due to the small rest mass their Lorentz factor γ , being 2000 times higher, ensures that the first term may be neglected and such machines operate always above transition.

2.5.3 Synchrotron Motion

If we consider the motion of a particle on the linear part of the voltage wave of an r.f. cavity it is not difficult to imagine that it approximates rather closely to a harmonic oscillator. Unlike to the situation in the transverse plane, however, the motion becomes more complicated when the oscillation amplitude is larger, and the particle feels the non-linear part of the sinusoidal RF wave, or even more, for part of its motion it finds itself over the crest of the wave. But first let us focus on a small amplitude solution.

It is not hard to deduce from special relativity that the momentum may be written

$$p = m_0c (\beta\gamma) . \tag{2.67}$$

The quantity $\Delta(\beta\gamma)$ serves as the momentum co-ordinate in longitudinal phase space. The other co-ordinate is the particle's arrival phase, ϕ , with respect to the zero crossing of the r.f. voltage at the cavity. Let us consider the simplest case of a small oscillation in a stationary bucket, $\phi_s = 0$ (when the particle is not being accelerated).

A particle with a small phase error will describe an ellipse in phase space which one may write parametrically as

$$\begin{aligned}\Delta(\beta\gamma) &= \Delta(\widehat{\beta\gamma}) \sin 2\pi f_s t, \\ \phi &= \hat{\phi} \cos 2\pi f_s t,\end{aligned}\tag{2.68}$$

where f_s is the frequency of execution of these oscillations in phase which we call the synchrotron frequency.

To reveal the differential equation behind this motion we must first remember that the angular frequency $2\pi f$ of an oscillator is nothing other than the rate of change of phase, $\dot{\phi}$ or to be exact $-\dot{\phi}$. (The negative sign stems from the fact that ϕ is a phase lag.) We may therefore relate the rate of change in arrival phase to the difference in revolution frequency of the particle, compared to that of the synchronous particle.

$$\dot{\phi} = -2\pi h [f(\Delta\beta\gamma) - f(0)] = -2\pi h \Delta f.\tag{2.69}$$

We have multiplied by, h , the harmonic number of the r.f. since ϕ is the phase angle of the r.f. swing while $f(\Delta\beta\gamma)$ is the revolution frequency. Here we can use the definition of the slip factor η and then simply use some standard relativistic relations to end up with Δf as a function of ΔE , the energy defect with respect to the synchronous particle:

$$\Delta f = \eta f \frac{\Delta p}{p} = \eta f \frac{\Delta(\beta\gamma)}{(\beta\gamma)} = \frac{\eta f}{\beta^2} \frac{\Delta\gamma}{\gamma} = \frac{\eta f}{E_0 \beta^2 \gamma} \Delta E.\tag{2.70}$$

where E_0 the total energy (including its rest mass) of the synchronous particle.

We differentiate once more to obtain a second order differential equation which we hope to resemble a simple oscillator.

$$\ddot{\phi} = -\frac{2\pi h \eta f}{E_0 \beta^2 \gamma} (\Delta \dot{E}).\tag{2.71}$$

The extra energy given per turn to a particle whose arrival phase is ϕ will be

$$\Delta E = e V_0 (\sin \phi - \sin \phi_s),\tag{2.72}$$

and the rate of change of energy will be this times, f , the revolution frequency. So we can write

$$\ddot{\phi} = -\frac{2\pi e V_0 h \eta f^2}{E_0 \beta^2 \gamma} (\sin \phi - \sin \phi_s).\tag{2.73}$$

This is a fundamental and exact description of the motion provided the parameters should change slowly (the adiabatic assumption). We can simply integrate to

find its solution numerically but to see an analytic solution for small amplitudes we set $\phi_s = 0$ and $\phi \approx \sin\phi$:

$$\ddot{\phi} + \frac{2\pi e V_0 h \eta f^2}{E_0 \beta^2 \gamma} \phi = 0. \quad (2.74)$$

The frequency of these synchrotron oscillations in longitudinal phase space is

$$f_s = \sqrt{\frac{|\eta| h e V_0}{2\pi E_0 \beta^2 \gamma}} f, \quad (2.75)$$

or writing $f_{\text{rf}} = hf$ we could equally express

$$f_s = \sqrt{\frac{|\eta| e V_0}{2\pi E_0 \beta^2 \gamma h}} f_{\text{rf}}. \quad (2.76)$$

In analogy to the transverse plane, we define a synchrotron tune, Q_s , as the number of such oscillations per revolution of the machine. This is analogous to Q in transverse phase space.

$$Q_s = \frac{f_s}{f} = \sqrt{\frac{|\eta| e h V_0 \cos\phi_s}{2\pi E_0 \beta^2 \gamma}}. \quad (2.77)$$

Usually Q_s is less than 10% of the revolution frequency. It drops down to zero at γ transition where η is zero and then rises again. In large proton machines it can be in the region 0 to 100 Hz and, but for the vacuum, one might hear it!

Close to γ_{tr} we cannot strictly assume that β , γ , η , and f vary slowly in comparison with the synchrotron oscillation which this equation describes. Hence we should use a more exact form of the equation of motion and approximate only when it seems that this is justified:

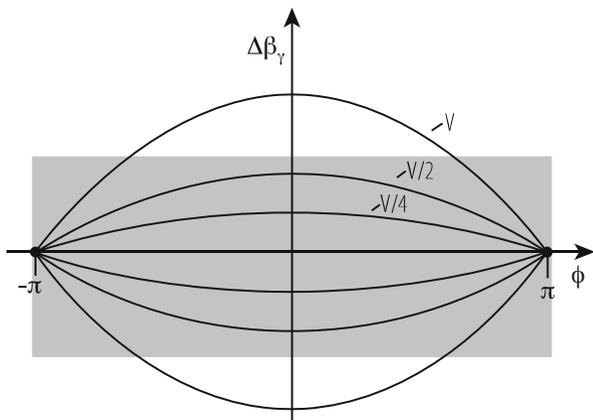
$$\frac{d}{dt} \left[\frac{E_0 \beta^2 \gamma \dot{\phi}}{2\pi \eta h f^2} \right] + e V_0 (\sin\phi - \sin\phi_s) = 0. \quad (2.78)$$

In a stationary bucket, when $\phi_s = 0$, this exact differential equation for large amplitude motion is identical to that for a rigid pendulum:

$$\ell \frac{d^2\theta}{dt^2} + g \sin\theta = 0. \quad (2.79)$$

There is an extra term, $\sin\phi_s$, on the right hand side of the synchrotron equation which is not there in the pendulum case but it could be introduced too for the pendulum by using a magnetic ‘bob’ and biasing its equilibrium position to one side by attaching a weight on a cantilever at right angles to the rod of the pendulum.

Fig. 2.24 Adiabatic trapping of coasting beam in growing stationary bucket



In fact the unbiased pendulum corresponds to synchrotron motion when there is no acceleration—we say the bucket is stationary. In Fig. 2.22 we saw how particles close to the edge of the stable area of the bucket follow a fish-shaped trajectory when $\phi_s = 0$; before acceleration starts or when the beam is held at the same energy in collider mode (see Fig. 2.24).

In order to accelerate ϕ_s must be made finite, in which case the figure changes somewhat. The stable area becomes smaller and shaped like a fish—or rather a series of fish chasing each other’s tails. Small amplitude motion will still be sinusoidal but the ellipse will be centred on the stable phase ϕ_s and not on $\phi = 0$.

2.5.4 Stationary Buckets

The size of the bucket depends on how close the stable phase, ϕ_s is to the crest of the sine-wave. It shrinks to zero if $\phi_s = 90^\circ$. There is a special case if ϕ_s is zero. This is often the case as a beam injected into a synchrotron before acceleration has started or in a collider where the r.f. simply holds the bunches together. The bucket is then said to be ‘stationary’ stretching over all phases from $-\pi$ to π . Its height is the range of energies $2\Delta E$ which the r.f. wave can constrain and this turns out to be dependent on \sqrt{V} for a given ϕ_s . If V is reduced, the more energetic particles spill out of the bucket.

Very often the particles are injected as a continuous ribbon without any longitudinal structure crosshatched in Fig. 2.24. Usually acceleration has not yet started, the magnetic field B is constant, and ϕ_s is zero. If V is increased slowly, the height of the stationary bucket grows, and more and more of the energy spread in the beam, ΔE , is trapped (Fig. 2.24). This is called “adiabatic trapping”.

References

1. E.D. Courant, H.S. Snyder: *Annals of Physics* 3 (1958) 1–48.
2. J.J. Livingood: *Principles of cyclic particle accelerators*, van Nostrand (1961).
3. R. Servranckx, K.L. Brown: SLAC Report 270 UC-288 (1984).
4. A.A. Garren, A.S. Kenney, E.D. Courant, M.J. Syphers: Fermilab report FN40 (1985).
5. L. Schachinger, R. Talman: SSC Report-52 (1985).
6. F.C. Iselin, H.G. Grote: CERN/SL/90-13 (1991).
7. P. Schmüser: CERN 87-10 (1987).
8. J. Rossbach, P. Schmüser: CERN 94-1 (1992).
9. P. Bryant: CERN ISR-MA/75-28 (1975).
10. J.P. Koutchouk: CERN ISR-OP/80-27 (1980).
11. G. Guignard: CERN 76-06 (1976).
12. S. Guiducci: CERN 94-01 (1992).
13. B.W. Montague: CERN 77-13 (1977).
14. J. Le Duff: CERN 94-01 (1992).
15. C. Bovet, R. Gouiran, I. Gumowski, H. Reich: CERN/MPS-SI/Int.DL/70/4 (1970).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 3

Non-linear Dynamics in Accelerators



Werner Herr and Etienne Forest

3.1 Introduction

Non-linear effects in accelerator physics are important both during the design stage and for successful operation of accelerators. Since both of these aspects are closely related, they will be treated together in this overview. Some of the most important aspects are well described by methods established in other areas of physics and mathematics. Given the scope of this handbook, the treatment will be focused on the problems in accelerators used for particle physics experiments. Although the main emphasis will be on accelerator physics issues, some of the aspects of more general interest will be discussed. In particular to demonstrate that in recent years a framework has been built to handle the complex problems in a consistent form, technically superior and conceptually simpler than the traditional techniques. The need to understand the stability of particle beams has substantially contributed to the development of new techniques and is an important source of examples which can be verified experimentally. Unfortunately the documentation of these developments is often poor or even unpublished, in many cases only available as lectures or conference proceedings.

This article is neither rigorous nor a complete treatment of the topic, but rather an introduction to a limited set of contemporary tools and methods we consider useful in accelerator theory.

W. Herr (✉)

CERN (European Organization for Nuclear Research), Meyrin, Genève, Switzerland
e-mail: werner.herr@cern.ch

E. Forest

KEK, High Energy Research Organisation, Tsukuba, Japan

© The Author(s) 2020

S. Myers, H. Schopper (eds.), *Particle Physics Reference Library*,
https://doi.org/10.1007/978-3-030-34245-6_3

3.1.1 *Motivation*

The most reliable tools to study (i.e. description of the machine) are simulations (e.g. tracking codes).

- Particle Tracking is a numerical solution of the (nonlinear) Initial Value Problem. It is a “integrator” of the equation of motion and a vast amount of tracking codes are available, together with analysis tools (Examples: Lyapunov, Chirikov, chaos detection, frequency analysis, ...)
- It is unfortunate that theoretical and computational tools exist side by side without an undertaking how they can be integrated.
- It should be undertaken to find an approach to link simulations with theoretical analysis, would allow a better understanding of the physics in realistic machines.
- A particularly promising approach is based on finite maps [1].

3.1.2 *Single Particle Dynamics*

The concepts developed here are used to describe single particle transverse dynamics in rings, i.e. circular accelerators or storage rings. This is not a restriction for the application of the presented tools and methods. In the case of linear betatron motion the theory is rather complete and the standard treatment [2] suffices to describe the dynamics. In parallel with this theory the well known concepts such as closed orbit and Twiss parameters are introduced and emerged automatically from the Courant-Snyder formalism [2]. The formalism and applications are found in many textbooks (e.g. [3–5]).

In many new accelerators or storage rings (e.g. LHC) the description of the machine with a linear formalism becomes insufficient and the linear theory must be extended to treat non-linear effects. The stability and confinement of the particles is not given a priori and should rather emerge from the analysis. Non-linear effects are a main source of performance limitations in such machines. A reliable treatment is required and the progress in recent years allows to evaluate the consequences. Very useful overview and details can be found in [6–8].

3.1.3 *Layout of the Treatment*

Following a summary of the sources of non-linearities in circular machine, the basic methods to evaluate the consequences of non-linear behaviour are discussed. Since the traditional approach has caused misconception and the simplifications led to wrong conclusions, more recent and contemporary tools are introduced to treat these problems. An attempt is made to provide the physical picture behind these tools rather than a rigorous mathematical description and we shall show how the

new concepts are a natural extension of the Courant-Snyder formalism to non-linear dynamics. An extensive treatment of these tools and many examples can be found in [7]. In the last part we summarize the most important physical phenomena caused by the non-linearities in an accelerator.

3.2 Variables

For what follows one should always use canonical variables!

In Cartesian coordinates:

$$R = (X, P_X, Y, P_Y, Z, P_Z, t) \quad (3.1)$$

If the energy is constant (i.e. $P_Z = \text{const.}$), we use:

$$(X, P_X, Y, P_Y, Z, t) \quad (3.2)$$

This system is rather inconvenient, what we want is the description of the particle in the neighbourhood of the reference orbit/trajectory:

$$R_d = (X, P_X, Y, P_Y, Z, t) \quad (3.3)$$

which are considered now the deviations from the reference and which are zero for a particle on the reference trajectory

It is very important that it is the reference not the design trajectory!
(so far it is a straight line along the Z -direction)

3.2.1 Trace Space and Phase Space

A confusion often arises about the terms Phase Space (x, p_x, \dots) or Trace Space (x, x', \dots)

It is not laziness nor stupidity to use one or the other:

- Beam dynamics is strictly correct only with (x, p_x, \dots) , (see later chapter) but in general quantities cannot be measured easily
- Beam dynamics with (x, x', \dots) needs special precaution, but quantities based on these coordinates are much easier to measure
- Some quantities are different (e.g. emittance)

It comes back to a remark made at the beginning, i.e. that we shall use rings for our arguments. In single pass machine, e.g. linac, beam lines, spectrometers, the beam is not circulating over many turns and several hours, therefore there is no interest in stability issues. Instead for most of these applications what counts is the

coordinates and angles at a given position (x, x', y, y') , e.g. at the end of a beam line or a small spot one an electron microscope. When “accelerator physicists” talk about concepts such as tune, resonances, β -functions, equilibrium emittances etc., all these are irrelevant for single pass machine. There is no need to study iterating systems. In these cases the use of the trace space is fully adequate, in fact preferred because the quantities can be measured. In the end, the mathematical tools are very different from the ones discussed in this article.

3.2.2 Curved Coordinate System

For a “curved” trajectory, in general not circular, with a local radius of curvature $\rho(s)$ in the horizontal (X - Z plane), we have to transform to a new coordinate system (x, y, s) (co-moving frame) with:

$$\begin{aligned} X &= (x + \rho) \cos\left(\frac{s}{\rho}\right) - \rho \\ Y &= y \\ Z &= (x + \rho) \sin\left(\frac{s}{\rho}\right) \end{aligned} \quad (3.4)$$

The new canonical momenta become:

$$\begin{aligned} p_x &= P_X \cos\left(\frac{s}{\rho}\right) + P_Z \sin\left(\frac{s}{\rho}\right) \\ p_y &= P_Y \\ p_s &= P_Z \left(1 + \frac{x}{\rho}\right) \cos\left(\frac{s}{\rho}\right) - P_X \left(1 + \frac{x}{\rho}\right) \sin\left(\frac{s}{\rho}\right) \end{aligned} \quad (3.5)$$

3.3 Sources of Non-linearities

Any object creating non-linear electromagnetic fields on the trajectory of the beam can strongly influence the beam dynamics. They can be generated by the environment or by the beam itself.

3.3.1 Non-linear Machine Elements

Non-linear elements can be introduced into the machine on purpose or can be the result of field imperfections. Both types can have adverse effects on the beam stability and must be taken into account.

3.3.1.1 Unwanted Non-linear Machine Elements

The largest fraction of machine elements are either dipole or quadrupole magnets. In the ideal case, these types of magnets have pure dipolar or quadrupolar fields and behave approximately as linear machine elements. Any systematic or random deviation from this linear field introduces non-linear fields into the machine lattice. These effects can dominate the aperture required and limit the stable region of the beam. The definition of tolerances on these imperfections is an important part of any accelerator design.

Normally magnets are long enough that a 2-dimensional field representation is sufficient. The components of the magnetic field can be derived from the potential and in cylindrical coordinates ($r, \Theta, s = 0$) can be written as:

$$B_r(r, \Theta) = \sum_{n=1}^{\infty} (B_n \sin(n\Theta) + A_n \cos(n\Theta)) \left(\frac{r}{R_{ref}} \right)^{n-1}, \quad (3.6)$$

$$B_{\Theta}(r, \Theta) = \sum_{n=1}^{\infty} (B_n \cos(n\Theta) - A_n \sin(n\Theta)) \left(\frac{r}{R_{ref}} \right)^{n-1}, \quad (3.7)$$

where R_{ref} is a reference radius and B_n and A_n are constants. Written in Cartesian coordinates we have:

$$B(z) = \sum_{n=1}^{\infty} (B_n + i A_n) \left(\frac{r}{R_{ref}} \right)^{n-1} \quad (3.8)$$

where $z = x + iy = r e^{i\Theta}$. The terms n correspond to $2n$ -pole magnets and the B_n and A_n are the normal and skew multipole coefficients. The beam dynamics set limits on the allowed multipole components of the installed magnets.

3.3.1.2 Wanted Non-linear Machine Elements

In most accelerators the momentum dependent focusing of the lattice (chromaticity) needs to be corrected with sextupoles [3, 4]. Sextupoles introduce non-linear fields into the lattice that are larger than the intrinsic non-linearities of the so-called linear elements (dipoles and quadrupoles). In a strictly periodic machine the correction can be done close to the origin and the required sextupole strengths can be kept small. For colliding beam accelerators usually special insertions are foreseen to host the experiments where the dispersion is kept small and the β -function is reduced to a minimum. The required sextupole correction is strong and can lead to a reduction of the dynamic aperture, i.e. the region of stability of the beam. In most accelerators the sextupoles are the dominant source of non-linearity. To minimize this effect is an important issue in any design of an accelerator.

Another source of non-linearities can be octupoles used to generate amplitude dependent detuning to provide Landau damping in case of instabilities.

3.3.2 *Beam–Beam Effects and Space Charge*

A strong source of non-linearities are the fields generated by the beam itself. They can cause significant perturbations on the same beam (space charge effects) or on the opposing beam (beam-beam effects) in the case of a colliding beam facility.

As an example, for the simplest case of round beams with the line density n and the beam size σ the field components can be written as:

$$E_r = -\frac{ne}{4\pi\epsilon_0} \cdot \frac{\partial}{\partial r} \int_0^\infty \frac{\exp(-\frac{r^2}{(2\sigma^2+q)})}{(2\sigma^2+q)} dq, \quad (3.9)$$

and

$$B_\Phi = -\frac{ne\beta c\mu_0}{4\pi} \cdot \frac{\partial}{\partial r} \int_0^\infty \frac{\exp(-\frac{r^2}{(2\sigma^2+q)})}{(2\sigma^2+q)} dq. \quad (3.10)$$

In colliding beams with high density and small beams these fields are the dominating source of non-linearities. The full treatment of beam-beam effects is complicated due to mutual interactions between the two beams and a self-consistent treatment is required. in the presence of all other magnets in the ring.

3.4 Map Based Techniques

In the standard approach to single particle dynamics in rings, the equations of motion are introduced together with an ansatz to solve these equations. In the case of linear motion, this ansatz is due to Courant-Snyder [2]. However, this treatment must assume that the motion of a particle in the ring is stable and confined. For a non-linear system this is a priori not known and the attempt to find a complete description of the particle motion must fail.

The starting point for the treatment of the linear dynamics in synchrotrons is based on solving a linear differential equation of the Hill type.

$$\frac{d^2x(s)}{ds^2} + \underbrace{\left(a_0 + 2 \sum_{n=1}^{\infty} a_n \cdot \cos(2ns) \right)}_{K(s)} x(s) = 0.$$

Each element at position s acts as a source of forces, i.e. we must write for the forces $K \rightarrow K(s)$ which is assumed to be a periodic function, i.e. $K(s + C) = K(s)_{ring}$

The solution of this Boundary Value Problem must be periodic too!

It is therefore not applicable in the general case (e.g. Linacs, Beamlines, FFAG, Recirculators, ...), much better to treat it as an Initial Value Problem.

In a more useful approach we do not attempt to solve such an overall equation but rather consider the fundamental objects of an accelerators, i.e. the machine elements themselves. These elements, e.g. magnets or other beam elements, are the basic building blocks of the machine. All elements have a well defined action on a particle which can be described independent of other elements or concepts such as closed orbit or β -functions. Mathematically, they provide a “map” from one face of a building block to the other, i.e. a description of how the particles move inside and between elements. In this context, a map can be anything from linear matrices to high order integration routines.

A map based technique is also the basis for the treatment of particle dynamics as an Initial value Problem (IVP).

It follow immediately that for a linear, 1st order equation of the type

$$\frac{dx(s)}{ds} = K(s) x(s) \quad (\text{and initial values at } s_0)$$

the solution can always be written as:

$$\begin{aligned} x(s) &= a \cdot x(s_0) + b \cdot x'(s_0) \\ x'(s) &= c \cdot x(s_0) + d \cdot x'(s_0) \end{aligned} \implies \begin{pmatrix} x \\ x' \end{pmatrix}_s = \overbrace{\begin{pmatrix} a & b \\ c & d \end{pmatrix}}^A \begin{pmatrix} x \\ x' \end{pmatrix}_{s_0}$$

where the function $K(s)$ does not have to be periodic. Furthermore, the determinant of the matrix A is always 1. Therefore it is an advantage to use maps (matrices) for a linear systems from the start, without trying to solve a differential equation.

The collection of all machine elements make up the ring pr beam line and it is the combination of the associated maps which is necessary for the description and analysis of the physical phenomena in the accelerator ring or beam line.

For a circular machine the most interesting map is the one which describes the motion once around the machine, the so-called One-Turn-Map. It contains all necessary information on stability, existence of closed orbit, and optical parameters. The reader is assumed to be familiar with this concept in the case of linear beam dynamics (Chap. 2) where all maps are matrices and the Courant-Snyder analysis of the corresponding one-turn-map produces the desired information such as e.g. closed orbit or Twiss parameters.

It should therefore be the goal to generalize this concept to non-linear dynamics. The computation of a reliable one-turn-map and the analysis of its properties will provide all relevant information.

Given that the non-linear maps can be rather complex objects, the analysis of the one-turn-map should be separated from the calculation of the map itself.

3.5 Linear Normal Forms

3.5.1 Sequence of Maps

Starting from a position s_0 and combining all matrices to get the matrix to position $s_0 + L$ (shown for 1D only):

$$\begin{pmatrix} x \\ x' \end{pmatrix}_{s_0 + L} = \underbrace{M_N \circ M_{N-1} \circ \dots \circ M_1}_{M(s_0, L)} \circ \begin{pmatrix} x \\ x' \end{pmatrix}_{s_0} \quad (3.11)$$

For a ring with circumference C one obtains the One-Turn-Matrix (OTM) at s_0

$$\begin{pmatrix} x \\ x' \end{pmatrix}_{s_0 + C} = \underbrace{\begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}}_{M_{OTM}} \circ \begin{pmatrix} x \\ x' \end{pmatrix}_{s_0} \quad (3.12)$$

Without proof, the scalar product:

$$\begin{pmatrix} x \\ x' \end{pmatrix}_{s_0} \cdot M_{OTM} \begin{pmatrix} x \\ x' \end{pmatrix}_{s_0} = \text{const.} = J \quad (3.13)$$

is a constant of the motion: invariant of the One Turn Map.

With this approach we have a strong argument that the construction of the One Turn Map is based on the properties of each element in the machine. It is entirely independent of the purpose of the machine and their global properties. It is not restricted to rings or in general to circular machine.

Once the One Turn Map is constructed, it can be analysed, but this analysis does not depend on how it was constructed.

As a paradigm: the construction of a map (being for a circular machine or not) and its analysis are conceptual and computational separated undertakings.

3.5.2 Analysis of the One Turn Map

The key for the analysis is that matrices can be transformed into **Normal Forms**. Starting with the One-Turn-Matrix, and try to find a (invertible) transformation A such that:

$$AMA^{-1} = R \quad (\text{or :} \quad A^{-1}RA = M)$$

- The matrix R is:
 - A “Normal Form”, (or at least a very simplified form of the matrix)
 - For example (most important case): R becomes a pure rotation
- The matrix R describes the same dynamics as M , but:
 - All coordinates are transformed by A
 - This transformation A “analyses” the complexity of the motion, it contains the structure of the phase space

$$M = A \circ R \circ A^{-1} \quad \text{or :} \quad R = A^{-1} \circ M \circ A$$

The motion on an ellipse becomes a motion on a circle (i.e. a rotation): R is the simple part of the map and its shape is dumped into the matrix A . R can be obtained by the evaluation of the Eigenvectors and Eigenvalues.

One finds for the two components of the original map:

$$A = \begin{pmatrix} \sqrt{\beta(s)} & 0 \\ -\frac{\alpha(s)}{\sqrt{\beta(s)}} & \frac{1}{\sqrt{\beta(s)}} \end{pmatrix} \quad \text{and} \quad R = \begin{pmatrix} \cos(\Delta\mu) & \sin(\Delta\mu) \\ -\sin(\Delta\mu) & \cos(\Delta\mu) \end{pmatrix} \quad (3.14)$$

Please note that the normal form analysis gives the eigenvectors (3.14) without any physical picture related to their interpretation. The formulation using α and β is due to Courant and Snyder. Amongst other advantages it can be used to “normalise” the position x : the normalised position x_n is the “non-normalized” divided by $\sqrt{\beta}$. The variation of the normalised position x_n is then smaller than in the non-normalized case. This is also better suited for analytical calculation, e.g. involving perturbation theory.

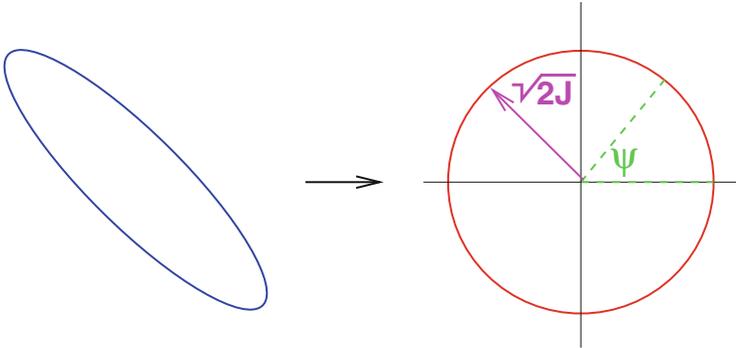
The Normal Form transformation together with this choice gives the required information:

- μ_x is the “tune” $Q_x \cdot 2\pi$ (now we can talk about phase advance!)
- β, α, \dots are the optical parameters and describe the ellipse
- The closed orbit (an invariant, identical coordinates after one turn!):
- $M_{OTM} \circ (x, x')_{co} \equiv (x, x')_{co}$

3.5.3 Action-Angle Variables

More appropriate for studies of beam dynamics is the use of Action-Angle variables.

Once the particles “travel” on a circle, the motion is better described by the canonical variables action J_x and angle Ψ_x :



with the definitions and the choice is (3.14):

$$\begin{aligned}
 x &= \sqrt{2J_x\beta_x} \cos(\Psi_x) \\
 p_x &= -\sqrt{\frac{2J_x}{\beta_x}} (\sin(\Psi_x) + \alpha_x \cos(\Psi_x)) \\
 J_x &= \frac{1}{2}(\gamma_x x^2 + 2\alpha_x x p_x + \beta_x p_x^2)
 \end{aligned} \tag{3.15}$$

- the angular position along the ring Ψ becomes the independent variable!
- The trajectory of a particle is now independent of the position s !
- The constant radius of the circle $\sqrt{2J}$ defines the action J (invariant of motion)

3.5.4 Beam Emittance

A sad and dismal story in accelerator physics is the definition of the emittance. Most foolish in this context is to relate emittance to single particles. This is true in particular when we have a beam line which is not periodic. In that case the Courant-Snyder parameters can be determined from the beam. These parameters are related to the moments of the beam, e.g. the beam size is directly related to the second order moment $\langle x^2 \rangle$. Using the expression above for the action and angle, we can write for this expression:

$$\langle x^2 \rangle = \langle 2J_x\beta_x \cdot \cos^2(\Psi_x) \rangle = 2\beta_x \langle J_x \cdot \cos^2(\Psi_x) \rangle. \tag{3.16}$$

The average of \cos^2 can immediately be evaluated as 0.5 and defining the emittance as:

$$\epsilon_x = \langle J_x \rangle, \quad (3.17)$$

we write

$$\langle x^2 \rangle = \beta_x \cdot \epsilon_x. \quad (3.18)$$

Using a similar procedure (details and derivation in e.g. [3], and to a much lesser extent in [1]) one can determine the moments

$$\langle p_x^2 \rangle = \gamma_x \cdot \epsilon_x, \quad (3.19)$$

and

$$\langle x \cdot p_x \rangle = -\alpha_x \cdot \epsilon_x. \quad (3.20)$$

Using these expressions, the emittance becomes readily

$$\epsilon_x = \sqrt{\langle x^2 \rangle \langle p_x^2 \rangle - \langle x \cdot p_x \rangle^2} \quad (3.21)$$

Therefore, once the emittance is measured, the Courant-Snyder parameters are determined by Eqs. (3.18), (3.19), and (3.20).

Since other definitions often refer to the treatment by Courant and Snyder, here a quote from Courant himself in [9]:

Interlude 1

The invariant J is simply related to the area enclosed by the ellipse:

$$\text{Area enclosed} = 2\pi J. \quad (3.22)$$

In accelerator and storage ring terminology there is a quantity called the *emittance* which is closely related to this invariant. The emittance, however, is a property of a distribution of particles, not a single particle. Consider a Gaussian distribution in amplitudes. Then the (rms) emittance, ϵ , is given by:

$$(x_{rms})^2 = \beta_x(s) \cdot \epsilon_x. \quad (3.23)$$

In terms of the action variable, J , this can be rewritten

$$\epsilon_x = \langle J \rangle. \quad (3.24)$$

where the bracket indicates an average over the distribution in J .

Other definitions based on handwaving arguments or those approximately valid only in special cases, should be discarded, in particular those relying on presumed distributions, e.g. Gaussian.

3.6 Techniques and Tools to Evaluate and Correct Non-linear Effects

The key to a more modern approach shown in this section is to avoid the prejudices about the stability and other properties of the ring. Instead, we must describe the machine in terms of the objects it consists of with all their properties, including the non-linear elements. The analysis will reveal the properties of the particles such as e.g. stability. In the simplest case, the ring is made of individual machine elements such as magnets which have an existence on their own, i.e. the interaction of a particle with a given element is independent of the motion in the rest of the machine. Also for the study of non-linear effects, the description of elements should be independent of concepts such as tune, chromaticity and closed orbit. To successfully study single particle dynamics, one must be able to describe the action of the machine element on the particle as well as the machine element.

3.6.1 Particle Tracking

The ring being a collection of maps, a particle tracking code, i.e. an integrator of the equation of motion, is the most reliable map for the analysis of the machine. Of course, this requires an appropriate description of the non-linear maps in the code. It is not the purpose of this article to describe the details of tracking codes and the underlying philosophy, such details can be found in the literature (see e.g. [6]). Here we review and demonstrate the basic principles and analysis techniques.

3.6.1.1 Symplecticity

If we define a map through $\vec{z}_2 = M_{12}(\vec{z}_1)$ as a propagator from a location “1” to a location “2” in the ring, we have to consider that not all possible maps are allowed. The required property of the map is called “symplecticity” and in the simplest case where M_{12} is a matrix, the symplecticity condition can be written as:

$$M \Rightarrow M^T \cdot S \cdot M = S \quad \text{where} \quad S = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \quad (3.25)$$

The physical meaning of this condition is that the map is area preserving in the phase space. The condition can easily be derived from a Hamiltonian treatment, closely related to Liouville's theorem.

3.6.2 *Approximations and Tools*

The concept of symplecticity is vital for the treatment of Hamiltonian systems. This is true in particular when the stability of a system is investigated using particle tracking. However, in practice it is difficult to accomplish for a given exact problem. As an example we may have the exact fields and potentials of electromagnetic elements. For a single pass system a (slightly) non-symplectic integrator may be sufficient, but for an iterative system the results are meaningless.

To track particles using the exact model may result in a non-symplectic tracking, i.e. the underlying model is correct, but the resulting physics is wrong.

It is much better to approximate the model to the extent that the tracking is symplectic. One might compromise on the exactness of the final result, but the correct physics is ensured.

As a typical example one might observe possible chaotic motion during the tracking procedure. However, there is always a non-negligible probability that this interpretation of the results may be wrong. To conclude that it is not a consequence of non-symplecticity of the procedure or a numerical artifact it is necessary to identify the physical mechanism leading to this observation.

This may not be possible to achieve using the exact model as input to a (possibly) non-symplectic procedure. Involving approximations to the definition of the problem should reveal the correct physics at the expense of a (hopefully) small error. Staying exact, the physics may be wrong.

As a result, care must be taken to positively identify the underlying process.

This procedure should be based on approximations as close as possible to the exact problem, but allowing a symplectic evaluation.

An example for this will be shown in Sect. 3.6.3.4.

3.6.3 *Taylor and Power Maps*

A non-linear element cannot be represented in the form of a linear matrix and more complicated maps have to be introduced [5]. In principle, any well behaved, non-linear function can be developed as a Taylor series. This expansion can be truncated at the desired precision.

Another option is the representation as Lie transformations [8, 10]. Both types are discussed in this section.

3.6.3.1 Taylor Maps

A Taylor map can be written using higher order matrices and in the case of two dimensions we have:

$$z_j(s_2) = \sum_{k=1}^4 R_{jk} z_k(s_1) + \sum_{k=1}^4 \sum_{l=1}^4 T_{jkl} z_k(s_1) z_l(s_1) \quad (3.26)$$

(where z_j , $j = 1, \dots, 4$, stand for x, x', y, y'). Let us call the collection: $A_2 = (R, T)$ the second order map A_2 . Higher orders can be defined as needed, e.g. for the 3rd order map $A_3 = (R, T, U)$ we add a third order matrix:

$$+ \sum_{k=1}^4 \sum_{l=1}^4 \sum_{m=1}^4 U_{jklm} z_k(s_1) z_l(s_1) z_m(s_1) \quad (3.27)$$

Since Taylor expansions are not matrices, to provide a symplectic map, it is the associated Jacobian matrix J which must fulfill the symplecticity condition:

$$J_{ik} = \frac{\partial z_i(s_2)}{\partial z_k(s_1)} \quad \text{and} \quad J \text{ must fulfill : } J^t \cdot S \cdot J = S \quad (3.28)$$

However, in general $J_{ik} \neq \text{const}$ and for a truncated Taylor map it can be difficult to fulfill this condition for all z . As a consequence, the number of independent coefficients in the Taylor expansion is reduced and the complete, symplectic Taylor map requires more coefficients than necessary [7].

The explicit maps for a sextupole is:

$$\begin{aligned} x_2 &= x_1 + Lx'_1 - k_2 \left(\frac{L^2}{4}(x_1^2 - y_1^2) + \frac{L^3}{12}(x_1x'_1 - y_1y'_1) + \frac{L^4}{24}(x_1'^2 - y_1'^2) \right) \\ x'_2 &= x'_1 - k_2 \left(\frac{L}{2}(x_1^2 - y_1^2) + \frac{L^2}{4}(x_1x'_1 - y_1y'_1) + \frac{L^3}{6}(x_1'^2 - y_1'^2) \right) \\ y_2 &= y_1 + Ly'_1 + k_2 \left(\frac{L^2}{4}x_1y_1 + \frac{L^3}{12}(x_1y'_1 + y_1x'_1) + \frac{L^4}{24}(x'_1y'_1) \right) \\ y'_2 &= y'_1 + k_2 \left(\frac{L}{2}x_1y_1 + \frac{L^2}{4}(x_1y'_1 + y_1x'_1) + \frac{L^3}{6}(x'_1y'_1) \right) \end{aligned} \quad (3.29)$$

Writing the explicit form of the Jacobian matrix:

$$J_{ik} = \begin{pmatrix} \frac{\partial x_2}{\partial x_1} & \frac{\partial x_2}{\partial x'_1} & \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y'_1} \\ \frac{\partial x'_2}{\partial x_1} & \frac{\partial x'_2}{\partial x'_1} & \frac{\partial x'_2}{\partial y_1} & \frac{\partial x'_2}{\partial y'_1} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x'_1} & \frac{\partial y_2}{\partial y_1} & \frac{\partial y_2}{\partial y'_1} \\ \frac{\partial y'_2}{\partial x_1} & \frac{\partial y'_2}{\partial x'_1} & \frac{\partial y'_2}{\partial y_1} & \frac{\partial y'_2}{\partial y'_1} \end{pmatrix} \rightarrow k_2 = 0 \quad \begin{pmatrix} 1 & L & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & L \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (3.30)$$

For $k_2 \neq 0$ coefficients depend on initial values, e.g.:

$$\frac{\partial y_2}{\partial y_1} = 1 + k_2 \left(\frac{L^2}{4} x_1 + \frac{L^3}{12 x_1} \right) \rightarrow \text{Power series are not symplectic, cannot be used}$$

The non-symplecticity can be recovered in the case of elements with $L = 0$. It becomes small (probably small enough) when the length is small.

As a result, the model is approximated by a small amount, but the symplecticity (and therefore the physics) is ensured. An exact model but compromised integration can fabricate non-existing features and conceal important underlying physics.

The situation is rather different in the case of single pass machines. The long term stability (and therefore symplecticity) is not an issue and the Taylor expansion around the closed orbit is what is really needed. Techniques like the one described in Sect. 3.7.6 provide exactly this in an advanced and flexible formalism.

3.6.3.2 Thick and Thin Lenses

All elements in a ring have a finite length and therefore should be treated as “thick lenses”. However, in general a solution for the motion in a thick element does not exist. It has become a standard technique to avoid using approximate formulae to track through thick lenses and rather perform exact tracking through thin lenses. This approximation is improved by breaking the thick element into several thin elements which is equivalent to a numerical integration. A major advantage of this technique is that “thin lens tracking” is automatically symplectic. In this context it becomes important to understand the implied approximations and how they influence the desired results. We proceed by an analysis of these approximations and show how “symplectic integration” techniques can be applied to this problem.

We demonstrate the approximation using a quadrupole. Although an exact solution of the motion through a quadrupole exists, it is a useful demonstration since it can be shown that all concepts developed here apply also to arbitrary non-linear elements.

Let us assume the transfer map (matrix) for a thick, linearized quadrupole of length L and strength K :

$$M_{s \rightarrow s+L} = \begin{pmatrix} \cos(L \cdot \sqrt{K}) & \frac{1}{\sqrt{K}} \cdot \sin(L \cdot \sqrt{K}) \\ -\sqrt{K} \cdot \sin(L \cdot \sqrt{K}) & \cos(L \cdot \sqrt{K}) \end{pmatrix} \quad (3.31)$$

This map is exact and can be expanded as a Taylor series for a “small” length L :

$$M_{s \rightarrow s+L} = L^0 \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + L^1 \cdot \begin{pmatrix} 0 & 1 \\ -K & 0 \end{pmatrix} + L^2 \cdot \begin{pmatrix} -\frac{1}{2}K & 0 \\ 0 & -\frac{1}{2}K \end{pmatrix} + \dots \quad (3.32)$$

If we keep only terms up to first order in L we get:

$$M_{s \rightarrow s+L} = L^0 \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + L^1 \cdot \begin{pmatrix} 0 & 1 \\ -K & 0 \end{pmatrix} + O(L^2) \quad (3.33)$$

$$M_{s \rightarrow s+L} = \begin{pmatrix} 1 & L \\ -K \cdot L & 1 \end{pmatrix} + O(L^2) \quad (3.34)$$

This map is precise to order $O(L^1)$, but since we have $\det M \neq 1$, this truncated expansion is not symplectic.

3.6.3.3 Symplectic Matrices and Symplectic Integration

However, the map (3.34) can be made symplectic by adding a term $-K^2L^2$. This term is of order $O(L^2)$, i.e. does not deteriorate the approximation because the inaccuracy is of the same order.

$$M_{s \rightarrow s+L} = \begin{pmatrix} 1 & L \\ -K \cdot L & 1 - K^2L^2 \end{pmatrix} \quad (3.35)$$

Following the same procedure we can compute a symplectic approximation precise to order $O(L^2)$ from (3.32) using:

$$M_{s \rightarrow s+L} = \begin{pmatrix} 1 - \frac{1}{2}KL^2 & L \\ -K \cdot L & 1 - \frac{1}{2}KL^2 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 - \frac{1}{2}KL^2 & L - \frac{1}{4}KL^3 \\ -K \cdot L & 1 - \frac{1}{2}KL^2 \end{pmatrix} \quad (3.36)$$

It can be shown that this ‘‘symplectification’’ corresponds to the approximation of a quadrupole by a single kick in the centre between two drift spaces of length $L/2$:

$$\begin{pmatrix} 1 & \frac{1}{2}L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -K \cdot L & 1 \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{2}L \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 - \frac{1}{2}KL^2 & L - \frac{1}{4}KL^3 \\ -K \cdot L & 1 - \frac{1}{2}KL^2 \end{pmatrix} \quad (3.37)$$

It may be mentioned that the previous approximation to 1st order corresponds to a kick at the **end** of a quadrupole, preceded by a drift space of length L . Both cases are illustrated in Fig. 3.1.

One can try to further improve the approximation by adding 3 kicks like in Fig. 3.2 where the distance between kicks and the kick strengths are optimized to obtain the highest order. The thin lens approximation in Fig. 3.2 with the constants:

$$a \approx 0.675602, b \approx -0.175602, \alpha \approx 1.351204, \beta \approx -1.702410 \quad (3.38)$$

provides an $O(L^4)$ integrator [11].

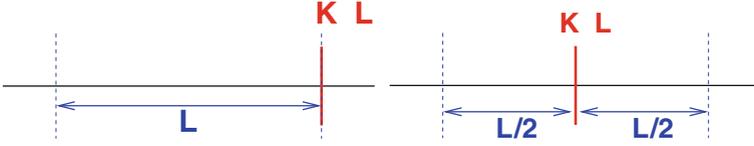


Fig. 3.1 Schematic representation of a symplectic kick of first order (left) and second order (right)

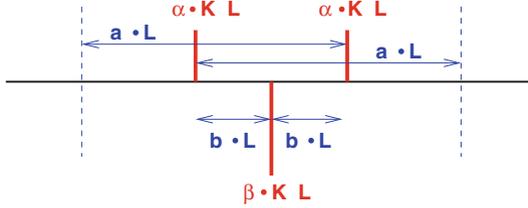


Fig. 3.2 Schematic representation of a symplectic integration with thin lenses of fourth order. The figure shows the size of drifts and thin lens kicks

This process is a Symplectic Integration [12] and is a formal procedure to construct higher order integrators from lower order ones. From a 2nd order scheme (1 kick) $S_2(t)$ we construct a 4th order scheme (3 kicks = 3×1 kick) like: $S_4(t) = S_2(x_1t) \circ S_2(x_0t) \circ S_2(x_1t)$ with:

$$x_0 = \frac{-2^{1/3}}{2 - 2^{1/3}} \approx -1.702410 \quad x_1 = \frac{1}{2 - 2^{1/3}} \approx 1.351204 \quad (3.39)$$

In general: If $S_{2k}(t)$ is a symmetric integrator of order $2k$, then we obtain a symmetric integrator of order $2k + 2$ by: $S_{2k+2}(t) = S_{2k}(x_1t) \circ S_{2k}(x_0t) \circ S_{2k}(x_1t)$ with:

$$x_0 = \frac{-2^{k+1}\sqrt{2}}{2 - 2^{k+1}\sqrt{2}} \quad x_1 = \frac{1}{2 - 2^{k+1}\sqrt{2}} \quad (3.40)$$

Higher order integrators can be obtained in a similar way in an iterative procedure. A very explicit example of the iterative construction of a higher order map from a lower order can be found in [7].

This method can be applied to any other non-linear map and we obtain the same integrators. The proof of this statement and the systematic extension can be done in the form of Lie operators [12].

It should be noted that higher order integrators require maps which drift backwards (3.38) as shown in Fig. 3.2 right. This has two profound consequences. First, a straightforward “physical” interpretation of thin lens models representing drifts and individual small “magnets” (a la MAD) makes no sense and prohibits the

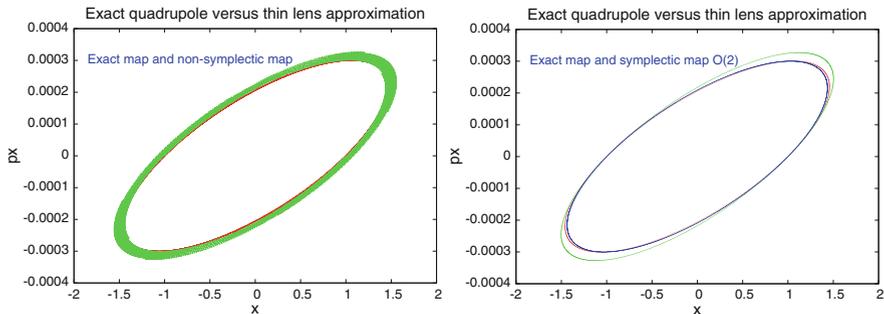


Fig. 3.3 Poincaré section for tracking through a quadrupole. Comparison between exact solution, non-symplectic (left) and symplectic (right) tracking. Shown are symplectic integrators of order 1 and 2

use of high order integrators. Secondly, models which require self-consistent time tracking or s tracking (e.g. space charge calculations) must use integrators for which $s(t)$ is monotonic in the magnets.

3.6.3.4 Comparison Symplectic Versus Non-symplectic Integration

A demonstration of a non-symplectic tracking is shown in Fig. 3.3. A particle is tracked through a quadrupole and the Poincaré section is shown. A quadrupole is chosen because it allows a comparison with the exact solution. The non-symplecticity causes the particle to spiral outwards. As comparison to the exact tracking is shown. In Fig. 3.3 (right) the symplectic integrators of order 1 and 2 as derived above are used instead. The trajectory is now constant and the difference to the exact solution is small. Although the model is approximated but symplectic, the underlying physics (i.e. constant energy in this case) is correct at the expense of a small discrepancy with respect to the exact solution.

3.7 Hamiltonian Treatment of Electro-Magnetic Fields

A frequently asked question is why one should not just use Newton’s laws and the Lorentz force. Some of the main reasons are:

- Newton requires rectangular coordinates and time, trajectories with e.g. “curvature” or “torsion” need to introduce “reaction forces”. (For example: LHC has locally non-planar (cork-screw) “design” orbits!).
- For linear dynamics done by ad hoc introduction of new coordinate frame.

- With Hamiltonian it is free: The formalism is “coordinate invariant”, i.e. the equations have the same form in every coordinate system.
- The basic equations ensure that the phase space is conserved

3.7.1 Lagrangian of Electro-Magnetic Fields

3.7.1.1 Lagrangian and Hamiltonian

It is common practice to use q for the coordinates when Hamiltonian and Lagrangian formalisms are used. This is deplorable because q is also used for particle charge.

The motion of a particle is usually described in classical mechanics using the Lagrange functional:

$$L(q_1(t), \dots, q_n(t), \dot{q}_1(t), \dots, \dot{q}_n(t), t) \quad \text{short:} \quad L(q_i, \dot{q}_i, t) \quad (3.41)$$

where $q_1(t), \dots, q_n(t)$ are generalized coordinates and $\dot{q}_1(t), \dots, \dot{q}_n(t)$ the corresponding generalized velocities. Here q_i can stand for any coordinate and any particle, and n can be a very large number.

The integral

$$S = \int L(q_i(t), \dot{q}_i(t), t) dt. \quad (3.42)$$

defines the action S .

The action S is used with the Hamiltonian principle: a system moves along a path such that the action S becomes stationary, i.e. $\delta S = 0$

Is fulfilled when:

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0 \quad (\text{Euler} - \text{Lagrange equation}) \quad (3.43)$$

It is unfortunate that the term action is used in different contexts and must not be confused with the action-angle variables defined earlier. The action above is a functional rather than a variable.

Without proof or derivation it should be stated that $L = T - V =$ kinetic energy – potential energy.

Given the Lagrangian, the Hamiltonian can be derived as:

$$H(\vec{q}, \vec{p}, t) = \sum_i [p_i \dot{q}_i - L(\vec{q}, \vec{\dot{q}}, t)]. \quad (3.44)$$

The coordinates q_i are identical to those in the Lagrangian (3.41), whereas the conjugate momenta p_i are derived from L as:

$$p_i = \frac{\partial L}{\partial \dot{q}_i}. \quad (3.45)$$

3.7.2 Hamiltonian with Electro-Magnetic Fields

Readers only interested in the final result can skip Eqs. (3.46)–(3.54).

A key for the correct Hamiltonian is the relativistic treatment. An intuitive derivation is presented here, a simpler and elegant derivation should be based on 4-vectors [13]. The action S must be a relativistic invariant and becomes (now using coordinates x and velocities v):

$$S = \int L(x_i(t), v_i(t), t) \gamma \cdot d\tau. \quad (3.46)$$

since the proper time τ is Lorentz invariant, and therefore also $\gamma \cdot L$.

The Lagrangian for a free particle is usually a function of the velocity (see classical formula of the kinematic term), but must not depend on its position.

The only Lorentz invariant with the velocity is [13]:

$$U^\mu U_\mu = c^2 \quad (3.47)$$

where U is the four-velocity.

For the Lagrangian of a (relativistic) free particle we must write

$$L_{free} = -mc^2 \sqrt{1 - \beta_r^2} = -mc^2 \sqrt{1 - \left(\frac{v}{c}\right)^2} = -\frac{mc^2}{\gamma} \quad (3.48)$$

Using for the electromagnetic Lagrangian a form (without derivation, any textbook):

$$L = \frac{e}{c} v \cdot \vec{A} - e\phi \quad (3.49)$$

Combining (3.48) and (3.49) we obtain the complete Lagrangian:

$$L = -\frac{mc^2}{\gamma} + \frac{e}{c} \cdot \vec{v} \cdot \vec{A} - e \cdot \phi \quad (3.50)$$

thus the conjugate momentum is derived as:

$$\vec{P} = \frac{\partial L}{\partial v_i} = \vec{p} + \frac{e}{c} \vec{A} \quad \left(\text{or} \quad \vec{P} = \vec{p} - \frac{q}{c} \vec{A} \right) \quad (3.51)$$

where \vec{p} is the ordinary kinetic momentum.

A consequence is that the canonical momentum cannot be written as:

$$P_x = mc\gamma\beta_x \quad (3.52)$$

Using the conjugate momentum the Hamiltonian takes the simple form:

$$H = \vec{P} \cdot \vec{v} - L \quad (3.53)$$

The Hamiltonian must be a function of the conjugate variables P and x and after a bit of algebra one can eliminate \vec{v} using:

$$\vec{v} = \frac{c\vec{P} - e\vec{A}}{\sqrt{(\vec{P} - \frac{e\vec{A}}{c})^2 + m^2c^2}} \quad (3.54)$$

With (3.50) and (3.54) we write for the Hamiltonian for a (ultra relativistic, i.e. $\gamma \gg 1$, $\beta \approx 1$) particle in an electro-magnetic field is given by:

$$H(\vec{x}, \vec{p}, t) = c\sqrt{(\vec{P} - e\vec{A}(\vec{x}, t))^2 + m^2c^2} + e\Phi(\vec{x}, t) \quad (3.55)$$

where $\vec{A}(\vec{x}, t)$, $\Phi(\vec{x}, t)$ are the vector and scalar potentials.

Interlude 2

A short interlude, one may want to skip to Eq. (3.60)

Equation (3.55) is the total energy E of the particle where the difference is the potential energy $e\phi$ and the new conjugate momentum $\vec{P} = (\vec{p} - \frac{e}{c}\vec{A})$, replacing \vec{p} .

From the classical expression

$$E^2 = p^2c^2 + (mc^2)^2 \quad (3.56)$$

one can re-write

$$(W - e\phi)^2 - (c\vec{P} - e\vec{A})^2 = (mc^2)^2 \quad (3.57)$$

(continued)

The expression $(mc^2)^2$ is the invariant mass [13], i.e.

$$p_\mu p^\mu = (mc)^2 \quad (3.58)$$

with the 4-vector for the momentum [13]:

$$p^\mu = \left(\frac{E}{c}, \vec{p} \right) = \left(\frac{1}{c}(W - e\phi), \vec{P} - \left(\frac{e}{c} \vec{A} \right) \right) \quad (3.59)$$

The changes are a consequence using 4-vectors in the presence of electromagnetic fields (potentials).

An interesting consequence of (3.51) is that the momentum is linked to the fields (\vec{A}) and the angle x' cannot easily be derived from the total momentum and the conjugate momentum. That is using (x, x') as coordinate are strictly speaking not valid in the presence of electromagnetic fields.

In this context using (x, x') or (x, p_x) is not equivalent. A general, strong statement that (x, x') is used in accelerator physics is at best bizarre.

3.7.3 Hamiltonian Used for Accelerator Physics

In a more convenient (and useful) form, using canonical variables x and p_x, p_y and the design path length s as independent variable (bending field B_0 in y -plane) and no electric fields (for details of the derivation see [14]):

$$H = \underbrace{-\left(1 + \frac{x}{\rho}\right)}_{\text{due to } t \rightarrow s} \cdot \underbrace{\sqrt{(1 + \delta)^2 - p_x^2 - p_y^2}}_{\text{kinematic}} + \underbrace{\left(\frac{x}{\rho} + \frac{x^2}{2\rho^2}\right)}_{\text{due to } t \rightarrow s} - \underbrace{\frac{A_s(x, y)}{B_0\rho}}_{\text{normalized}} \quad (3.60)$$

where $p = \sqrt{E^2/c^2 - m^2c^2}$ total momentum, $\delta = (p - p_0)/p_0$ is relative momentum deviation and $A_s(x, y)$ (normalized) longitudinal (along s) component of the vector potential. Only transverse field and no electric fields are considered.

After square root expansion and sorting the A_s contributions:

$$H = \overbrace{\frac{p_x^2 + p_y^2}{2(1 + \delta)}}^{\text{kinematic}} - \overbrace{\left(\frac{x\delta}{\rho} + \frac{x^2}{2\rho^2} \right)}^{\text{dipole}} + \overbrace{\frac{k_1}{2}(x^2 - y^2)}^{\text{quadrupole}} + \overbrace{\frac{k_2}{6}(x^3 - 3xy^2)}^{\text{sextupole}} + \dots$$

$\underbrace{\hspace{10em}}_{\text{bending}}$
 $\underbrace{\hspace{10em}}_{\text{focusing}}$

(3.61)

$$\text{using : } k_n = k_n^{(n)} = \frac{1}{B\rho} \frac{\partial^n B_y}{\partial x^n} \quad \left(k_n^{(s)} = \frac{1}{B\rho} \frac{\partial^n B_x}{\partial x^n} \right)$$
(3.62)

- The Hamiltonian describes the motion of a particle through an element
- Each element has a component in the Hamiltonian
- Basis to extend the linear to a nonlinear formalism

A short list of Hamiltonians of some machine elements (3D)

In general for multipoles of order n :

$$H_n = \frac{1}{1+n} \text{Re} \left[(k_n + ik_n^{(s)})(x + iy)^{n+1} \right] + \frac{p_x^2 + p_y^2}{2(1 + \delta)}$$
(3.63)

We get for some important types (normal components k_n only):

$$\text{drift space : } H = -\sqrt{(1 + \delta)^2 - p_x^2 - p_y^2} \approx \frac{p_x^2 + p_y^2}{2(1 + \delta)}$$
(3.64)

$$\text{dipole : } H = -\frac{-x\delta}{\rho} + \frac{x^2}{2\rho^2} + \frac{p_x^2 + p_y^2}{2(1 + \delta)}$$
(3.65)

$$\text{quadrupole : } H = \frac{1}{2}k_1(x^2 - y^2) + \frac{p_x^2 + p_y^2}{2(1 + \delta)}$$
(3.66)

$$\text{sextupole : } H = \frac{1}{3}k_2(x^3 - 3xy^2) + \frac{p_x^2 + p_y^2}{2(1 + \delta)}$$
(3.67)

$$\text{octupole : } H = \frac{1}{4}k_3(x^4 - 6x^2y^2 + y^4) + \frac{p_x^2 + p_y^2}{2(1 + \delta)}$$
(3.68)

Interlude 3

A few remarks are required after this list of Hamiltonian for particular elements.

- Unlike said in many introductory textbooks and lectures, a multipole of order n is not required to drive a n th order resonance—nothing could be more wrong!!
- In leading order perturbation theory, only elements with an even order (and larger than 2) in the Hamiltonian can produce an amplitude dependent tune shift and tune spread.

3.7.3.1 Lie Maps and Transformations

In this chapter we would like to introduce Lie algebraic tools and Lie transformations [15–17]. We use the symbol $z_i = (x_i, p_i)$ where x and p stand for canonically conjugate position and momentum. We let $f(z)$ and $g(z)$ be any function of x, p and can define the Poisson bracket for a differential operator [18]:

$$[f, g] = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial x_i} \right) \quad (3.69)$$

Assuming that the motion of a dynamic system is defined by a Hamiltonian H , we can now write for the equations of motion [18]:

$$[x_i, H] = \frac{\partial H}{\partial p_i} = \frac{dx_i}{dt} \quad (3.70)$$

$$[p_i, H] = -\frac{\partial H}{\partial x_i} = \frac{dp_i}{dt} \quad (3.71)$$

If H does not explicitly depend on time then:

$$[f, H] = 0 \quad (3.72)$$

implies that f is an invariant of the motion. To proceed, we can define a Lie operator $: f :$ via the notation:

$$: f : g = [f, g] \quad (3.73)$$

where $: f :$ is an operator acting on the function g .

We can define powers as:

$$(: f :)^2 g =: f : (: f : g) = [f, [f, g]] \quad \text{etc.} \quad (3.74)$$

One can collect a set of useful formulae for calculations:

Some common special (very useful) cases for f :

$$\begin{aligned} : x : &= \frac{\partial}{\partial p} & : p : &= -\frac{\partial}{\partial x} \\ : x :^2 &= \overbrace{: x :: x :}^{\text{applied twice}} = \frac{\partial^2}{\partial p^2} & : p :^2 &= \overbrace{: p :: p :}^{\text{applied twice}} = \frac{\partial^2}{\partial x^2} \\ : xp : &= p \frac{\partial}{\partial p} - x \frac{\partial}{\partial x} & : x :: p : &= : p :: x : = -\frac{\partial^2}{\partial x \partial p} \\ : x^2 : &= 2x \frac{\partial}{\partial p} & : p^2 : &= -2p \frac{\partial}{\partial x} \\ : x^n : &= n \cdot x^{n-1} \frac{\partial}{\partial p} & : p^n : &= -n \cdot p^{n-1} \frac{\partial}{\partial x} \end{aligned} \quad (3.75)$$

Once powers of the Lie operators are defined, they can be used to formulated an exponential form:

$$e^{:f:} = \sum_{i=0}^{\infty} \frac{1}{i!} (: f :)^i \quad (3.76)$$

This expression is call a ‘‘Lie transformation’’.

Give the Hamiltonian H of an element, the generator f is this Hamiltonian multiplied by the length L of the element.

To evaluate a simple example, for the case $H = -p^2/2$ using the exponential form and (3.75):

$$\begin{aligned} e^{-Lp^2/2} : x &= x - \frac{1}{2}L : p^2 : x + \frac{1}{8}L^2 (: p^2 :)^2 x + .. \\ &= x + Lp \end{aligned} \quad (3.77)$$

$$\begin{aligned} e^{-Lp^2/2} : p &= p - \frac{1}{2}L : p^2 : p + ... \\ &= p \end{aligned} \quad (3.78)$$

One can easily verify that for 1D and $\delta = 0$ this is the transformation of a drift space of length L (if $p \approx x'$) as introduced previously. The function $f(x, p) = -Lp^2/2$ is the generator of this transformation.

Interlude 4

The exact Hamiltonian in two transverse dimensions and with a relative momentum deviation δ is (full Hamiltonian with $\vec{A}(\vec{x}, t) = 0$):

$$H = -\sqrt{(1 + \delta)^2 - p_x^2 - p_y^2} \quad \longrightarrow \quad f_{drift} = L \cdot H$$

The exact map for a drift space is now:

$$x^{new} = x + L \cdot \frac{p_x}{\sqrt{(1 + \delta)^2 - p_x^2 - p_y^2}}$$

$$p_x^{new} = p_x$$

$$y^{new} = y + L \cdot \frac{p_y}{\sqrt{(1 + \delta)^2 - p_x^2 - p_y^2}}$$

$$p_y^{new} = p_y$$

In 2D and with $\delta \neq 0$ it is more complicated than Eq. (3.78). In practice the map can (often) be simplified to the well known form.

More general, acting on the phase space coordinates:

$$e^{:f:}(x, p)_1 = (x, p)_2 \quad (3.79)$$

is the Lie transformation which describes how to go from one point to another.

While a Lie operator propagates variables over an infinitesimal distance, the Lie transformation propagates over a finite distance.

To illustrate this technique with some simple examples, it can be shown easily, using the formulae above, that the transformation:

$$e^{:-\frac{1}{2f}x^2:} \quad (3.80)$$

corresponds to the map of a thin quadrupole with focusing length f , i.e.

$$\begin{aligned} x_2 &= x_1 \\ p_2 &= p_1 - \frac{1}{f}x_1 \end{aligned}$$

A transformation of the form:

$$e^{\dot{-}\frac{1}{2}L(k^2x^2 + p^2)} : \quad (3.81)$$

corresponds to the map of a thick quadrupole with length L and strength k :

$$x_2 = x_1 \cos(kL) + \frac{p_1}{k} \sin(kL) \quad (3.82)$$

$$p_2 = -kx_1 \sin(kL) + p_1 \cos(kL) \quad (3.83)$$

The linear map using Twiss parameters in Lie representation (we shall call it : f_2 : from now on) is always of the form:

$$e^{\dot{f}_2} : \text{ with } : f_2(x) = -\frac{\mu}{2}(\gamma x^2 + 2\alpha x p + \beta p^2) \quad (3.84)$$

In case of a general non-linear function $f(x)$, i.e. with a (thin lens) kick like:

$$x_2 = x_1 \quad (3.85)$$

$$p_2 = p_1 + f(x_1) \quad (3.86)$$

the corresponding Lie operator can be written as:

$$e^{\dot{h}} : = e^{\dot{\int}_0^x f(u)du} : \quad \text{or} \quad e^{\dot{F}} : \quad \text{with} \quad F = \int_0^x f(u)du. \quad (3.87)$$

An important property of the Lie transformation is that the one turn map is the exponential of the effective Hamiltonian and the circumference C :

$$M_{ring} = e^{\dot{-}CH_{eff}}. \quad (3.88)$$

The main advantages of Lie transformations are that the exponential form is always symplectic and that a formalism exists for the concatenation of transformations. An overview of this formalism and many examples can be found in [7]. As for the Lie operator, one can collect a set of useful formulae. Another neat package with useful formulae:

With a constant and f, g, h arbitrary functions:

$$\begin{aligned} : a : &= 0 \quad \longrightarrow \quad e^{\dot{a}} : = 1 \\ : f : a &= 0 \quad \longrightarrow \quad e^{\dot{f}} : a = a \end{aligned}$$

$$e^{\cdot f} \cdot [g, h] = [e^{\cdot f} \cdot g, e^{\cdot f} \cdot h]$$

$$e^{\cdot f} \cdot (g \cdot h) = e^{\cdot f} \cdot g \cdot e^{\cdot f} \cdot h$$

and very important:

$$M g(x) = e^{\cdot f} \cdot g(x) = g(e^{\cdot f} \cdot x) \quad \text{e.g.} \quad e^{\cdot f} \cdot x^2 = (e^{\cdot f} \cdot x)^2$$

$$M^{-1} g(x) = (e^{\cdot f} \cdot)^{-1} g(x) = e^{-\cdot f} \cdot g(x) \quad \text{note:} \quad \frac{1}{e^{\cdot f} \cdot} \neq (e^{\cdot f} \cdot)^{-1}$$

(3.89)

3.7.3.2 Concatenation of Lie Transformations

The concatenation is very easy when f and g commute (i.e. $[f, g] = [g, f] = 0$) and we have:

$$e^{\cdot h} \cdot = e^{\cdot f} \cdot e^{\cdot g} \cdot = e^{\cdot f + g} \cdot \tag{3.90}$$

The generators of the transformations can just be added.

To combine two transformations in the general case (i.e. $[f, g] \neq 0$) we can use the Baker–Campbell–Hausdorff formula (BCH) which in our convention can be written as:

$$h = f + g + \frac{1}{2} \cdot f \cdot g + \frac{1}{12} \cdot f \cdot^2 g + \frac{1}{12} \cdot g \cdot^2 f$$

$$+ \frac{1}{24} \cdot f \cdot \cdot g \cdot^2 f - \frac{1}{720} \cdot g \cdot^4 f$$

$$- \frac{1}{720} \cdot f \cdot^4 g + \frac{1}{360} \cdot g \cdot \cdot f \cdot^3 g + \dots \tag{3.91}$$

In many practical cases, non-linear perturbations are localized and small compared to the rest of the (often linear) ring, i.e. one of f or g is much smaller, e.g. f corresponds to one turn, g to a small, local distortion.

In that case we can sum up the BCH formula to first order in the perturbation g and get:

$$e^{\cdot h} \cdot = e^{\cdot f} \cdot e^{\cdot g} \cdot = \exp \left[\cdot f + \left(\frac{\cdot f \cdot}{1 - e^{-\cdot f} \cdot} \right) g + O(g^2) \cdot \right] \tag{3.92}$$

When g is small compared to f , the first order is a good approximation.

For example, we may have a full ring $e^{\cdot f_2 \cdot}$ with a small (local) distortion, e.g. a multipole $e^{\cdot g \cdot}$ with $g = kx^n$ then the expression:

$$e^{\cdot h} \cdot = e^{\cdot f_2 \cdot} \cdot e^{\cdot kx^n \cdot} ; \tag{3.93}$$

allows the evaluation of the invariant h for a single multipole of order n in this case.

In the case that f_2, f_3, f_4 , are 2nd, 3rd, 4th order polynomials (Dragt-Finn factorization [19]):

$$e : f : = e : f_2 : e : f_3 : e : f_4 : , \quad (3.94)$$

each term is symplectic and the truncation at any order does not violate symplecticity.

One may argue that this method is clumsy when we do the analysis of a linear system. The reader is invited to prove this by concatenating by hand a drift space and a thin quadrupole lens. However, the central point of this method is that the technique works whether we do linear or non-linear beam dynamics and provides a formal procedure. Lie transformations are the natural extension of the linear matrix formalism to a non-linear formalism. There is no need to move from one method to another as required in the traditional treatment.

In the case an element is described by a Hamiltonian H , the Lie map of an element of length L and the Hamiltonian H is:

$$e^{-L : H :} = \sum_{i=0}^{\infty} \frac{1}{i!} (-L : H :)^i \quad (3.95)$$

For example, the Hamiltonian for a thick sextupole is:

$$H = \frac{1}{3}k(x^3 - 3xy^2) + \frac{1}{2}(p_x^2 + p_y^2) \quad (3.96)$$

To find the transformation we search for:

$$e^{-L : H :} : x \quad \text{and} \quad e^{-L : H :} : p_x \quad \text{i.e. for} \quad (3.97)$$

$$e^{-L : H :} : x = \sum_{i=0}^{\infty} \frac{-L^i}{i!} (: H :)^i x \quad (3.98)$$

We can compute:

$$: H :^i x \quad \text{for each } i \quad (3.99)$$

to get:

$$: H :^1 x = -p_x, \quad (3.100)$$

$$: H :^2 x = -k(x^2 - y^2), \quad (3.101)$$

$$: H :^3 x = 2k(xp_x - yp_y), \tag{3.102}$$

$$\dots \tag{3.103}$$

Putting the terms together one obtains:

$$e^{-L} : H : x = x + p_x L - \frac{1}{2}kL^2(x^2 - y^2) - \frac{1}{3}kL^3(xp_x - yp_y) + \dots \tag{3.104}$$

3.7.4 Analysis Techniques: Poincare Surface of Section

Under normal circumstances it is not required to examine the complete time development of a particle trajectory around the machine. Given the experimental fact that the trajectory can be measured only at a finite number of positions around the machine, it is only useful to sample the trajectory periodically at a fixed position. The plot of the rate of change of the phase space variables at the beginning (or end) of each period is the appropriate method and also known as Poincare Surface of Section [20]. An example of such a plot is shown in Fig. 3.4 where the one-dimensional phase space is plotted for a completely linear machine (Fig. 3.4, left) and close to a 5th order resonance in the presence of a single non-linear element (in this case a sextupole) in the machine (Fig. 3.4, right).

It shows very clearly the distortion of the phase space due to the non-linearity, the appearance of resonance islands and chaotic behaviour between the islands. From this plot is immediately clear that the region of stability is strongly reduced in the

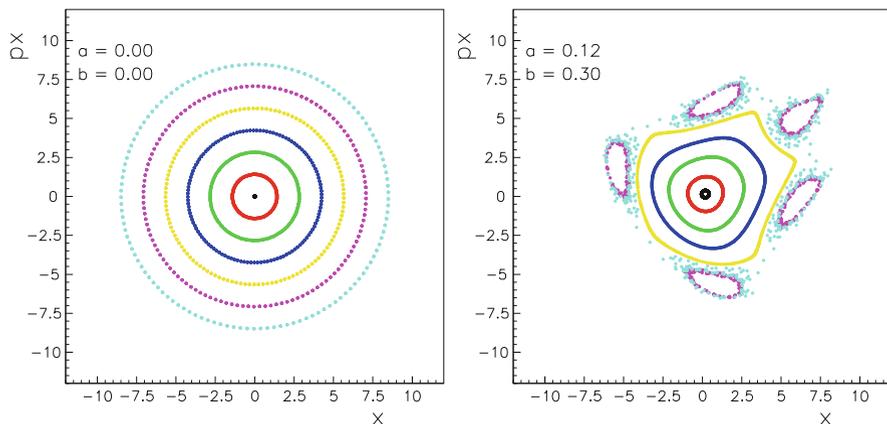


Fig. 3.4 Poincare surface of section of a particle near the 5th order resonances. Left without non-linear elements, right with one sextupole

presence of the non-linear element. The main features we can observe in Fig. 3.4 are that particles can:

- Move on closed curves
- Lie on islands, i.e. jump from one island to the next from turn to turn
- Move on chaotic trajectories

The introduction of these techniques by Poincare mark a paradigm shift from the old classical treatment to a more modern approach. The question of long term stability of a dynamic system is not answered by getting the solution to the differential equation of motion, but by the determination of the properties of the surface where the motion is mapped out. Independent how this surface of section is obtained, i.e. by analytical or numerical methods, its analysis is the key to understand the stability.

3.7.5 Analysis Techniques: Normal Forms

The idea behind this technique is that maps can be transformed into Normal Forms. This tool can be used to:

- Study invariants of the motion and the effective Hamiltonian
- Extract non-linear tune shifts (detuning)
- Perform resonance analysis

In the following we demonstrate the use of normal forms away from resonances. The treatment of the beam dynamics close to resonances is beyond the scope of this review and can be found in the literature (see e.g. [6, 7]).

3.7.5.1 Normal Form Transformation: Linear Case

The strategy is to make a transformation to get a simpler form of the map M , e.g. a pure rotation $R(\Delta\mu)$ as schematically shown in Fig. 3.5 using a transformation like:

$$M = U \circ R(\Delta\mu) \circ U^{-1} \quad \text{or} : \quad R(\Delta\mu) = U^{-1} \circ M \circ U \quad (3.105)$$

with

$$U = \begin{pmatrix} \sqrt{\beta(s)} & 0 \\ -\frac{\alpha(s)}{\sqrt{\beta(s)}} & \frac{1}{\sqrt{\beta(s)}} \end{pmatrix} \quad \text{and} \quad R = \begin{pmatrix} \cos(\Delta\mu) & \sin(\Delta\mu) \\ -\sin(\Delta\mu) & \cos(\Delta\mu) \end{pmatrix} \quad (3.106)$$

This transformation corresponds to the Courant-Snyder analysis in the linear case and directly provides the phase advance and optical parameters. The optical

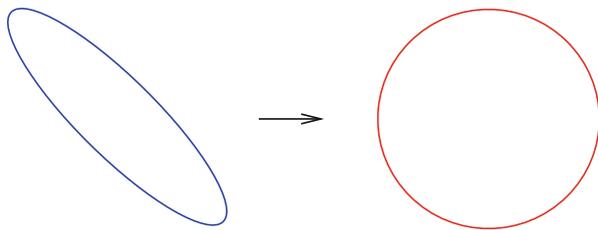


Fig. 3.5 Normal form transformation in the linear case, related to the Courant-Snyder analysis

parameters emerge automatically from the normal form analysis of the one-turn-map.

Although not required in the linear case, we demonstrate how this normal form transformation is performed using the Lie formalism. Starting from the general expression:

$$R(\Delta\mu) = U^{-1} \circ M \circ U \quad (3.107)$$

we know that a linear map M in Lie representation is always:

$$e: f_2: \quad \text{with:} \quad f_2 = -\frac{\mu}{2}(\gamma x^2 + 2\alpha x p_x + \beta p_x^2) \quad (3.108)$$

therefore:

$$\begin{aligned} R(\Delta\mu) &= U^{-1} \circ e: f_2(x): \circ U \\ &= e^{U^{-1}}: f_2: U = e: U^{-1} f_2: \end{aligned} \quad (3.109)$$

and (with $U^{-1} f_2$) f_2 expressed in the new variables X, P_x it assumes the form:

$$f_2 = -\frac{\mu}{2}(X^2 + P_x^2) \quad \text{because:} \quad \begin{pmatrix} X \\ P_x \end{pmatrix} = U^{-1} \begin{pmatrix} x \\ p_x \end{pmatrix} \quad (3.110)$$

i.e. with the transformation U^{-1} the rotation $: f_2 :$ becomes a circle in the transformed coordinates. We transform to action and angle variables J and Φ , related to the variables X and P_x through the transformations:

$$X = \sqrt{2J\beta} \sin \Phi, \quad P_x = \sqrt{\frac{2J}{\beta}} \cos \Phi \quad (3.111)$$

With this transformation we get a simple representation for the linear transfer map f_2 :

$$f_2 = -\mu J \quad \text{and} \quad R(\Delta\mu) = e^{i-\Delta\mu J} : \tag{3.112}$$

3.7.5.2 Normal Form Transformation: Non-linear Case

In the more general, non-linear case the transformation is more complicated and one must expect that the rotation angle becomes amplitude dependent (see e.g. [6]). A schematic view of this scheme is shown in Fig. 3.6 where the transformation leads to the desired rotation, however the rotation frequency (phase advance) is now amplitude dependent.

We demonstrate the power by a simple example in one dimension, but the treatment is similar for more complex cases. In particular, it demonstrates that this analysis using the algorithm based on Lie transforms leads easily to the desired result. A very detailed discussion of this method is found in [6].

From the general map we have made a transformation such that the transformed map can be expressed in the form e^{ih_2} : where the function h_2 is now a function only of J_x , J_y , and δ and it is the effective Hamiltonian.

In the non-linear case and away from resonances we can get the map in a similar form:

$$N = e^{i h_{eff}(J_x, J_y, \delta)} : \tag{3.113}$$

where the effective Hamiltonian h_{eff} depends only on J_x , J_x , and δ .

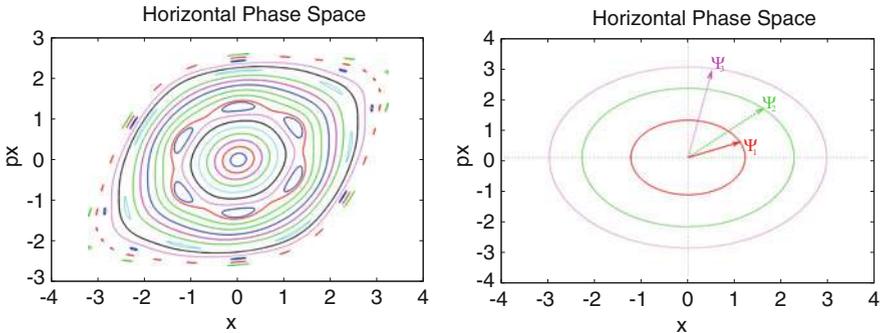


Fig. 3.6 Normal form transformation in the non-linear case, leading to amplitude dependent phase advance. The transformation was done for non-resonant amplitudes

If the map for h_{eff} corresponds to a one-turn-map, we can write for the tunes:

$$Q_x(J_x, J_y, \delta) = \frac{1}{2\pi} \frac{\partial h_{eff}}{\partial J_x} \quad (3.114)$$

$$Q_y(J_x, J_y, \delta) = \frac{1}{2\pi} \frac{\partial h_{eff}}{\partial J_y} \quad (3.115)$$

and the change of path length:

$$\Delta s = -\frac{\partial h_{eff}}{\partial \delta} = \alpha_c \delta \quad (3.116)$$

In the non-linear case, particles with different J_x, J_y, δ have different tunes. Their dependence on J_x, J_y is the amplitude detuning, the dependence on δ are the chromaticities.

The effective Hamiltonian can always be written (here to 3rd order) in a form:

$$h_{eff} = +\mu_x J_x + \mu_y J_y + \frac{1}{2} \alpha_c \delta^2 \quad (3.117)$$

$$+ c_{x1} J_x \delta + c_{y1} J_y \delta + c_3 \delta^3 \quad (3.118)$$

$$+ c_{xx} J_x^2 + c_{xy} J_x J_y + c_{yy} J_y^2 + c_{x2} J_x \delta^2 + c_{y2} J_y \delta^2 + c_4 \delta^4 \quad (3.119)$$

and then tune depends on action J and momentum deviation δ :

$$Q_x(J_x, J_y, \delta) = \frac{1}{2\pi} \frac{\partial h_{eff}}{\partial J_x} = \frac{1}{2\pi} \left(\underbrace{\mu_x + 2c_{xx} J_x + c_{xy} J_y}_{\text{detuning}} + \underbrace{c_{x1} \delta + c_{x2} \delta^2}_{\text{chromaticity}} \right) \quad (3.120)$$

$$Q_y(J_x, J_y, \delta) = \frac{1}{2\pi} \frac{\partial h_{eff}}{\partial J_y} = \frac{1}{2\pi} \left(\underbrace{\mu_y + 2c_{yy} J_y + c_{xy} J_x}_{\text{detuning}} + \underbrace{c_{y1} \delta + c_{y2} \delta^2}_{\text{chromaticity}} \right) \quad (3.121)$$

The meaning of the different contributions are:

- μ_x, μ_y : linear phase advance or $2\pi \cdot$ i.e. the tunes for rings
- $\frac{1}{2} \alpha_c, c_3, c_4$: linear and nonlinear “momentum compaction”
- c_{x1}, c_{y1} : first order chromaticities
- c_{x2}, c_{y2} : second order chromaticities

– c_{xx}, c_{xy}, c_{yy} : detuning with amplitude

The coefficients are the various aberrations of the optics.

As a first example one can look at the effect of a single (thin) sextupole. The map is:

$$\mathcal{M} = e^{- : \mu J_x + \mu J_y + \frac{1}{2} \alpha_c \delta^2 : } e^{ : k(x^3 - 3xy^2) + \frac{p_x^2 + p_y^2}{2(1+\delta)} : } \quad (3.122)$$

we get for h_{eff} (see e.g. [6, 7]):

$$h_{eff} = \mu_x J_x + \mu_y J_y + \frac{1}{2} \alpha_c \delta^2 - k D^3 \delta^3 - 3k \beta_x J_x D \delta + 3k \beta_y J_y D \delta$$

Then it follows:

$$Q_x(J_x, J_y, \delta) = \frac{1}{2\pi} \frac{\partial h_{eff}}{\partial J_x} = \frac{1}{2\pi} (\mu_x - 3k \beta_x D \delta) \quad (3.123)$$

$$Q_y(J_x, J_y, \delta) = \frac{1}{2\pi} \frac{\partial h_{eff}}{\partial J_y} = \frac{1}{2\pi} (\mu_y + 3k \beta_y D \delta) \quad (3.124)$$

Since it was developed to first order only, there is no non-linear detuning with amplitude.

As a second example one can use a linear rotation followed by an octupole, the Hamiltonian is:

$$H = \frac{\mu}{2} (x^2 + p_x^2) + \delta (s - s_0) \frac{x^4}{4} = \mu J + \delta (s - s_0) \frac{x^4}{4} \quad \text{with : } J = \frac{(x^2 + p_x^2)}{2} \quad (3.125)$$

The first part of the Hamiltonian corresponds to the generator of a linear rotation and the second part to the localized octupole.

The map, written in Lie representation becomes:

$$M = e^{(-\frac{\mu}{2} : x^2 + p_x^2 :)} e^{ : \frac{x^4}{4} : } = e^{ : -\mu J : } e^{ : \frac{x^4}{4} : } = R e^{ : \frac{x^4}{4} : } \quad (3.126)$$

The purpose is now to find a generator F for a transformation

$$e^{- : F : } M e^{ : F : } = e^{- : F : } e^{ : \frac{x^4}{4} : } e^{ : F : } \quad (3.127)$$

such that the exponents of the map depend only on J and not on x .

Fig. 3.7 Schematic view of tracking through a complex element



Without going through the algebra (advanced tools exist for this purpose, see e.g. [6]) we quote the result and with

$$F = -\frac{1}{64} \{-5x^4 + 3p_x^4 + 6x^2 p_x^2 + x^3 p_x (8 \cot(\mu) + 4 \cot(2\mu)) + x p_x^3 (8 \cot(\mu) - 4 \cot(2\mu))\} \tag{3.128}$$

we can write the map:

$$M = e^{- : F : e} : -\mu J + \frac{3}{8} J^2 : e : F : \tag{3.129}$$

the term $\frac{3}{8} J^2$ implies a tune shift with amplitude for an octupole.

3.7.6 Truncated Power Series Algebra Based on Automatic Differentiation

It was argued that an appropriate technique to evaluate the behaviour of complex, non-linear systems is by numerically tracking through the individual elements. Schematically this is shown in Fig. 3.7 and the tracking through a complicated system relates the output **numerically** to the input. When the algorithm depicted in Fig. 3.7 represents the full turn in a ring, we obtained the most reliable one-turn-map through this tracking procedure, assuming we have chosen an appropriate representation of the maps for the individual elements.

3.7.6.1 Automatic Differentiation: Concept

This procedure may not be fully satisfactory in all cases and one might like to get an analytical expression for the one-turn-map or equivalent. Could we imagine something that relates the output algebraically to the input? This might for example be a Taylor series of the type:

$$z_2 = \sum C_j z_1^j = \sum d_j f^{(n)} z_1^j \tag{3.130}$$

Then we have an analytic map (for all z_1).

To understand why this could be useful, we can study the paraxial behaviour. In Fig. 3.8 we show schematically the trajectories of particles close to the ideal orbit. The red line refers to the ideal trajectory while the other lines show the motion of

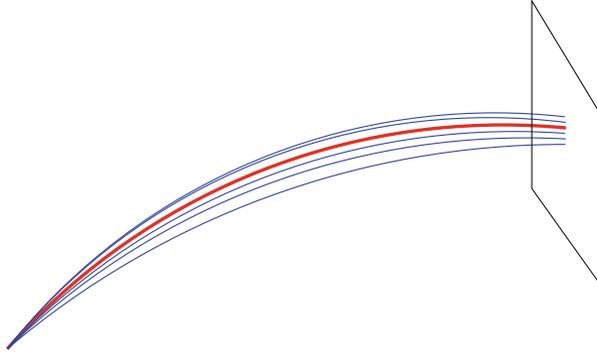


Fig. 3.8 Pictorial view of a paraxial analysis. Red line represents the ideal trajectory

individual particles with small deviations from the ideal path. The idea is that if we understand how small deviations behave, we understand the system much better.

If we now remember the definition of the Taylor series:

$$f(x + \Delta x) = f(x) + \sum_{n=1}^{\infty} \frac{\Delta x^n}{n!} f^{(n)}(x) \quad (3.131)$$

we immediately realize that the coefficients determine the behaviour of small deviations Δx from the ideal orbit x . Therefore the Taylor expansion does a paraxial analysis of the system and the main question is how to get these coefficients without extra work?

The problem is getting the derivatives $f^{(n)}(a)$ of $f(x)$ at a :

$$f'(a) = \lim_{\epsilon \rightarrow 0} \frac{f(a + \epsilon) - f(a)}{\epsilon} \quad (3.132)$$

Numerically this corresponds to the need to subtract almost equal numbers and divide by a small number. For higher orders f'' , f''' .., one must expect numerical problems. An elegant solution to this problem is the use of Differential Algebra (DA) [21].

3.7.6.2 Automatic Differentiation: The Algebra

Here we demonstrate the concept, for more details the literature should be consulted [6, 7, 21].

1. Define a pair (q_0, q_1) , where q_0, q_1 are real numbers
2. Define operations on such pairs like:

$$(q_0, q_1) + (r_0, r_1) = (q_0 + r_0, q_1 + r_1) \quad (3.133)$$

$$c \cdot (q_0, q_1) = (c \cdot q_0, c \cdot q_1) \quad (3.134)$$

$$(q_0, q_1) \cdot (r_0, r_1) = (q_0 \cdot r_0, q_0 \cdot r_1 + q_1 \cdot r_0) \quad (3.135)$$

3. We define the ordering like:

$$(q_0, q_1) < (r_0, r_1) \quad \text{if } q_0 < r_0 \quad \text{or} \quad (q_0 = r_0 \quad \text{and} \quad q_1 < r_1) \quad (3.136)$$

$$(q_0, q_1) > (r_0, r_1) \quad \text{if } q_0 > r_0 \quad \text{or} \quad (q_0 = r_0 \quad \text{and} \quad q_1 > r_1) \quad (3.137)$$

4. This implies that:

$$(0, 0) < (0, 1) < (r, 0) \quad (\text{for any } r) \quad (3.138)$$

This means that $(0,1)$ is between 0 and ANY real number, i.e. it is infinitely small, corresponding to the “ ϵ ” in standard calculus.

Therefore we call this special pair “differential unit” $d = (0, 1)$.

With our rules we can further see that:

$$(1, 0) \cdot (q_0, q_1) = (q_0, q_1) \quad \text{and} \quad (q_0, q_1)^{-1} = \left(\frac{1}{q_0}, -\frac{q_1}{q_0^2} \right) \quad (3.139)$$

In general the inverse of a function $f(q_0, q_1)$ can be derived like:

$$f((q_0, q_1)) \cdot (r_0, r_1) = (1, 0) \quad (3.140)$$

using the multiplication rules. The inverse is then (r_0, r_1) . For example:

$$(q_0, q_1)^2 \cdot (r_0, r_1) = (1, 0) \quad (3.141)$$

gives for the inverse:

$$(r_0, r_1) = \left(\frac{1}{q_0^2}, \frac{-2q_1}{q_0^3} \right) \quad (3.142)$$

3.7.6.3 Automatic Differentiation: The Application

Of course $(q, 0)$ is just the real number q and we define the “real” and the “differential part”:

$$q_0 = R(q_0, q_1) \quad \text{and} \quad q_1 = D(q_0, q_1) \quad (3.143)$$

For a function $f(x)$ we have (without proof, see e.g. [21]):

$$D[f(x + d)] = D[f((x, 0) + (0, 1))] = f'(x) \quad (3.144)$$

We use an example instead to demonstrate this with the function:

$$f(x) = x^2 + \frac{1}{x} \quad (3.145)$$

Using school calculus we have for the derivative:

$$f'(x) = 2x - \frac{1}{x^2} \quad \text{and for } x = 2 \text{ we get: } f(2) = \frac{9}{2}, \quad f'(2) = \frac{15}{4} \quad (3.146)$$

We now apply Automatic Differentiation instead. For the variable x in (3.145) we substitute $x \rightarrow (x, 1) = (2, 1)$ and using our rules:

$$\begin{aligned} f[(2, 1)] &= (x, 1)^2 + (x, 1)^{-1} = (2, 1)^2 + (2, 1)^{-1} \\ &= (4, 4) + \left(\frac{1}{2}, -\frac{1}{4}\right) = \left(\frac{9}{2}, \frac{15}{4}\right) = (f(2), f'(2)) \end{aligned}$$

we arrive at a vector containing the differentials at $x = 2$. The computation of derivatives becomes an algebraic problem, without need for small numbers. No numerical difficulties are expected and the differential is exact.

3.7.6.4 Automatic Differentiation: Higher Orders

To obtain higher orders, we need higher derivatives, i.e. larger dimension for our vectors:

1. The pair $(q_0, 1)$, becomes a vector of length N and with equivalent rules:

$$(q_0, 1) \Rightarrow (q_0, 1, 0, 0, \dots, 0) \quad (3.147)$$

$$(q_0, q_1, q_2, \dots, q_N) + (r_0, r_1, r_2, \dots, r_N) = (s_0, s_1, s_2, \dots, s_N) \quad (3.148)$$

$$c \cdot (q_0, q_1, q_2, \dots, q_N) = (c \cdot q_0, c \cdot q_1, c \cdot q_2, \dots, c \cdot q_N) \tag{3.149}$$

$$(q_0, q_1, q_2, \dots, q_N) \cdot (r_0, r_1, r_2, \dots, r_N) = (s_0, s_1, s_2, \dots, s_N) \tag{3.150}$$

with:

$$s_i = \sum_{k=0}^i \frac{i!}{k!(i-k)!} q_k r_{i-k} \tag{3.151}$$

If we had started with:

$$x = (a, 1, 0, 0, 0 \dots) \tag{3.152}$$

we would get:

$$f(x) = (f(a), f'(a), f''(a), f'''(a), \dots, f^{(n)}(a)) \tag{3.153}$$

Some special cases are:

$$(x, 0, 0, 0, \dots)^n = (x^n, 0, 0, 0, \dots) \tag{3.154}$$

$$(0, 1, 0, 0, \dots)^n = (0, 0, 0, \dots, \overbrace{n!}^{n+1}, 0, 0, \dots) \tag{3.155}$$

$$(x, 1, 0, 0, \dots)^2 = (x^2, 2x, 2, 0, \dots) \tag{3.156}$$

$$(x, 1, 0, 0, \dots)^3 = (x^3, 3x^2, 6x, 6, \dots) \tag{3.157}$$

As another exercise one can consider the function $f(x) = x^{-3}$

$$f(x) \rightarrow f(x, 1, 0, 0, \dots) = \underbrace{(x, 1, 0, 0, \dots)^{-3}} = (f_0, f' f'', f''', \dots)$$

next : multiply both sides with $(x, 1, 0, 0)^3$

$$(1, 0, 0, \dots) = (x, 1, 0, 0, \dots)^3 \cdot (f_0, f', f'', f''', \dots)$$

$$(1, 0, 0, \dots) = (x^3, 3x^2, 6x, 6, \dots) \cdot (f_0, f', f'', f''', \dots) \text{ using (3.157)}$$

$$(1, 0, 0, \dots) = (\underbrace{x^3 \cdot f_0}_{q_0 = 1}, \underbrace{3x^2 \cdot f_0 + x^3 \cdot f'}_{q_1 = 0}, \underbrace{6x \cdot f_0 + 2 \cdot 3x^2 \cdot f' + x^3 \cdot f''}_{q_2 = 0}, \dots)$$

This can easily be solved by forward substitution:

$$\begin{aligned} 1 &= x^3 \cdot f_0 && \rightarrow f_0 = x^{-3} \\ 0 &= 3x^2 \cdot f_0 + x^3 \cdot f' && \rightarrow f' = -3x^{-4} \\ 0 &= 6x \cdot f_0 + 2 \cdot 3x^2 \cdot f' + x^3 \cdot f'' && \rightarrow f'' = 12x^{-5} \\ &\dots && \end{aligned}$$

$$\tag{3.158}$$

Using the same procedure for $f(x) = x^{-1}$ one obtains:

$$(x, 1, 0, 0, \dots)^{-1} = \left(\frac{1}{x}, -\frac{1}{x^2}, \frac{2}{x^3}, \dots\right) \tag{3.159}$$

For the function we have used before (3.145) :

$$f(x) = x^2 + \frac{1}{x} \tag{3.160}$$

and using (adding!) the expressions (3.156) and (3.159) one has:

$$(f_0, f', f'', f''') = \left(x^2 + \frac{1}{x}, 2x - \frac{1}{x^2}, 2 + \frac{2}{x^3}, \dots\right) \tag{3.161}$$

3.7.6.5 Automatic Differentiation: More Variables

It can be extended to more variables x, y and a function $f(x, y)$:

$$x = (a, 1, 0, 0, 0 \dots) \tag{3.162}$$

$$y = (b, 0, 1, 0, 0 \dots) \tag{3.163}$$

and get (with more complicated multiplication rules):

$$f((x + dx), (y + dy)) = \left(f, \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial^2 f}{\partial x^2}, \frac{\partial^2 f}{\partial x \partial y}, \dots\right)(x, y) \tag{3.164}$$

3.7.6.6 Differential Algebra: Applications to Accelerators

Of course it is not the purpose of these tools to compute analytical expressions for the derivatives, the examples were used to demonstrate the techniques. The application of these techniques (i.e. Truncated Power Series Algebra [6, 21]) is schematically shown in Fig.3.9. Given an algorithm, which may be a complex simulation program with several thousands of lines of code, we can use the techniques to “teach” the code how to compute the derivatives automatically.



Fig. 3.9 Schematic view of application of Truncated Power Series Algebra

When we push $f(x) = (a, 1, 0, 0, 0 \dots)$ through the algorithm, using our rules, we get all derivatives around a , i.e. we get the Taylor coefficients and can construct the map!

What is needed is to replace the standard operations performed by the computer on real numbers by the algebra defined above. The maps are provided with the desired accuracy and to any order.

Given a Taylor series to high accuracy, the wanted information about stability of the system, global behaviour and optical parameters can be derived more easily. It should be stressed again that the origin is the underlying tracking code, just acting on different data types with different operations.

3.7.6.7 Differential Algebra: Simple Example

A simple example is shown below where the original “tracking code” is shown in the left column (DATEST1) and the corresponding modified code in the right column (DATEST2). The operation is rather trivial to demonstrate the procedure more easily. The code is written in standard FORTRAN 95 which allows operator overloading, but an object oriented language such as C++ or Python are obviously well suited for this purpose. Standard FORTRAN-95 is however more flexible overloading arbitrary operations. The DA-package used for demonstration only is loaded by the command *use myownda* in the code. To make the program perform the wanted operation we have to make two small modifications:

1. Replace the types *real* by the type *mytaylor* (defined in the package).
2. Add the “differential unit” (0, 1), the monomial in this implementation to the variable.

```
PROGRAM DATEST1
use my_own_da
real(8) x,z, dx
my_order=3
dx=0.0
x=3.141592653_8/6.0_8+dx
call track(x, z)
call print(z,6)
END PROGRAM DATEST1
```

```
SUBROUTINE TRACK(a, b)
use my_own_da
real(8) a,b
b = sin(a)
END SUBROUTINE TRACK
```

```
PROGRAM DATEST2
use my_own_da
type(my_taylor) x,z, dx
my_order=3
dx=1.0d0.mono.1 ! this is our (0,1)
x=3.141592653_8/6.0_8+dx
call track(x, z)
call print(z,6)
END PROGRAM DATEST2
```

```
SUBROUTINE TRACK(a, b)
use my_own_da
type(my_taylor) a,b
b = sin(a)
END SUBROUTINE TRACK
```

Running these two programs we get the results in the two columns below. In the left column we get the expected result from the real calculation of the expression

$\sin(\pi/6) = 0.5$, while in the right column we get additional numbers sorted according to the array index.

(0,0) 0.50000000E+00	(0,0) 0.50000000E+00
	(1,0) 0.86602540E+00
	(0,1) 0.00000000E+00
	(2,0) -0.25000000E+00
	(0,2) 0.00000000E+00
	(1,1) 0.00000000E+00
	(3,0) -0.14433756E+00
	(0,3) 0.00000000E+00
	(2,1) 0.00000000E+00
	(1,2) 0.00000000E+00

The inspection shows that these numbers are the coefficients of the Taylor expansion of $\sin(x)$ around $x = \pi/6$:

$$\sin\left(\frac{\pi}{6} + \Delta x\right) = \sin\left(\frac{\pi}{6}\right) + \cos\left(\frac{\pi}{6}\right)\Delta x - \frac{1}{2}\sin\left(\frac{\pi}{6}\right)\Delta x^2 - \frac{1}{6}\cos\left(\frac{\pi}{6}\right)\Delta x^3 \quad (3.165)$$

We have indeed obtained the derivatives of our “algorithm” through the tracking code.

Some examples related to the analysis of accelerator physics lattices.

In example 1 a lattice with 8 FODO cells is constructed and the quadrupole is implemented as a thin lens “kick” in the center of the element. Note that the example is implemented in the horizontal and the longitudinal planes. For the second example an octupole kick is added to demonstrate the correct computation of the non-linear effect, i.e. the detuning with amplitude.

The procedure is:

1. Track through the lattice and get Taylor coefficients
2. Produce a map from the coefficients
3. Perform a Normal Form Analysis on the map

program ex1

	do j = 1,8
use my_own_da	z(1) = z(1)+DL/2*z(2)
use my_analysis	z(2) = z(2)-kf*DL*z(1)/(1 + z(3))
type(my_taylor) z(3)	z(1) = z(1)+DL/2*z(2)
type(normalform) NORMAL	
type(my_map) M,id	z(1) =z(1)+LC*z(2)
real(dp) L,DL,k1,k3,fix(3)	z(1) = z(1)+DL/2*z(2)
	z(2) = z(2)-kd*DL*z(1)/(1 + z(3))

```

! set up initial parameters
my_order=4 ! maximum order
4
fix=0.0 ! fixed point
id=1
z=fix+id
    
```

```

! set up lattice parameters
LC=62.5 ! half cell length
DL=3.0 ! quadrupole length
kf= 0.00295278 ! strength
kd=-0.00295278 ! strength
    
```

(0,0,0) 0.9369211296691E-01
 (0,0,1) -0.9649503806747E-01

(1,0,0) 0.9083165810508E-01
 (0,1,0) 0.1667704101367E+03

(1,0,1) 0.1238115392391E+01
 (0,1,1) -0.3527698956093E+02
 (1,0,2) -0.1567062442887E+01
 (0,1,2) 0.3478356898518E+02
 (1,0,3) 0.1896009493384E+01
 (0,1,3) -0.3429014840944E+02

(1,0,0) -0.5139797664004E-02
 (0,1,0) 0.1572511594903E+01
 (1,0,1) 0.1027959532801E-01
 (0,1,1) -0.5648018984066E+00
 (1,0,2) -0.1541939299201E-01
 (0,1,2) 0.5570922019106E+00
 (1,0,3) 0.2055919065602E-01
 (0,1,3) -0.5493825054146E+01

```

program ex2
    
```

```

use my_own_da
    
```

```

z(1) = z(1)+DL/2*z(2)
    
```

```

z(1) =z(1)+LC*z(2)
enddo
    
```

```

call print(z(1),6)
call print(z(2),6)
    
```

```

M=z ! overloads coefficient with the map
normal=m ! overloads map with normal
form
    
```

```

write(6,*) normal%tune, normal%dtune_da
end program ex1
    
```

From the elements in the Taylor expansion, the result for the matrix per cell:

$$\Delta x_f = 0.09083\Delta x_i + 166.77\Delta p_i$$

$$\Delta p_f = -0.00514\Delta x_i + 1.5725\Delta p_i$$

The output from the normal form analysis are (per cell!):

Tune = (0,0,0) = 0.093692

Chromaticity = (0,0,1) = -0.096495

```

do j = 1,8
    
```

```

z(1) = z(1)+DL/2*z(2)
    
```

```

z(2) = z(2)-kf*DL*z(1)/(1 + z(3))
    
```

```

z(1) = z(1)+DL/2*z(2)
    
```

```

use my_analysis
type(my_taylor) z(3)
type(normalform) NORMAL
type(my_map) M,id

real(dp) L,DL,k1,k3,fix(3)

! set up initial parameters
my_order=4 ! maximum order 4
fix=0.0 ! fixed point
id=1
z=fix+id
    
```

```

! set up lattice parameters
LC=62.5 ! half cell length
DL=3.0 ! quadrupole length
kf= 0.00295278 ! strength
kd=-0.00295278 ! strength
    
```

(0,0,0) 0.9369211296691E-01
(0,0,1) -0.9649503806747E-01

(2,0,0) 0.5383744464902E+02
(0,2,0) 0.5383744464902E+02
(0,0,2) 0.1009289258270E+00
(2,0,1) 0.2116575633218E+02
.....
(1,0,0) 0.9083165810508E-01
(0,1,0) 0.1667704101367E+03
(1,0,1) 0.1238115392391E+01
(0,1,1)-0.3527698956093E+02
(3,0,0)-0.1578216232118E+01
(2,1,0)-0.1429958442579E+02
(1,2,0)-0.4318760015031E+02

.....
(1,0,0)-0.5139797664004E-02
(0,1,0) 0.1572511594903E+01
(1,0,1) 0.1027959532801E-01
(0,1,1)-0.5648018984066E+00
(3,0,0)-0.1505298087837E-01

```

z(2) =z(2)*k3*z(1)**3/1+z(3) ! add
octupole kick
z(1) =z(1)+LC*z(2)

z(1) = z(1)+DL/2*z(2)
z(2) = z(2)-kd*DL*z(1)/(1+z(3))
z(1) = z(1)+DL/2*z(2)

z(1) =z(1)+LC*z(2)
enddo
call print(z(1),6)
call print(z(2),6)
    
```

```

M=z ! overloads coefficient with the map
normal=m ! overloads map with normal
form
    
```

```

write(6,*) normal%tune, normal%dtune_da
end program ex2
    
```

From the elements in the Taylor expansion, the result for the matrix per cell:

$$\Delta x_f = 0.09083\Delta x_i + 166.77\Delta p_i$$

$$\Delta p_f = -0.00514\Delta x_i + 1.5725\Delta p_i$$

The output from the normal form analysis are (per cell!):

Tune = (0,0,0) = 0.093692
Chromaticity = (0,0,1) = -0.096495

The added octupole kick results in a detuning with amplitude of $dQ/dJ = 53.837$

Defined assignments:

$M = z,$ constructs a map M using the coefficients z
 $NORMAL = M,$ computes a normal form $NORMAL$ using the map M

In FORTRAN95 derived “type” plays the role of “structures” in C, and $NORMAL$ contains:

$NORMAL\%tune$ is the tune Q
 $NORMAL\%dtune_da$ is the detuning with amplitude $\frac{dQ}{da}$

$NORMAL\%R,$ $NORMAL\%A,$ $NORMAL\%A^{**}-1$ are the matrices such that: $M = A R A^{-1}$

from the normal form transformation one obtains $\alpha, \beta, \gamma \dots$

Below a comparison is shown in Fig. 3.10 using the lattice function (e.g. β) obtained with this procedure and the optical functions from the corresponding MAD-X output.

One finds that $\beta_{max} \approx 300 m,$ $\beta_{min} \approx 170 m$ and perfect agreement.

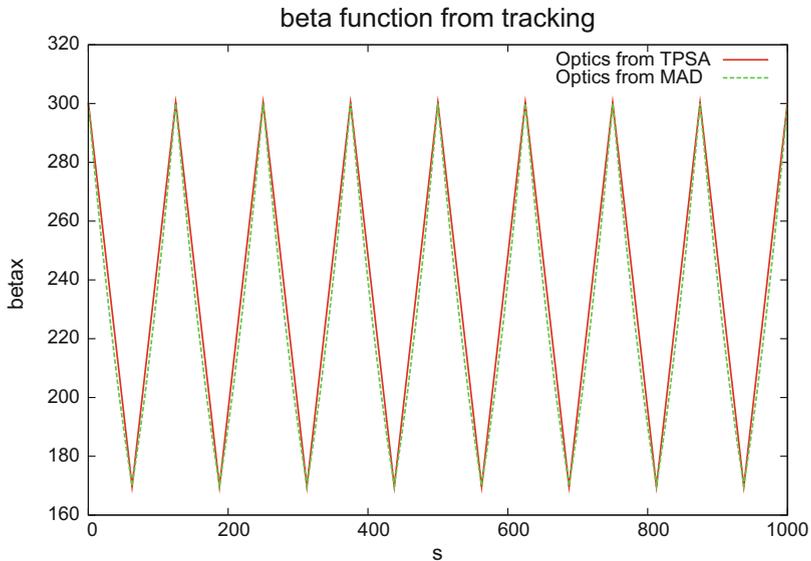


Fig. 3.10 Comparison: β -function from the model and the corresponding result from MAD-X

3.8 Beam Dynamics with Non-linearities

Following the overview of the evaluation and analysis tools, it is now possible to analyse and classify the behaviour of particles in the presence of non-linearities. The tools presented beforehand allow a better physical insight to the mechanisms leading to the various phenomena, the most important ones being:

- Amplitude detuning
- Excitation of non-linear resonances
- Reduction of dynamic aperture and chaotic behaviour.

This list is necessarily incomplete but will serve to demonstrate the most important aspects.

To demonstrate these aspects, we take a realistic case and show how the effects emerge automatically.

3.8.1 Amplitude Detuning

It was discussed in Sect. 3.7.5 that the one-turn-map can be transformed into a simpler map where the rotation is separated. A consequence of the non-linearities was that the rotation frequency becomes amplitude dependent to perform this transformation. Therefore the amplitude detuning is directly obtained from this normal form transformation.

3.8.1.1 Amplitude Detuning due to Non-linearities in Machine Elements

Non-linear elements cause an amplitude dependent phase advance. The computational procedure to derive this detuning was demonstrated in the discussion on normal form transformations in the case of an octupole Eqs. (3.125) and (3.129). This formalism is valid for any non-linear element.

Numerous other examples can be found in [6] and [5].

3.8.1.2 Amplitude Detuning due to Beam–Beam Effects

For the demonstration we use the example of a beam-beam interaction because it is a very complex non-linear problem and of large practical importance [7, 22].

In this simplest case of one beam-beam interaction we can factorize the machine in a linear transfer map $e^{i f_2}$ and the beam-beam interaction $e^{i F}$, i.e.:

$$e^{i f_2} \cdot e^{i F} = e^{i h} \quad (3.166)$$

with

$$f_2 = -\frac{\mu}{2}\left(\frac{x^2}{\beta} + \beta p_x^2\right) \quad (3.167)$$

where μ is the overall phase, i.e. the tune Q multiplied by 2π , and β is the β -function at the interaction point. We assume the waist of the β -function at the collision point ($\alpha = 0$). The function $F(x)$ corresponds to the beam-beam potential (3.87):

$$F(x) = \int_0^x f(u)du \quad (3.168)$$

For a round Gaussian beam we use for $f(x)$ the well known expression:

$$f(x) = \frac{2Nr_0}{\gamma x} \left(1 - e^{-\frac{x^2}{2\sigma^2}}\right) \quad (3.169)$$

Here N is the number of particles per bunch, r_0 the classical particle radius, γ the relativistic parameter and σ the transverse beam size.

For the analysis we examine the invariant h which determines the one-turn-map (OTM) written as a Lie transformation $e^{:h:}$. The invariant h is the effective Hamiltonian for this problem.

As usual we transform to action and angle variables J and Φ , related to the variables x and p_x through the transformations:

$$x = \sqrt{2J\beta}\sin\Phi, \quad p_x = \sqrt{\frac{2J}{\beta}}\cos\Phi \quad (3.170)$$

With this transformation we get a simple representation for the linear transfer map f_2 :

$$f_2 = -\mu J \quad (3.171)$$

The function $F(x)$ we write as Fourier series:

$$F(x) \Rightarrow \sum_{n=-\infty}^{\infty} c_n(J)e^{in\Phi} \quad \text{with} \quad c_n(J) = \frac{1}{2\pi} \int_0^{2\pi} e^{-in\Phi} F(x)d\Phi \quad (3.172)$$

For the evaluation of (3.172) see [7]. We take some useful properties of Lie operators (e.g. [6, 7]):

$$: f_2 : g(J) = 0, \quad : f_2 : e^{in\Phi} = in\mu e^{in\Phi}, \quad g(: f_2 :)e^{in\Phi} = g(in\mu)e^{in\Phi} \quad (3.173)$$

and the CBH-formula for the concatenation of the maps (3.92):

$$e \circ f_2 \circ e \circ F \circ e = e \circ h \circ e = \exp \left[: f_2 + \left(\frac{: f_2 :}{1 - e^{-: f_2 :}} \right) F + O(F^2) : \right] \quad (3.174)$$

which gives immediately for h :

$$h = -\mu J + \sum_n c_n(J) \frac{in\mu}{1 - e^{-in\mu}} e^{in\Phi} = -\mu J + \sum_n c_n(J) \frac{n\mu}{2 \sin(\frac{n\mu}{2})} e^{(in\Phi + i\frac{n\mu}{2})} \quad (3.175)$$

Equation (3.175) is the beam-beam perturbed invariant to first order in the perturbation using (3.92).

From (3.175) we observe that for $\nu = \frac{\mu}{2\pi} = \frac{p}{n}$ resonances appear for all integers p and n when $c_n(J) \neq 0$.

Away from resonances a normal form transformation gives:

$$h = -\mu J + c_0(J) = \text{const.} \quad (3.176)$$

and the oscillating term disappears. The first term is the linear rotation and the second term gives the amplitude dependent tune shift (see (3.114)):

$$\Delta\mu(J) = -\frac{1}{2\pi} \frac{dc_0(J)}{dJ} \quad (3.177)$$

The computation of this tuneshift from the equation above can be found in the literature [7, 23].

3.8.1.3 Phase Space Structure

To demonstrate how this technique can be used to reconstruct the phase space structure in the presence of non-linearities, we continue with the very non-linear problem of the beam-beam interaction treated above. To test our result, we compare the invariant h to the results of a particle tracking program.

The model we use in the program is rather simple:

- linear transfer between interactions
- beam-beam kick for round beams
- compute action $J = \frac{\beta^*}{2\sigma_x} \left(\frac{x^2}{\beta^*} + p_x^2 \beta^* \right)$
- compute phase $\Phi = \arctan\left(\frac{p_x}{x}\right)$
- compare J with h as a function of the phase Φ

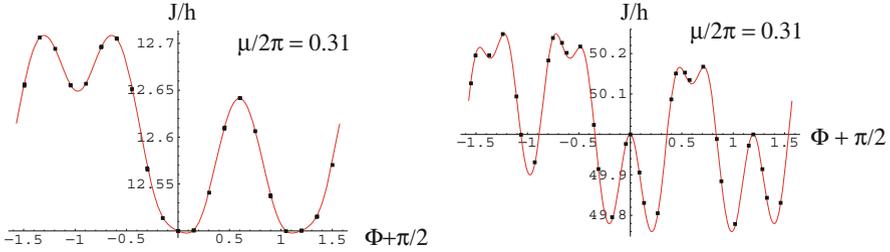


Fig. 3.11 Comparison: numerical and analytical model for one interaction point. Shown for $5\sigma_x$ (left) and $10\sigma_x$ (right). Full symbols from numerical model and solid lines from invariant (3.175)

The evaluation of the invariant (3.175) is done numerically with Mathematica. The comparison between the tracking results and the invariant h from the analytical calculation is shown in Fig. 3.11 in the (J, Φ) space. One interaction point is used in this comparison and the particles are tracked for 1024 turns. The symbols are the results from the tracking and the solid lines are the invariants computed as above. The two figures are computed for amplitudes of 5σ and 10σ . The agreement between the models is excellent. The analytic calculation was done up to the order $N = 40$. Using a lower number, the analytic model can reproduce the envelope of the tracking results, but not the details. The results can easily be generalized to more interaction points [22]. Close to resonances these tools can reproduce the envelope of the phase space structure [22].

3.8.2 Non-linear Resonances

Non-linear resonances can be excited in the presence of non-linear fields and play a vital role for the long term stability of the particles.

3.8.2.1 Resonance Condition in One Dimension

For the special case of the beam-beam perturbed invariant (3.175) we have seen that the expansion (3.175) diverges when the resonance condition for the phase advance is fulfilled, i.e.:

$$\nu = \frac{\mu}{2\pi} = \frac{p}{n} \quad (3.178)$$

The formal treatment would imply to use the n -turn map with the n -turn effective Hamiltonian or other techniques. This is beyond the scope of this handbook and can be found in the literature [6, 7]. We should like to discuss the consequences of resonant behaviour and possible applications in this section.

3.8.2.2 Driving Terms

The treatment of the resonance map is still not fully understood and a standard treatment using first order perturbation theory leads to a few wrong conclusions. In particular it is believed that a resonance cannot be excited unless a driving term for the resonance is explicitly present in the Hamiltonian. This implies that the related map must contain the term for a resonance in leading order to reproduce the resonance. This regularly leads to the conclusion that 3rd order resonances are driven by sextupoles, 4th order are driven by octupoles etc. This is only a consequence of the perturbation theory which is often not carried beyond leading order, and e.g. a sextupole can potentially drive resonances of any order. Such a treatment is valid only for special operational conditions such as resonant extraction where strong resonant effects can be well described by a perturbation theory. A detailed discussion of this misconception is given in [6]. A correct evaluation must be carried out to the necessary orders and the tools presented here allow such a treatment in an easier way.

3.8.3 Chromaticity and Chromaticity Correction

For reasons explained earlier, sextupoles are required to correct the chromaticities. In large machines and in particular in colliders with insertions, these sextupoles dominate over the non-linear effects of so-called linear elements.

3.8.4 Dynamic Aperture

Often in the context of the discussion of non-linear resonance phenomena the concept of *dynamic aperture* is introduced. This is the maximum stable oscillation amplitude in the transverse (x, y) -space due to non-linear fields. It must be distinguished from the physical aperture of the vacuum chamber or other physical restrictions such as collimators.

One of the most important tasks in the analysis of non-linear effects is to provide answers to the questions:

- Determination of the dynamic aperture
- Maximising the dynamic aperture

The computation of the dynamic aperture is a very difficult task since no mathematical methods are available to calculate it analytically except for the trivial cases. Following the concepts described earlier, the theory is much more complete from the simulation point of view. Therefore the standard approach to compute the dynamic aperture is done by numerical tracking of particles.

The same techniques can be employed to maximise the dynamic aperture, in the ideal case beyond the limits of the physical aperture. Usually one can define tolerances for the allowed multipole components of the magnets or the optimized parameters for colliding beams when the dominant non-linear effect comes from beam-beam interactions.

3.8.4.1 Long Term Stability and Chaotic Behaviour

In accelerators such as particle colliders, the beams have to remain stable for many hours and we may be asked to answer the question about stability for as many as 10^9 turns in the machine. This important question cannot be answered by perturbative techniques. In the discussion of Poincare surface-of-section we have tasted the complexity of the phase space topology and the final question is whether particles eventually reach the entire region of the available phase space.

It was proven by Kolmogorov, Arnol'd and Moser (KAM theorem) that for weakly perturbed systems invariant surfaces exist in the neighbourhood of integrable ones. Poincare gave a first hint that stochastic behaviour may be generated in non-linear systems. In fact, higher order resonances change the topology of the phase space and lead to the formation of island chains on an increasingly fine scale. Satisfactory insight to the fine structure of the phase space can only be gained with numerical computation. Although the motion near resonances may be stochastic, the trajectories are constrained by nearby KAM surfaces (at least in one degree of freedom) and the motion remains confined.

3.8.4.2 Practical Implications

In numerical simulations where particles are tracked for millions of turns we would like to determine the region of stability, i.e. dynamic aperture. Since we cannot track ad infinitum, we have to specify criteria whether a particle is stable or not. A straightforward method is to test the particle amplitudes against well defined apertures and declare a particle lost when the aperture is reached. A sufficient number of turns, usually determined by careful testing, is required with this method.

Usually this means to find the particle survival time as a function of the initial amplitude. In general the survival time decreases as the amplitude increases and should reach an asymptotic value at some amplitude. The latter can be identified as the dynamic aperture.

Other methods rely on the assumption that a particle that is unstable in the long term, exhibits features such as a certain amount of chaotic motion.

Typical methods to detect and quantify chaotic motion are:

- Frequency Map Analysis [24, 25].
- Lyapunov exponent [26].
- Chirikov criterion [27].

In all cases care must be taken to avoid numerical problems due to the computation techniques when a simulation over many turns is performed.

References

1. W. Herr, **Mathematical and Numerical Methods for Nonlinear Dynamics** Proc. CERN Accelerator School: **Advanced Accelerator Physics** (2013), published as CERN Yellow Report CERN-2014-009, arXiv:1601.07311.
2. E. Courant and H. Snyder, **Theory of the Alternating Gradient Synchrotron**, Ann. Phys. 3 (1958) 1.
3. A. Wolski, **Beam Dynamics in High Energy Particle Accelerators**, Imperial College Press (2014).
4. H. Wiedemann, **Particle Accelerator Physics—basic Principles and Linear Beam Dynamics**, Springer-Verlag (1993).
5. A. Chao and M. Tigner, **Handbook of Accelerator Physics and Engineering**, World Scientific (1998).
6. E. Forest, **Beam Dynamics**, Harwood Academic Publishers (1998).
7. A. Chao, **Lecture Notes on Topics in Accelerator Physics**, SLAC (2001).
8. A. Dragt, **Lie Methods for Nonlinear Dynamics with Applications to Accelerator Physics**, in preparation, Univ. of Maryland (Nov. 2019). <https://www.physics.umd.edu/dsat/dsatliemethods.html>.
9. E. Courant and R. Ruth, **Stability in Dynamical Systems**, Summer School on High Energy Particle Accelerators, Upton, New York, July 6–16, 1983.
10. A. Dragt and E. Forest, **Computation of Nonlinear Behaviour of Hamiltonian Systems using Lie Algebraic Methods**, J.Math.Phys. **24**, 2734 (1983).
11. E. Forest and R. Ruth, Physica D43, (1990) 105.
12. H. Yoshida, Phys. Lett A150 (1990) 262.
13. W. Herr, **Relativity**, lecture at CAS-CERN Accelerator School on “Introduction to accelerator Physics”, Budapest, Hungary (2016).
14. S. Sheehy, **Motion of Particles in Electro-magnetic Fields**, lecture at CAS-CERN Accelerator School on “Introduction to accelerator Physics”, Budapest, Hungary (2016).
15. A. Dragt, AIP Proc. 87, Phys. High Energy Accelerators, Fermilab, (1981) 147.
16. A. Dragt et al., Ann. Rev. Nucl. Part. Sci 38 (1988) 455.
17. A. Dragt et al, Phys. Rev. A45, (1992) 2572.
18. H. Goldstein, **Classical Mechanics**, Addison Wesley, (2001).
19. A. Dragt and J. Finn, J. Math. Phys., **17**, 2215 (1976); A. Dragt et al., Ann. Rev. Nucl. Part. Sci., **38**, 455 (1988).
20. H. Poincare, **Les Methodes Nouvelles de la Mecanique Celeste**, Gauthier-Villars, Paris (1892).
21. M. Berz, Particle Accelerators 24, (1989) 109.
22. W. Herr, D. Kaltchev, **Effect of phase advance between interaction points in the LHC on the beam-beam interaction**, LHC Project Report 1082, unpublished, (2008).
23. T. Pieloni, **A Study of Beam-Beam Effects in Hadron Colliders with a Large Number of Bunches**, PhD thesis Nr. 4211, EPFL Lausanne, (2008).
24. H.S. Dumas, J. Laskar, Phys. Rev. Lett. 70 (1989) 2975.
25. J. Laskar, D. Robin, Particle Accelerators 54 (1996)183.
26. G. Benettin et al., Phys. Rev. A14 (1976) 2338.
27. B.V. Chirikov, At. Energ. 6, (1959) 630.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 4

Impedance and Collective Effects



E. Metral, G. Rumolo, and W. Herr

As the beam intensity increases, the beam can no longer be considered as a collection of non-interacting single particles: in addition to the “single-particle phenomena”, “collective effects” become significant. At low intensity a beam of charged particles moves around an accelerator under the Lorentz force produced by the “external” electromagnetic fields (from the guiding and focusing magnets, RF cavities, etc.). However, the charged particles also interact with themselves (leading to space charge effects) and with their environment, inducing charges and currents in the surrounding structures, which create electromagnetic fields called wake fields. In the ultra-relativistic limit, causality dictates that there can be no electromagnetic field in front of the beam, which explains the term “wake”. It is often useful to examine the frequency content of the wake field (a time domain quantity) by performing a Fourier transformation on it. This leads to the concept of impedance (a frequency domain quantity), which is a complex function of frequency. The charged particles can also interact with other charged particles present in the accelerator (leading to two-stream effects, and in particular to electron cloud effects in positron/hadron machines) and with the counter-rotating beam in a collider (leading to beam–beam effects). As the beam intensity increases, all these “perturbations” should be properly quantified and the motion of the charged particles will eventually still be governed by the Lorentz force but using the total electromagnetic fields, which are the sum of the external and perturbation fields. Note that in some cases a perturbative treatment is not sufficient and the problem has to be solved self consistently. These perturbations can lead to both incoherent (i.e. of

Coordinated by E. Metral

E. Metral (✉) · G. Rumolo · W. Herr
CERN (European Organization for Nuclear Research), Geneva, Switzerland
e-mail: Elias.Metral@cern.ch; Giovanni.Rumolo@cern.ch; Werner.Herr@cern.ch

© The Author(s) 2020
S. Myers, H. Schopper (eds.), *Particle Physics Reference Library*,
https://doi.org/10.1007/978-3-030-34245-6_4

a single particle) and coherent (i.e. of the centre of mass) effects, in the longitudinal and in one or both transverse directions, leading to beam quality degradation or even partial or total beam losses. Fortunately, stabilising mechanisms exist, such as Landau damping, electronic feedback systems and linear coupling between the transverse planes (as in the case of a transverse coherent instability, one plane is usually more critical than the other).

Beam instabilities cover a wide range of effects in particle accelerators and they have been the subjects of intense research for several decades. As the machines performance was pushed new mechanisms were revealed and nowadays the challenge consists in studying the interplays between all these intricate phenomena, as it is very often not possible to treat the different effects separately [1, 2]. This field is still very active as can be revealed by the recent (and future) international workshops devoted to this subject [3–5].

This chapter is structured as follows: space charge is discussed in Sect. 4.1, wake fields (and related impedances) in Sect. 4.2, the induced coherent instabilities in Sect. 4.3 and the Landau damping mechanism in Sect. 4.4. The two-stream effects are analyzed in Sect. 4.5, concentrating mainly on electron cloud, while beam–beam effects are reviewed in Sect. 4.6, before concluding in Sect. 4.7 by the numerical modelling of collective effects.

4.1 Space Charge

E. Metral

4.1.1 Direct Space Charge

Two space charge effects are distinguished: the direct space charge and the indirect (or image) one ([6–9], and references therein). The direct space charge comes from the interaction between the particles of a single beam, without interaction with the surrounding vacuum chamber. Consider two particles with the same charge (for instance protons) in vacuum. They will feel two forces: the Coulomb repulsion (as they have the same charge) and the magnetic attraction (as they represent currents moving in the same direction, leading to an azimuthal magnetic field). Let's assume that particle 1 is moving with speed $v_1 = \beta_1 c$ with respect to the laboratory frame, with β the relativistic velocity factor and c the speed of light. In its rest frame, particle 1 produces only an electrostatic field, which can be computed, and applying the relativistic transformation of the electromagnetic fields between the rest and laboratory frames, the magnetic contribution can be obtained. Note that there is no magnetic contribution in the longitudinal plane ($B_s = 0$), which leads to the longitudinal Lorentz force $F_s = eE_s$, where e is the elementary charge, s the azimuthal coordinate and E_s the longitudinal electric field. The transverse

(horizontal and vertical) Lorentz force on particle 2, moving with speed $v_2 = \beta_2 c$ with respect to the laboratory frame, is written

$$F_{x,y} = eE_{x,y} \begin{cases} (1 - \beta_1\beta_2), & \text{if 2 moves in same direction as 1} \\ (1 + \beta_1\beta_2), & \text{if 2 moves in opposite direction as 1} \end{cases} \quad (4.1)$$

The first case corresponds to the space charge case where both particles move in the same direction, while the second corresponds to the beam–beam case (see Sect. 4.6) where the particles move in opposite direction. In both cases, the first term comes from the electric field while the second comes from the magnetic one. The main difference between the two regimes is that for the space charge case there is a partial compensation of the two forces, while for the beam–beam case the two forces add. The space charge force is maximum at low energy and vanishes at high energy, while the beam–beam force is maximum at high energy. Considering the space charge regime and assuming the same speed for both beams, the Lorentz force simplifies to

$$F_{x,y} = eE_{x,y} (1 - \beta^2) = e \frac{E_{x,y}}{\gamma^2}, \quad F_s = eE_s, \quad (4.2)$$

where γ is the relativistic mass factor. Assuming a circular beam pipe with radius b (which is important only for the computation of the longitudinal force) and applying Gauss's law, the electromagnetic fields can be computed for a bunch with Gaussian radial density (with rms $\sigma_x = \sigma_y = \sigma$) using the cylindrical coordinates (r, θ, s) . The associated Lorentz forces are given by

$$F_r = \frac{e}{\gamma^2} E_r = \frac{e\lambda(z)}{2\pi\epsilon_0\gamma^2} \left(\frac{1 - e^{-\frac{r^2}{2\sigma^2}}}{r} \right), \quad F_s = -\frac{e}{2\pi\epsilon_0\gamma^2} \frac{d\lambda(z)}{dz} \int_{r'=r}^b \frac{1 - e^{-\frac{r'^2}{2\sigma^2}}}{r'} dr', \quad (4.3)$$

where $\lambda(z)$ is the longitudinal line density, $z = s - vt$ with t being the time, and ϵ_0 the vacuum permittivity. A first observation is that the space charge forces are highly nonlinear. Another important observation is that the radial force is proportional to the longitudinal density while the longitudinal one is proportional to the derivative of the longitudinal density. Linearizing both forces (for very small amplitudes where $r \ll \sigma$) leads to

$$F_r \approx \frac{e\lambda(z)}{2\pi\epsilon_0\gamma^2} \frac{r}{2\sigma^2}, \quad F_s \approx -\frac{e}{4\pi\epsilon_0\gamma^2} \frac{d\lambda(z)}{dz} \left[1 + 2 \ln \left(\frac{b}{\sqrt{2}\sigma} \right) \right]. \quad (4.4)$$

This means that the transverse space charge force is linear for small amplitudes and defocusing. Due to the additional space charge force, e.g. the horizontal betatron tune will no longer be the unperturbed tune Q_{x0} but will be $Q_x = Q_{x0} + \Delta Q_x$, where ΔQ_x is the horizontal incoherent betatron tune shift. Similarly, the new synchrotron

tune will be $Q_s = Q_{s0} + \Delta Q_s$, where ΔQ_s is the incoherent synchrotron tune shift. The betatron and synchrotron linearized incoherent space charge tune shifts are given by

$$\Delta Q_x^{\text{Lin}} = -\frac{N_b r_p}{4\pi\beta\gamma^2\varepsilon_{x,\text{rms}}^{\text{norm}}B}, \quad \Delta Q_s^{\text{Lin}} = +\frac{\eta N_b e^2 R^2}{8\pi\sqrt{2\pi}\varepsilon_0 E_{\text{total}}\beta^2\gamma^2\sigma_z^3 Q_{s0}} \left[1 + 2 \ln\left(\frac{b}{\sqrt{2}\sigma}\right) \right], \quad (4.5)$$

where N_b is the number of protons in the bunch, r_p the classical proton radius, $\varepsilon_{x,\text{rms}}^{\text{norm}} = \beta\gamma\varepsilon_{x,\text{rms}}$ the normalized rms horizontal emittance, with $\varepsilon_{x,\text{rms}} = \sigma^2/\beta_x$ at a place of zero dispersion with β_x the horizontal betatron function, $B = \sqrt{2\pi}\sigma_z/2\pi R$ is the bunching factor (assuming a Gaussian longitudinal distribution with rms σ_z) with R the average machine radius, $\eta = \gamma_{\text{tr}}^{-2} - \gamma^{-2}$ the slip factor (where γ_{tr} stands for γ at transition energy) and E_{total} is the total particles' energy. It is shown from Eq. (4.5) that the transverse betatron tune shift is always negative, revealing that the space charge force is always defocusing. Note that for the case of an ion with mass number A and charge state Z , the transverse tune shift has to be multiplied by Z^2/A .

Contrary to the transverse case, the longitudinal space charge is defocusing below transition (as $\eta < 0$) and focusing above (as $\eta > 0$). One can therefore already anticipate some longitudinal mismatch issues when crossing transition with high-intensity bunches, i.e. the bunch length will not be in equilibrium anymore and will oscillate inside the RF buckets. Such a case is depicted in Fig. 4.1(left) for the particular case of the high-intensity bunch in the CERN PS machine, which is sent to the nTOF (neutron Time-Of-Flight) experiment [10]. The computed quadrupolar oscillation is induced when transition is crossed and is a consequence of the longitudinal mismatch: below transition space charge is defocusing which reduces the bucket height and increases the bunch length, while above transition space charge is focusing which increases the bucket height and decreases the bunch length.

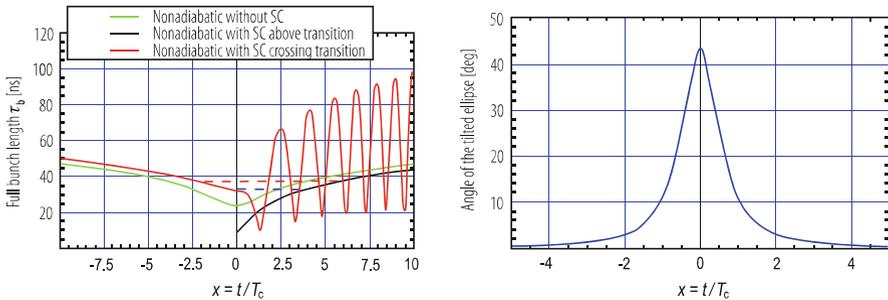


Fig. 4.1 (Left) Computation of the evolution of the full bunch length vs. time for the case of the nonadiabatic theory (the adiabatic theory is not valid anymore close to transition) with and without space charge, applied to the PS nTOF bunch [10]. T_c is the nonadiabatic time, equal to ~ 2 ms in the present case. (Right) Evolution of the angle of the tilted ellipse around transition (without space charge)

Therefore, there is an intensity-dependent step in the equilibrium bunch length at transition, which leads to a longitudinal mismatch and subsequent quadrupolar oscillations. If these bunch shape oscillations are not damped they will eventually result in filamentation and longitudinal emittance blow-up. It's worth mentioning that in presence of significant space charge, the minimum of bunch length is not reached right at transition anymore, but after about one nonadiabatic time T_c , i.e. after ~ 2 ms in the present case [11, 12]. The same kind of mechanism appears with the inductive part of the longitudinal machine impedance (see Sect. 4.2). The only difference is that in this case, the equilibrium bunch length is shorter below transition and longer above transition.

If transition crossing cannot be avoided, the γ_t jump is the only (known) method to overcome all the intensity limitations. It consists in an artificial increase of the transition crossing speed by means of fast pulsed quadrupoles. The idea is that quadrupoles at nonzero dispersion locations can be used to adjust the momentum compaction factor. The change in momentum compaction (called γ_t jump) depends on the unperturbed and perturbed dispersion functions at the kick-quadrupole locations. These schemes were pioneered by the CERN PS group [13–16]. Such a γ_t jump scheme makes it possible to keep the beam at a safe distance from transition, except for the very short time during which the transition region is crossed at a speed increased by one or two orders of magnitude. Looking at Fig. 4.1(left) clearly reveals why an asymmetric jump was proposed in the past [14] to damp the longitudinal quadrupolar oscillations arising from the space charge induced mismatch: the idea is to jump rapidly from an equilibrium bunch length below transition to the same value above. The amplitude of the jump is defined by the time needed to go to the same equilibrium bunch length above transition. The minimum amplitude of the jump corresponds to the case represented with the dashed blue line starting right at transition. However, in this case the initial longitudinal phase space ellipse is tilted (see Fig. 4.1(right)), while the final one is almost not, which is not ideal. One might want therefore to start the jump earlier, when the longitudinal phase space is almost not tilted, for instance at $x \approx -2$, which requires a larger jump (see the dashed orange line in Fig. 4.1(left)).

Coming back to the transverse space charge, in the case of an elliptical beam (instead of a round one), one has to replace $2\sigma_x^2$ by $\sigma_x(\sigma_x + \sigma_y)$ and $2\sigma_y^2$ by $\sigma_y(\sigma_x + \sigma_y)$. Furthermore, due to the nonlinear nature of the space charge, the tune shifts of the different particles will not be the same, which will lead to a tune spread: plotted in the tune diagram it is called a tune footprint. The latter has to be accommodated in the tune diagram, without crossing harmful resonance lines, which might lead to emittance growth and/or beam losses. The exact tune footprint depends on the distribution and to illustrate this effect we consider in the following a round beam with quasi-parabolic distribution function, whose particle density extends up to $\sim 3.2\sigma$ [17, 18]. The corresponding horizontal 2D (i.e. neglecting the longitudinal distribution) space charge force is plotted in Fig. 4.2(left), and the tune footprint in Fig. 4.2(right). The unperturbed (low-intensity) working point is in the top right corner, the small-amplitude particles have the largest tune shifts while the largest amplitude particles have the smallest tune shifts. If the longitudinal

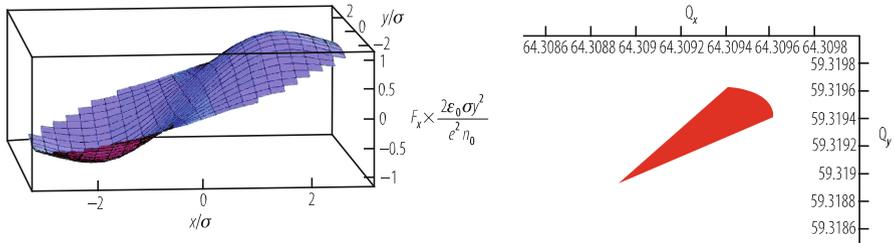


Fig. 4.2 Horizontal 2D (i.e. neglecting the longitudinal distribution) space charge force (left) and tune footprint for the case of the CERN LHC at injection, assuming the tunes in collision (64.31, 59.32). The parameter n_0 is the constant term in the particle density [18]

distribution is taken into account, the longitudinal variation (due to synchrotron oscillations) of the transverse space-charge force fills the gap until the low-intensity working point. However, it is interesting to plot it like this to clearly see the region occupied by the large synchrotron amplitude particles, because the interaction with a nonlinear resonance will depend on the overlapping position. Several possibilities exist with core emittance blow-up, creation of tails and/or beam losses. In particular, if the resonance interacts with the small amplitude particles, there could be a regime of loss-free (core-)emittance blow-up, while if the resonance interacts with the particles with large synchrotron amplitudes (i.e. if the resonance line is in the gap between the 2D tune footprint and the low-intensity working point) there could be a regime with continuous loss due to the trapping–detrapping mechanisms, as observed both in the PS [19] and at SIS18 [20].

Finally, another space charge mechanism, which could be important in high-intensity synchrotrons with unsplit transverse tunes (i.e. having the same integer) is the Montague resonance which can lead to emittance transfer from one plane to the other and might lead to losses if the beam fills the aperture [21, 22].

4.1.2 Indirect Space Charge

In the case of a beam off-axis in a perfectly conducting circular beam pipe (with radius b), a coherent (or dipolar, i.e. of the centre of mass) force arises, which can be found by using the method of the images (to satisfy the boundary condition on a perfect conductor, i.e. of a vanishing tangential electrical field). The electric field is always assumed to be non-penetrating. However, for the magnetic field, the situation is more complicated as it may or may not penetrate the vacuum chamber: the high-frequency components, called “ac” will not penetrate, while the low-frequency ones, called “dc” will penetrate and form images on the magnet pole faces (if there are some; otherwise they will go to infinity and will not act back on the beam). In the case of a non-penetrating “ac” magnetic field, one finally obtains (keeping only the linear terms, i.e. $\bar{x} \ll b$ and $\bar{y} \ll b$, where \bar{x} and \bar{y} are the transverse displacements

of the centre of mass):

$$F_x = \Lambda_c \bar{x}, \quad F_y = \Lambda_c \bar{y}, \quad \Lambda_c = \frac{\lambda e}{2\pi \epsilon_0 \gamma^2 b^2}. \quad (4.6)$$

It can be seen that the transverse coherent space charge force of Eq. (4.6) is similar to the transverse incoherent space charge force of Eq. (4.4, left): $2\sigma^2$ has been replaced by b^2 .

The same analysis can be performed in the case of two infinite (horizontal) parallel plates spaced by $2h$ and the results are the following (assuming that the transverse beam sizes are much smaller than h , assuming only the “ac” magnetic part and keeping only the linear terms)

$$F_x = \Lambda_c \left(\frac{\pi^2}{24} \bar{x} - \frac{\pi^2}{24} x \right), \quad F_y = \Lambda_c \left(\frac{\pi^2}{12} \bar{y} + \frac{\pi^2}{24} y \right). \quad (4.7)$$

Therefore, compared to the circular case, the coherent force is smaller by $\pi^2/24 \approx 0.4$ in the horizontal plane and $\pi^2/12 \approx 0.8$ in the vertical one. Furthermore, there is a second incoherent (or quadrupolar, as it is linear with the particle position) term with opposite sign in both planes. The coefficients are linked to the Laslett coefficients usually used in the literature [23], and they are the same as the ones obtained by Yokoya [24] in the case of a resistive beam pipe under some assumptions (see Sect. 4.2). General formulae exist for the “real” tune shifts of coasting or bunched beams in pipes with different geometries, considering both the “ac” and “dc” magnetic parts and can be found for instance in Refs. [6, 7].

4.2 Wake Fields and Impedances

E. Metral

If the wall of the beam pipe is perfectly conducting and smooth, as it was the case in the previous section, a ring of negative charges is formed on the walls of the beam pipe where the electric field ends, and these induced charges travel at the same pace with the particles, creating the so-called “image” (or induced) current, which leads to real tune shifts. However, if the wall of the beam pipe is not perfectly conducting or contains discontinuities, the movement of the induced charges will be slowed down, thus leaving electromagnetic fields (which are proportional to the beam intensity) mainly behind: this is why these electromagnetic fields are called wake fields. The latter will create complex tune shifts leading to instabilities (see Sect. 4.3). What needs to be computed are the wake fields at the distance $z = s - vt$ behind the source particle (which is at position $s_{\text{source}} = vt$; with this convention, one has $z < 0$) and their effects on the test or witness particles that compose the

beam. The computation of these wake fields is quite involved and two fundamental approximations are introduced:

1. *The rigid-beam approximation:* The beam traverses a piece of equipment rigidly, i.e. the wake field perturbation does not affect the motion of the beam during the traversal of the impedance. The distance z of the test particle behind some source particle does not change.
2. *The impulse approximation:* As the test particle moves at the fixed velocity $v = \beta c$ through a piece of equipment, the important quantity is the impulse (and not the force) given by

$$\Delta \mathbf{p}(x, y, z) = \int_{-\infty}^{+\infty} dt \mathbf{F}(x, y, s = z + \beta ct, t) = \int_{-\infty}^{+\infty} dt e(\mathbf{E} + \mathbf{v} \times \mathbf{B}), \quad (4.8)$$

where vectors are designated by boldtype letters. Starting from the four Maxwell equations for a particle in the beam, it can be shown that for a constant β (which does not need to be 1) [9]

$$\nabla \times \Delta \mathbf{p}(x, y, z) = 0, \quad (4.9)$$

which is known as Panofsky-Wenzel theorem. This relation is very general, as no boundary conditions have been imposed. Only the two fundamental approximations have been made. Another important relation can be obtained when $\beta = 1$, taking the divergence of the impulse, which is

$$\nabla_{\perp} \cdot \Delta \mathbf{p}_{\perp} = 0. \quad (4.10)$$

Considering the case of a cylindrically symmetric chamber (using the cylindrical coordinates r, θ, z), yields the following three equations from Panofsky-Wenzel theorem

$$\frac{1}{r} \frac{\partial \Delta p_z}{\partial \theta} = \frac{\partial \Delta p_{\theta}}{\partial z}, \quad \frac{\partial \Delta p_r}{\partial z} = \frac{\partial \Delta p_z}{\partial r}, \quad \frac{\partial (r \Delta p_{\theta})}{\partial r} = \frac{\partial \Delta p_r}{\partial \theta}. \quad (4.11)$$

The fourth relation when $\beta = 1$ writes

$$\frac{\partial (r \Delta p_r)}{\partial r} = -\frac{\partial \Delta p_{\theta}}{\partial \theta}. \quad (4.12)$$

Consider now as a source charge density a macro-particle of charge $Q = N_b e$ moving along the pipe (in the s -direction) with an offset $r = a$ in the $\vartheta = 0$ direction

and with velocity $v = \beta c$

$$\begin{aligned} \rho(r, \vartheta, s; t) &= \frac{Q}{a_\infty} \delta(r - a) \delta_\rho(\vartheta) \delta(s - vt) \\ &= \sum_{m=0}^{\infty} \frac{Q_m \cos(m\vartheta)}{\pi a^{m+1} (1 + \delta_{m0})} \delta(r - a) \delta(s - vt) = \sum_{m=0}^{\infty} \rho_m, \end{aligned} \quad (4.13)$$

with $\mathbf{J}_m = \rho_m \mathbf{v}$, where $Q_m = Q a^m$ and δ is the Dirac function. In this case the whole solution can be written as, for $m \geq 0$ and $\beta = 1$ (with q the charge of the test particle and L the length of the structure)

$$\begin{aligned} v \Delta p_s(r, \theta, z) &= \int_0^L F_s ds = -q Q a^m r^m \cos m\theta W'_m(z) \\ v \Delta p_r(r, \theta, z) &= \int_0^L F_r ds = -q Q a^m m r^{m-1} \cos m\theta W_m(z), \\ v \Delta p_\theta(r, \theta, z) &= \int_0^L F_\theta ds = q Q a^m m r^{m-1} \sin m\theta W_m(z). \end{aligned} \quad (4.14)$$

The function $W_m(z)$ is called the transverse wake function (whose unit is $\text{VC}^{-1} m^{-2m}$) and $W'_m(z)$ is called the longitudinal wake function (whose unit is $\text{VC}^{-1} m^{-2m+1}$) of azimuthal mode m . They describe the shock response of the vacuum chamber environment to a δ -function beam which carries a m th moment. The integrals (on the left) are called wake potentials. The longitudinal wake function for $m = 0$ and transverse wake function for $m = 1$ are therefore given by

$$\begin{aligned} W'_0(z) &= -\frac{1}{qQ} \int_0^L F_s ds = -\frac{1}{Q} \int_0^L E_s ds, \\ W_1(z) &= -\frac{1}{qQa} \int_0^L F_x ds = -\frac{1}{Qa} \int_0^L (E_x - vB_y) ds. \end{aligned} \quad (4.15)$$

The Fourier transform of the wake function is called the impedance. The idea of representing the accelerator environment by an impedance was introduced by Vaccaro [25] and Sessler [26]. As the conductivity, permittivity and permeability of a material depend in general on frequency, it is usually better (or easier) to treat the problem in the frequency domain, i.e. compute the impedance instead of the wake function. It is also easier to treat the case $\beta \neq 1$. Then, an inverse Fourier transform is applied to obtain the wake function in the time domain. The different relations linking the wake functions and the impedances are given by (with $k = \omega/v$, $\omega = 2\pi f$ with f the frequency, and j the imaginary unit)

$$\begin{aligned} Z_m^\parallel(\omega) &= \int_{-\infty}^{\infty} W'_m(z) e^{jkz} \frac{dz}{v} = \int_{-\infty}^{\infty} W'_m(t) e^{jks} e^{-j\omega t} dt, \\ Z_m^\perp(\omega) &= -j \int_{-\infty}^{\infty} W_m(z) e^{jkz} \frac{dz}{v} = -j \int_{-\infty}^{\infty} W_m(t) e^{jks} e^{-j\omega t} dt, \\ W'_m(z) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} Z_m^\parallel(\omega) e^{-jkz} d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} Z_m^\parallel(\omega) e^{-jks} e^{j\omega t} d\omega, \\ W_m(z) &= \frac{j}{2\pi} \int_{-\infty}^{\infty} Z_m^\perp(\omega) e^{-jkz} d\omega = \frac{j}{2\pi} \int_{-\infty}^{\infty} Z_m^\perp(\omega) e^{-jks} e^{j\omega t} d\omega. \end{aligned} \quad (4.16)$$

The unit of the longitudinal impedance $Z_m^{\parallel}(\omega)$ is Ωm^{-2m} while the unit of the transverse impedance $Z_m^{\perp}(\omega)$ is Ωm^{-2m+1} . Furthermore, two important properties of impedances can be derived. The first is a consequence of the fact that the wake function is real, which leads to

$$\left[Z_m^{\parallel}(\omega) \right]^* = Z_m^{\parallel}(-\omega), \quad -\left[Z_m^{\perp}(\omega) \right]^* = Z_m^{\perp}(-\omega), \quad (4.17)$$

where $*$ stands for the complex conjugate. The second is a consequence of Panofsky-Wenzel theorem

$$Z_m^{\parallel}(\omega) = k Z_m^{\perp}(\omega). \quad (4.18)$$

Another interesting property of the impedances is the directional symmetry (Lorentz reciprocity theorem): the same impedance is obtained from both sides if the entrance and exit are the same.

A more general definition of the impedances (still for a cylindrically symmetric structure) is the following

$$Z_m^{\parallel}(\omega) = -\frac{1}{Q_m^2} \int dV E_m^{\parallel} J_m^*, \quad Z_m^{\perp}(\omega) = -\frac{1}{k Q_m^2} \int dV E_m^{\parallel} J_m^*, \quad (4.19)$$

where $dV = r dr d\vartheta ds$. For the previous ring-shape source it yields

$$\begin{aligned} Z_0^{\parallel}(\omega) &= -\frac{1}{Q_0} \int_0^L ds E_s(r=a) e^{jks}, \\ Z_1^{\perp}(\omega) &= -\frac{L}{k\pi a Q_1} \int_0^{2\pi} d\vartheta E_s(r=a, \vartheta, s) \cos \vartheta e^{jks}. \end{aligned} \quad (4.20)$$

The situation is more involved in the case of non axi-symmetric structures (due in particular to the presence of the quadrupolar wake field, already discussed in Sect. 4.1) and for $\beta \neq 1$, as in this case some electromagnetic fields also appear in front of the source particle. In the case of axi-symmetric structures, a current density with some azimuthal Fourier component creates electromagnetic fields with the same azimuthal Fourier component. In the case of non axi-symmetric structures, a generalized notion of impedances was introduced by Tsutsui [27], where a current density with some azimuthal Fourier component may create an electromagnetic field with various different azimuthal Fourier components. If the source particle 1 and test particle 2 have the same charge q , and in the ultra-relativistic case, the transverse wake potentials can be written (taking into account only the linear terms with respect to the source and test particles and neglecting the constant, coupling and high order terms) [28]

$$\begin{aligned} \int_0^L F_x ds &= -q^2 \left[x_1 W_x^{\text{driving}}(z) - x_2 W^{\text{detuning}}(z) \right], \\ \int_0^L F_y ds &= -q^2 \left[y_1 W_y^{\text{driving}}(z) + y_2 W^{\text{detuning}}(z) \right], \end{aligned} \quad (4.21)$$

where the driving term is used here instead of dipolar and detuning instead of quadrupolar (or incoherent). In the frequency domain, Eq. (4.21) leads to the following generalized impedances

$$\begin{aligned} Z_x [\Omega] &= x_1 Z_x^{\text{driving}} - x_2 Z^{\text{detuning}}, \\ Z_y [\Omega] &= y_1 Z_x^{\text{driving}} + y_2 Z^{\text{detuning}}. \end{aligned} \quad (4.22)$$

Note that in the case $\beta \neq 1$, another quadrupolar term is found [29].

From Eqs. (4.21) and (4.22), the procedure to simulate or measure the driving and detuning contributions can be deduced. In the time domain, using some time-domain electromagnetic codes like for instance CST Particle Studio [30], the driving and detuning contributions can be disentangled. A first simulation with $x_2 = 0$ gives the dipolar part while a second one with $x_1 = 0$ provides the quadrupolar part. It should be noted that if the simulation is done with $x_1 = x_2$, only the sum of the dipolar and quadrupolar parts is obtained. The situation is more involved in the frequency domain, which is used for instance for impedance measurements on a bench [31]. Two measurement techniques can be used to disentangle the transverse driving and detuning impedances, which are both important for the beam dynamics (this can also be simulated with codes like Ansoft-HFSS [32]). The first uses two wires excited in opposite phase (to simulate a dipole), which yields the transverse driving impedance only. The second consists in measuring the longitudinal impedance, as a function of frequency, for different transverse offsets using a single displaced wire. The sum of the transverse driving and detuning impedances is then deduced applying the Panofsky-Wenzel theorem in the case of top/bottom and left/right symmetry [33]. Subtracting finally the transverse driving impedance from the sum of the transverse driving and detuning impedances obtained from the one-wire measurement yields the detuning impedance only. If there is no top/bottom or left/right symmetry the situation is more involved [34].

Both longitudinal and transverse resistive-wall impedances were already calculated 40 years ago by Laslett, Neil and Sessler [35]. However, a new physical regime was revealed by the CERN LHC collimators. A small aperture paired with a large wall thickness asks for a different physical picture of the transverse resistive-wall effect from the classical one. The first unstable betatron line in the LHC is around 8 kHz, where the skin depth for graphite (whose measured isotropic DC resistivity is $10 \mu\Omega\text{m}$) is 1.8 cm. It is smaller than the collimator thickness of 2.5 cm. Hence one could think that the resistive thick-wall formula would be about right. In fact it is not. The resistive impedance is about two orders of magnitude lower at this frequency, as can be seen on Fig. 4.3. A number of papers have been published on this subject using the field matching technique starting from the Maxwell equations and assuming a circular geometry [36–40]. New results have been also obtained for flat chambers, extending the (constant) Yokoya factors to frequency and material dependent ones [41], as was already found with some simplified kicker impedance models [42, 43]. Note that the material resistivity may vary with the magnetic field

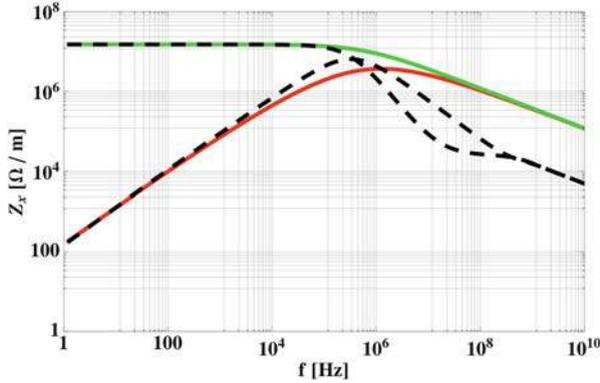


Fig. 4.3 Horizontal (driving) impedance of a cylindrical one-meter long LHC collimator (even if in reality the LHC collimators are composed of two parallel plates), with $b = 2$ mm and $\rho = 10 \mu\Omega\text{m}$. The real part is in red while the imaginary part is in green (note that in the classical thick-wall regime, the real and imaginary parts are equal). The dashed curves correspond to the case with a copper coating of $5 \mu\text{m}$ [1]

through the magneto-resistance and the surface impedance can also increase due to the anomalous skin effect ([44] and references therein).

In the case of a cavity, an equivalent RLC circuit can be used with R_s the longitudinal shunt impedance, C the capacity and L the inductance. In a real cavity, these three parameters cannot be separated easily and some other related parameters are used, which can be measured directly such as the resonance (angular) frequency $\omega_r = 1/\sqrt{LC}$, the quality factor $Q = R_s\sqrt{C/L} = R_s/(L\omega_r) = R_s C\omega_r$ and the damping rate $\alpha = \omega_r/(2Q)$. When $Q = 1$, the resonator impedance is called “broad-band”, and this model was extensively used in the past in many analytical computations. The longitudinal and transverse impedances and wake functions (with R_\perp the transverse shunt impedance) are given by [45] (see Fig. 4.4):

$$\begin{aligned}
 Z_m^\parallel(\omega) &= \frac{R_s}{1+jQ\left(\frac{\omega}{\omega_r}-\frac{\omega_r}{\omega}\right)}, & W_m^\parallel(t) &= \frac{\omega_r R_s}{Q} e^{-\alpha t} \left[\cos(\bar{\omega}_r t) - \frac{\alpha}{\omega_r} \sin(\bar{\omega}_r t) \right] \\
 Z_m^\perp(\omega) &= \frac{\omega_r}{1+jQ\left(\frac{\omega}{\omega_r}-\frac{\omega_r}{\omega}\right)} \frac{R_\perp}{\omega}, & W_m^\perp(t) &= \frac{\omega_r^2 R_\perp}{Q\bar{\omega}_r} e^{-\alpha t} \sin(\bar{\omega}_r t), & \bar{\omega}_r &= \omega_r \sqrt{1 - \frac{1}{4Q^2}}.
 \end{aligned}
 \tag{4.23}$$

Finally, all the transverse impedances (dipolar or driving and quadrupolar or detuning) should be weighted by the betatron function at the location of the impedances, as this is what matters for the effect on the beam, i.e. for the beam dynamics. Furthermore, all the weighted impedances can be summed and lumped at a single location around the ring (as the betatron phase advance does not play a role [46]) such that the transverse impedance-induced instabilities can be studied by considering only one interaction per turn, as it was confirmed in the past by performing macro-particle tracking simulations and comparing the cases of

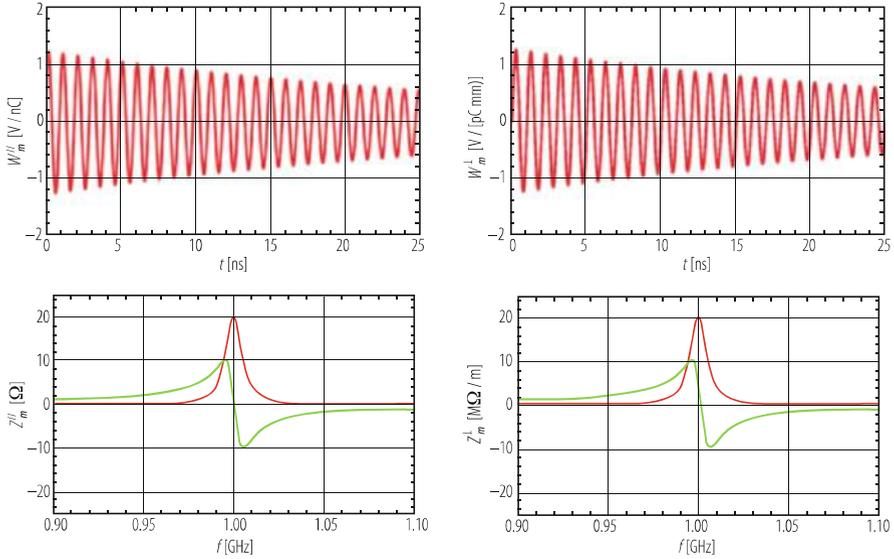


Fig. 4.4 Longitudinal and transverse impedances and wake functions, in the case of resonator impedances ($f_r = 1$ GHz, $Q = 100$, $R_s = 20 \Omega$, and $R_\perp = 20$ M Ω /m)

distributed kickers in the CERN SPS with the corresponding lumped impedance: exactly the same result was obtained [47].

4.3 Coherent Instabilities

E. Metral

The wake fields can influence the motion of trailing particles, in the longitudinal and in one or both transverse directions, leading to energy loss, beam instabilities, or producing undesirable secondary effects such as excessive heating of sensitive components at or near the chamber wall. Therefore, in practice the elements of the vacuum chamber should be designed to minimise the self-generated electromagnetic fields. For example, chambers with different cross-sections should be connected with tapered transitions; bellows need to be separated from the beam by shielding; plates should be grounded or terminated to avoid reflections; high-resistivity materials should be coated with a thin layer of very good conductor (such as copper) when possible; etc.

Two approaches are usually used to deal with collective instabilities. One starts from the single-particle equation while the other solves the Vlasov equation, which is nothing else but an expression for the Liouville conservation of phase-space density seen by a stationary observer. In the second approach, the motion of the

beam is described by a superposition of modes, rather than a collection of individual particles. The detailed methods of analysis in the two approaches are different, the particle representation is usually conveniently treated in the time domain, while in the mode representation the frequency domain is more convenient, but in principle they necessarily give the same final results. The advantage of the mode representation is that it offers a formalism that can be used systematically to treat the instability problem.

The first formalism was used by Courant and Sessler to describe the transverse coupled-bunch instabilities [48]. In most accelerators, the RF acceleration mechanism generates an azimuthal non-uniformity of the particle density and consequently the work of Laslett, Neil and Sessler for continuous beams [35] is not applicable in the case of bunched beams. Courant and Sessler studied the case of rigid (point-like) bunches, i.e. bunches oscillating as rigid units, and they showed that the transverse electromagnetic coupling of bunches of particles with each other can lead (due to the imperfectly conducting vacuum chamber walls) to a coherent instability. The physical basis of the instability is that in a resistive vacuum tank, fields due to a particle decay only very slowly in time after the particle has left (this leads to a long-range interaction). The decay can be so slow that when a bunch returns after one (or more) revolutions it is subject to its own residual wake field which, depending upon its phase relative to the wake field, can lead to damped or anti-damped transverse motion. For M equi-populated equi-spaced bunches, M coupled-bunch mode numbers exist ($n = 0, 1, \dots, M - 1$), characterized by the integer number of waves of the coherent motion around the ring. Therefore the coupled-bunch mode number resembles the azimuthal mode number for coasting beams, except that for coasting beams there is an infinite number of modes. The bunch-to-bunch phase shift $\Delta\phi$ is related to the coupled-bunch mode number n by $\Delta\phi = 2\pi n/M$.

Pellegrini [49] and, independently, Sands [50, 51] then showed that short-range wake fields (i.e. fields that provide an interaction between the particles of a bunch but have a negligible effect on subsequent passages of the bunch or of other bunches in the beam) together with the internal circulation of the particles in a bunch can cause internal coherent modes within the bunch to become unstable. The important point here is that the betatron phase advance per unit of time (or betatron frequency) of a particle depends on its instantaneous momentum deviation (from the ideal momentum) in first order through the chromaticity and the slip factor. Considering a non-zero chromaticity couples the betatron and synchrotron motions, since the betatron frequency varies around a synchrotron orbit. The betatron phase varies linearly along the bunch (from the head) and attains its maximum value at the tail. The total betatron phase shift between head and tail is the physical origin of the head tail instability. The head and the tail of the bunch oscillate therefore with a phase difference, which reduces to rigid-bunch oscillations only in the limit of zero chromaticity. A new (within-bunch) mode number $m = \dots, -1, 0, 1, \dots$, also called head-tail (or azimuthal) mode number, was introduced. This mode describes the number of betatron wavelengths (with sign) per synchrotron period.

The work of Courant and Sessler, or Pellegrini and Sands, was done for particular impedances and oscillation modes. Using the Vlasov formalism, Sacherer unified

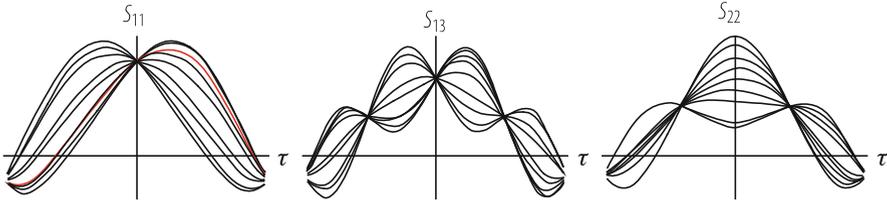


Fig. 4.5 Signal at the pick-up electrode for three different modes shown for several superimposed turns (the red line corresponds to one particular turn), for the case of the parabolic amplitude distribution and a constant inductive impedance (exhibiting therefore no growing oscillations!)

the two previous approaches, introducing a third mode number $q \equiv |m| + 2k$ (with $0 \leq k < +\infty$), called radial mode number, which comes from the distribution of synchrotron oscillation amplitudes [52, 53]. It can be obtained by superimposing several traces of the directly observable average displacement along the bunch at a particular pick-up. The number of nodes is the mode number q (see Figs. 4.5 and 4.8). The advantage of this formalism is that it is valid for generic impedances and any high order head–tail modes. This approach starts from a distribution of particles (split into two different parts, a stationary distribution and a perturbation), on which Liouville theorem is applied. After linearization of the Vlasov equation, one ends up with Sacherer’s integral equation or Laclare’s eigenvalue problem to be solved [53]. Because there are two degrees of freedom (phase and amplitude), the general solution is a twofold infinity of coherent modes of oscillation (m, q) . At sufficiently low intensity, only the most coherent mode (largest value for the coherent tune shift) is generally considered, leading to the classical Sacherer’s formulae in both transverse and longitudinal planes. Note that contrary to the space charge case, these tune shifts are now complex, the imaginary part being linked to the instability growth rate. For protons a parabolic density distribution is generally assumed and the corresponding oscillation modes are sinusoidal (or close to it). For electrons, the distribution is usually Gaussian, and the oscillation modes are described in this case by Hermite polynomials. In reality, the oscillation modes depend both on the distribution function and the impedance, and can only be found numerically by solving the (infinite) eigenvalue problem. However, the mode frequencies are usually not very sensitive to the accuracy of the eigenfunctions. Similar results are obtained for the longitudinal plane.

It is worth mentioning that the CERN ISR suffered from a beam instability brought about by beams having different revolution frequencies. They could be in the same vacuum chamber or coupled by the beam–beam effect. The name of “overlap knock-out” [54] has been given to this phenomenon by which the stack is subjected to transverse kicks from the bunches. This produces blow-up of the stacked beam when the longitudinal frequency spectrum of the bunches overlaps with the betatron frequency spectrum of the coasting stacked beam. Similar problems limit the energy range of RHIC and proton lead in LHC [55].

4.3.1 Longitudinal

The most fundamental longitudinal instability encountered in circular accelerators is called the Robinson instability. The (Radio-Frequency) RF frequency accelerating cavities in a circular accelerator are tuned so that the resonant frequency of the fundamental mode is very close to an integral multiple of the revolution frequency of the beam. This necessarily means that the wake field excited by the beam in the cavities contains a major frequency component near a multiple of the revolution frequency. The exact value of the resonant frequency relative to the multiple of the revolution frequency is of critical importance for the stability of the beam. Above the transition energy, the beam will be unstable if the resonant frequency is slightly above it and stable if slightly below. This is the opposite below transition. This instability mechanism was first analyzed by Robinson [56]. Physically, the Robinson instability comes from the fact that the revolution frequency of an off-momentum beam is not given by the on-momentum revolution frequency, but by a quantity slightly different, depending on both the slip factor and the energy deviation.

Let's assume in the following that the Robinson criterion is met. A bunch is longitudinally stable if the longitudinal profile observed at a wall-current monitor is constant turn after turn and it is unstable if the longitudinal profile is not constant turn after turn. In the case of instability, the way the longitudinal profile oscillates gives some information about the type of instabilities. This was studied in detail by Laclare [53], who explained theoretically such pictures of "longitudinal (single-bunch) instability" starting from the single-particle longitudinal signal at a pick-up electrode (assuming infinite bandwidth). The current signal induced by the test particle is a series of impulses delivered on each passage

$$s_z(t, \vartheta) = e \sum_{k=-\infty}^{k=+\infty} \delta \left(t - \tau - \frac{\vartheta}{\Omega_0} - \frac{2k\pi}{\Omega_0} \right), \quad (4.24)$$

where τ is the time interval between the passage of the synchronous particle and the test particle, for a fixed observer at azimuthal position ϑ and Ω_0 is the angular revolution frequency. In the frequency domain, the single-particle spectrum is therefore a line spectrum at (angular) frequencies $\omega_{pm} = p\Omega_0 + m\omega_{s0}$, where ω_{s0} is the small-amplitude synchrotron frequency. Around every harmonic of the revolution frequency $p\Omega_0$, there is an infinite number of synchrotron satellites m (it is different from the one used in Sect. 4.2!), whose spectral amplitude is given by the Bessel function $J_m(p\Omega_0\hat{\tau})$, where $\hat{\tau}$ is the synchrotron amplitude. The spectrum is centred at the origin and because the argument of the Bessel functions is proportional to $\hat{\tau}$, the width of the spectrum behaves like $\hat{\tau}^{-1}$. Applying the Vlasov equation, linearizing it, and studying the effect of the impedance on the unperturbed distribution leads to the potential-well effect: a new fixed point is reached, with a new synchronous phase, a new effective voltage, a new synchrotron frequency, a new bunch length and a new momentum spread, which all depend on intensity. Studying a perturbation on top of the new stationary distribution, one ends up at low

intensity, i.e. considering independently the modes m (which is valid up to a certain intensity), with the following eigenvalue system

$$\Delta\omega_{cmq}\sigma_{mq}(l) = \sum_{p=-\infty}^{p=+\infty} K_{lp}^m \sigma_{mq}(p), \text{ with } \Delta\omega_{cmq} = \omega_{cmq} - m\omega_s, \quad (4.25)$$

where

$$K_{lp}^m = -\frac{2\pi I_b m \omega_s}{\Omega_0^2 \hat{V}_T h \cos \phi_s} j \frac{Z_l(p)}{p} \int_{\hat{\tau}=0}^{\hat{\tau}=+\infty} \frac{dg_0}{d\hat{\tau}} J_m(p\Omega_0\hat{\tau}) J_m(l\Omega_0\hat{\tau}) d\hat{\tau}. \quad (4.26)$$

Here, $\Delta\omega_{cmq}$ is the coherent complex synchrotron frequency shift to be determined, $I_b = N_b e \Omega_0 / 2\pi$ is the bunch current, ω_s , \hat{V}_T and ϕ_s are the new synchrotron frequency, total voltage and synchronous phase (taking into account the potential-well distortion), $Z_l = Z_0^{\parallel}$ is the longitudinal impedance, g_0 is the longitudinal amplitude density function and h the RF harmonic number. The procedure to obtain first order exact solutions, with realistic modes and a general interaction, thus consists of finding the eigenvalues and eigenvectors of the infinite complex matrix whose elements are given by Eq. (4.26). The result is an infinite number of modes mq of oscillation. To each mode, one can associate a coherent frequency shift $\Delta\omega_{cmq} = \omega_{cmq} - m\omega_s$ (which is the q th eigenvalue), a coherent spectrum $\sigma_{mq}(p)$ (which is the q th eigenvector) and a perturbation distribution $g_{mq}(\hat{\tau})$. For numerical reasons, the matrix needs to be truncated, and thus only a finite frequency domain is explored. For the case of the parabolic amplitude distribution and a constant inductive impedance (which leads to real tune shifts only and therefore no instability), the signal at the pick-up electrode shown for several superimposed turns is depicted on Fig. 4.5. In the case of a complex impedance, the real part will lead in addition to a growing amplitude with an associated instability rise-time. The spectrum of mode mq is peaked at $f_q \approx (q+1)/(2\tau_b)$ and extends $\sim \pm\tau_b^{-1}$, where τ_b is the full bunch length (in second). It can be seen from Fig. 4.5 that there are q nodes on these ‘‘standing-wave’’ patterns. The longitudinal signal at the pick-up electrode is given by

$$S_{mq}(t, \vartheta) = S_{z0}(t, \vartheta) + \Delta S_{zmq}(t, \vartheta), \quad (4.27)$$

$$S_{z0}(t, \vartheta) = 2\pi I_b \sum_{p=-\infty}^{p=+\infty} \sigma_0(p) e^{jp\Omega_0 t} e^{-jp\vartheta}, \quad \sigma_0(p) = \int_{\hat{\tau}=0}^{\hat{\tau}=+\infty} J_0(p\Omega_0\hat{\tau}) g_0(\hat{\tau}) \hat{\tau} d\hat{\tau}, \quad (4.28)$$

$$\Delta S_{zmq}(t, \vartheta) = 2\pi I_b \sum_{p=-\infty}^{p=+\infty} \sigma_{mq}(p) e^{j(p\Omega_0 + m\omega_s + \Delta\omega_{cmq})t} e^{-jp\vartheta}. \quad (4.29)$$

Finding the eigenvalues and eigenvectors of a complex matrix by computer can be difficult in some cases, and a simple approximate formula for the eigenvalues is useful in practice to have a rough estimate. This is known as Sacherer's (longitudinal) formula [52]. Sacherer's formula is also valid for coupled-bunch instability with M equally-populated equally-spaced bunches, assuming multi-bunch modes with only one type of internal motion. In the case of gaps between bunch trains, a time-domain approach is usually better suited.

As the bunch intensity increases, the different longitudinal modes can no longer be treated separately and the situation is more involved. In the longitudinal plane, the microwave instability for coasting beams is well understood. It leads to a stability diagram, which is a graphical representation of the solution of the dispersion relation (taking into account the momentum spread) depicting curves of constant growth rates, and especially a threshold contour in the complex plane of the driving impedance (see Sect. 4.4) [57]. When the real part of the driving impedance is much greater than the modulus of the imaginary part, a simple approximation, known as the Keil-Schnell (or circle) stability criterion, may be used to estimate the threshold curve [58]. For bunched beams, it has been proposed by Boussard to use the coasting-beam formalism with local values of bunch current and momentum spread [59]. A first approach to explain this instability, without coasting-beam approximations, has been suggested by Sacherer through Longitudinal Mode-Coupling (LMC) [60]. The equivalence between LMC Instabilities (LMCI) and microwave instabilities has been pointed out by Sacherer and Laclare [53] in the case of broad-band driving resonator impedances, when the bunch length is much greater than the inverse of twice the resonance frequency. Furthermore, due to the potential well-distortion, a bunch is more stable below transition than above [53, 61, 62]. Typical pictures of LMCI are shown in Fig. 4.6 [63] and a comparison with macroparticle tracking simulations, which revealed a good agreement, is discussed in Ref. [64]. Experimentally, the most evident signature of the LMCI is the intensity-dependent longitudinal beam emittance blow-up to remain just below the threshold [65], as revealed also by macroparticle tracking simulations (see Fig. 4.7 [64]).

4.3.2 Transverse

A similar analysis as the one done for the longitudinal plane can be done in the transverse plane [53, 63]. Following the same procedure, the horizontal coherent oscillations (over several turns) of a "water-bag" bunch (i.e. with constant longitudinal amplitude density) interacting with a constant inductive impedance are shown in Fig. 4.8.

The main difference with the longitudinal plane is that there is no effect of the stationary distribution and the bunch spectrum is now centered at the chromatic frequency $f_\xi = Q_x f_0 \xi / \eta$, where f_0 is the revolution frequency and ξ is the relative chromaticity. The sign of the chromatic frequency is very important and to avoid the head-tail instability (of mode 0) it should be slightly positive, meaning that

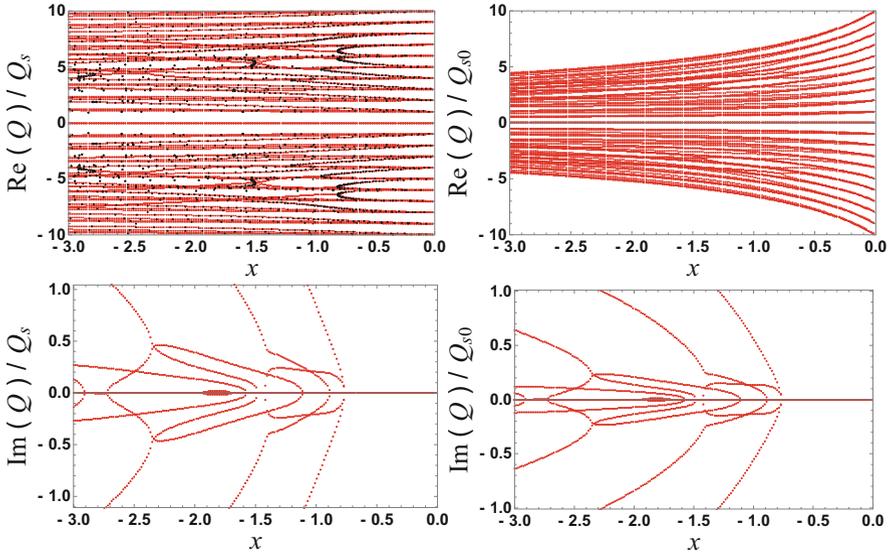


Fig. 4.6 (Left) Comparison between GALACLIC Vlasov Solver [63] (in red) and Laclare’s approach [53] (in black) of the normalised mode-frequency shifts vs. the normalised parameter x (proportional to the bunch intensity [63]), in the case of a broad-band resonator impedance (with a quality factor of 1 and a resonance frequency f_r such that $f_r \tau_b = 2.8$), above transition, without taking into account the potential-well distortion (this is why the intensity-dependent synchrotron tune Q_s is used) and for a “Parabolic Amplitude Density” (PAD) longitudinal distribution [53]. (Right) Similar plot from GALACLIC only, taking into account the potential-well distortion (this is why the low-intensity synchrotron tune Q_{s0} is used)

the chromaticity should be negative below transition and positive above. Sacherer’s formula is also valid for coupled-bunch instability with M equally-populated equally-spaced bunches, assuming multi-bunch modes with only one type of internal motion (i.e. the same head–tail mode number). This analysis was extended in Ref. [66] to include also the coupling between the modes (and the possibility to have different head–tail modes in the different bunches). In the case of gaps between bunch trains, a time-domain approach is usually better suited.

At low intensity (i.e. below a certain intensity threshold), the standing-wave patterns (head–tail modes) are treated independently. This leads to instabilities where the head and the tail of the bunch exchange their roles (due to synchrotron oscillation) several times during the rise-time of the instability. The (approximate) complex transverse coherent betatron frequency shift of bunched-beam modes is given by Sacherer’s formula for round pipes [52]. For flat chambers a quadrupolar effect (see Sect. 4.2) has to be added to obtain the real part of the coherent tune shift, which explains why the horizontal coherent tune shift is zero in horizontally flat chambers (of good conductors). As an example, a head–tail instability with mode $q = 10$ is shown in Fig. 4.9(left). It is worth mentioning that there is also a head–tail instability in the longitudinal plane. The longitudinal head–tail instability,

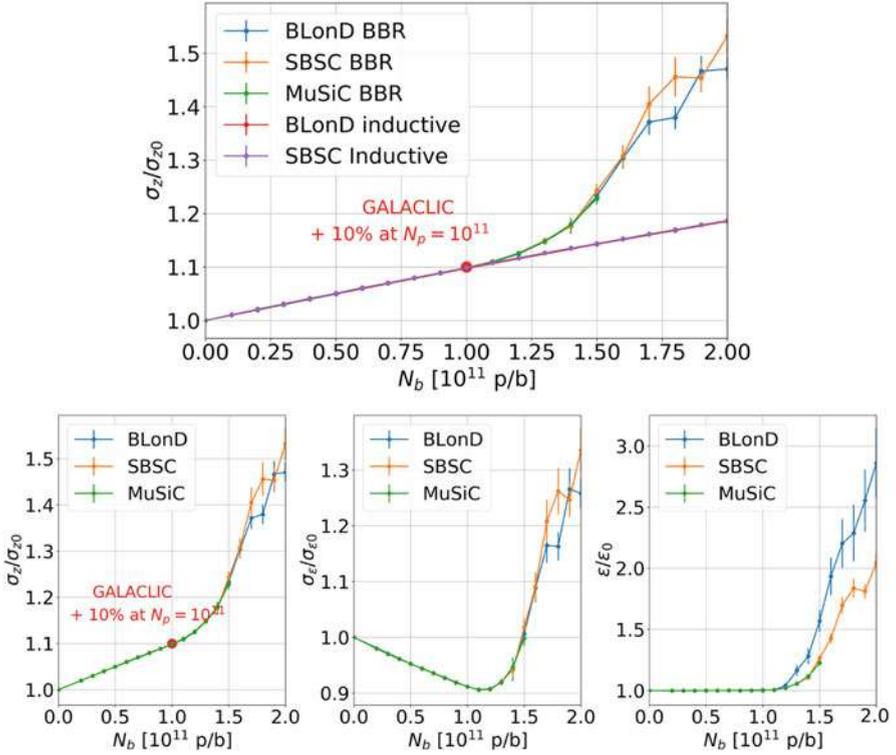


Fig. 4.7 Simulation results from BLonD, SBSC and MuSiC codes [64] corresponding to the case mentioned above: (upper) evolution of the normalised rms bunch length vs. bunch intensity for the cases of broad-band resonator and constant inductive impedances; (lower) evolution of the normalised rms bunch length, energy spread and longitudinal emittance vs. bunch intensity for the case of the broad-band resonator impedance. Courtesy of M. Migliorati [64]

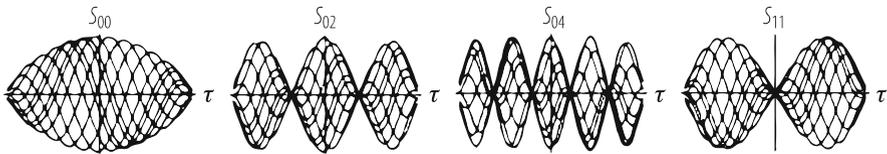


Fig. 4.8 Transverse signal at the pick-up electrode for four different modes shown for several superimposed turns, for the case of the “water-bag” bunch (i.e. with constant longitudinal amplitude density) and a constant inductive impedance. In the present example, the total phase shift between the head and the tail is given $\chi = \omega_\xi \tau_b = 10$ (see below)

first suggested by Hereward [67] and possibly observed at the CERN SPS [68] results from the fact that the slip factor is not strictly a constant: it depends on the instantaneous energy error just as the betatron frequency does. The longitudinal beam distribution then acquires a head–tail phase, and instability may arise as a result.

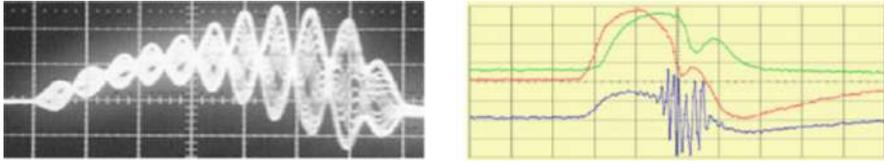


Fig. 4.9 (Left) Signal from a radial beam position monitor during 20 consecutive turns observed in the CERN PS at 1.4 GeV kinetic energy in 1999. Time scale: 20 ns/div. (Right) Fast instability observed in the CERN PS near transition (~ 6 GeV total energy) in 2000. Single-turn signals from a wide-band pick-up. From top to bottom: Σ , Δx , and Δy . Time scale: 10 ns/div. The head of the bunch is stable and only the tail is unstable in the vertical plane. The particles lost at the tail of the bunch can be seen from the hollow in the bunch profile

As the bunch intensity increases, the different head–tail modes can no longer be treated separately. In this regime, the wake fields couple the modes together and a wave pattern travelling along the bunch is created: this is the Transverse Mode Coupling Instability (TMCI). The TMCI for circular accelerators has been first described by Kohaupt [69] in terms of coupling of Sacherer’s head–tail modes. This extended to the transverse motion, the theory proposed by Sacherer to explain the longitudinal microwave instability through coupling of the longitudinal coherent bunch modes. The TMCI is the manifestation in synchrotrons of the Beam Break-Up (BBU) mechanism observed in linacs. The only difference comes from the synchrotron oscillation, which stabilises the beam in synchrotrons below a threshold intensity by swapping the head and the tail continuously. In fact, several analytical formalisms exist for fast (compared to the synchrotron period) instabilities, but the same formula is obtained (within a factor smaller than two) from five, seemingly diverse, formalisms in the case of a broad-band resonator impedance in the “long-bunch” regime [70], as recently confirmed in Ref. [71]: (i) Coasting-beam approach with peak values, (ii) Fast blow-up, (iii) BBU (for 0 chromaticity), (iv) Post head–tail, and (v) TMCI with 2 modes in the “long-bunch” regime (for 0 chromaticity). Two regimes are indeed possible for the TMCI according to whether the total bunch length is larger or smaller than the inverse of twice the resonance frequency of the impedance. The simple (approximate) formula reveals the scaling with the different parameters. In particular it can be seen that the instability does not disappear at high energy but saturates like the slip factor (what is important is not the energy but the distance from the transition energy) [72]. This means that the TMCI intensity threshold can be raised by changing the transition energy, i.e. by modifying the optics. Furthermore, the intensity threshold increases with the resonance frequency (as high-order head–tail modes will couple), with longitudinal emittance and with chromaticity. Note that the coherent synchrotron resonances, important in large machines, are not discussed here. This was checked with the MOSES Vlasov solver [73], which is a program computing the coherent bunched-beam modes. Below is a comparison between the MOSES code and the HEADTAIL code [74], which is a tracking code simulating single-bunch phenomena, in the case of a LHC-type single bunch at SPS injection [75]. As can be seen from Fig. 4.10, a very good

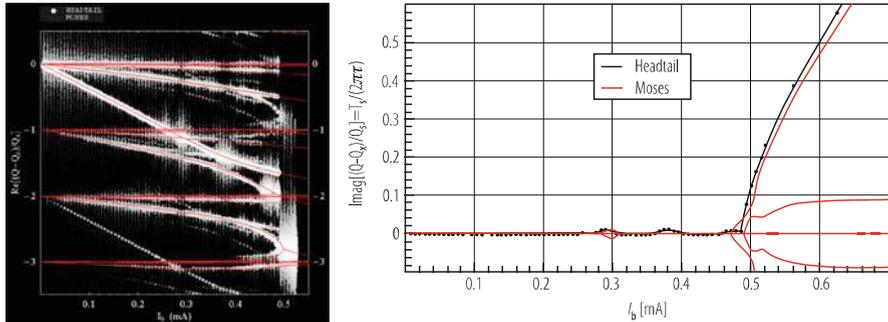


Fig. 4.10 Comparison between MOSES (in red) and HEADTAIL (in white) in the case of a broad-band resonator (Courtesy of Benoit Salvant [75]). Evolution of the real (left) and imaginary (right) parts of the shifts of the transverse modes (with respect to the unperturbed betatron tune), normalized by the synchrotron tune, vs. bunch intensity

agreement between the two was found. For a general impedance (i.e. not a resonator impedance) the situation is more involved and MOSES cannot be used: one should rely on HEADTAIL simulations or on the recently developed Vlasov solvers such as NHT [76] or DELPHI [77]. In the case of flat chambers, the intensity threshold is higher in one plane than in the other and linear coupling can be used to raise the TMCI intensity threshold [78]. Note finally that with many bunches the TMCI intensity threshold can be considerably reduced [66].

It is worth mentioning also all the work done for the TMCI in LEP, as Chin's work (with MOSES) came later. It was proposed to cure the TMCI with a reactive feedback that would prevent the zero mode frequency from changing with increasing beam intensity [79]. In [80, 81] a theory of reactive feedback has been developed in the two-particle approach and with the Vlasov equation. Theory has revealed that the reactive feedback can really appreciably increase the TMCI intensity threshold, which was confirmed by simulation [82, 83]. On the contrary, the resistive feedback was found to be "completely" ineffective as a cure for the TMCI [81]. An action of a feedback on the TMCI intensity threshold was later examined experimentally at PEP [84]. It was confirmed that a reactive feedback is indeed capable to increase the TMCI intensity threshold. But it turned out unexpectedly that a resistive feedback can also increase the TMCI intensity threshold and even more effectively [84]. In [85], an attempt was made to develop an advanced transverse feedback theory capable to elucidate the conditions at which the resistive or reactive or some intermediate feedback can cure the TMCI. Positive chromaticity above transition helps, but depending on the coupling impedance, beam stability may require a large value of the chromaticity either unattainable or which reduces the beam lifetime. It was proposed to have a negative chromaticity (what is usually avoided), where the zero mode is unstable (by head-tail instability) and all the other modes are damped, and stabilise this mode by a resistive feedback, keeping the higher order modes stable. In this case, the TMCI intensity threshold could be increased by a factor 3–5 [85]. In the last few years, several Vlasov solvers

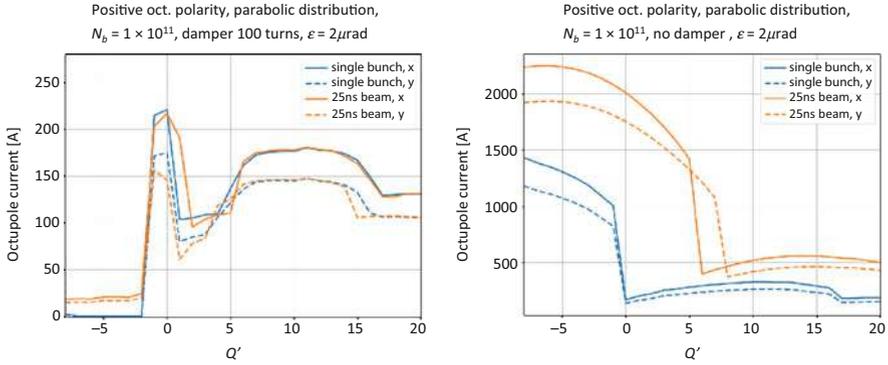


Fig. 4.11 Required Landau octupole current to stabilise the 2018 CERN LHC beam vs. chromaticity: (left) with resistive transverse damper and (right) without resistive transverse damper. Courtesy of N. Mounet [87]

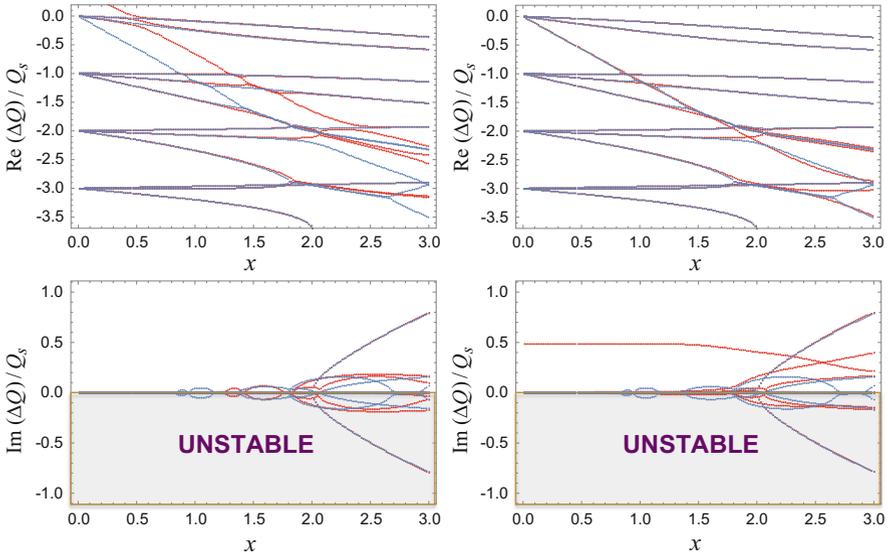


Fig. 4.12 Usual TMCI plots (for $f_r \tau_b = 2.8$, i.e. in the “long-bunch” regime) showing the real and imaginary parts of the normalised complex tune shift vs. the normalised parameter x (which is proportional to the bunch intensity [63]) without (in blue) and with (in red) a transverse damper: (left) reactive and (right) resistive [86]

were developed to take into account the effect of a transverse damper, such as NHT [76], DELPHI [77] and GALACTIC [86]. An example of DELPHI for the case of the LHC in 2018 is shown in Fig. 4.11, where the beneficial effect of the transverse resistive damper (on the required Landau octupole current needed to stabilise the beam) can be clearly seen. A comparison between a reactive and a resistive damper is shown in Figs. 4.12 and 4.13 using GALACTIC [86] (and a comparison between GALACTIC and Laclare’s approach [53] is discussed in Ref. [63]).

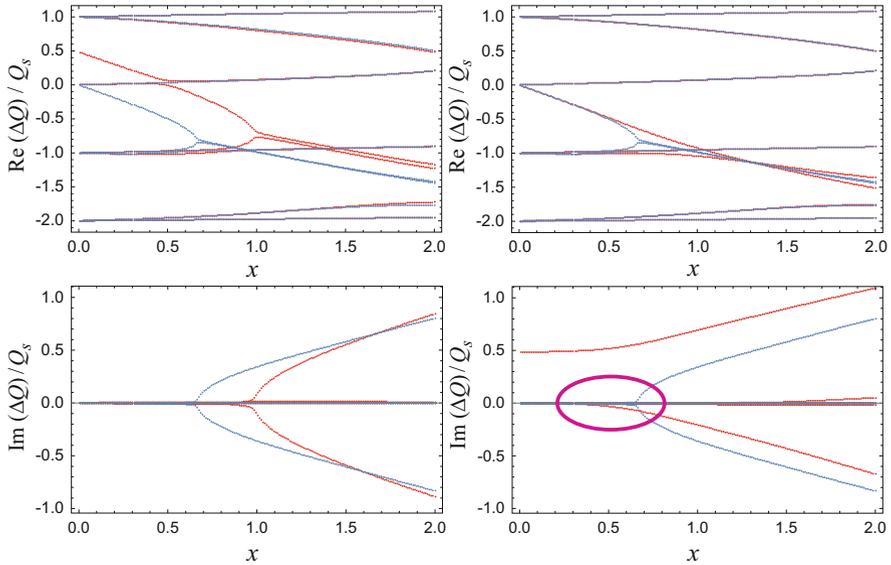
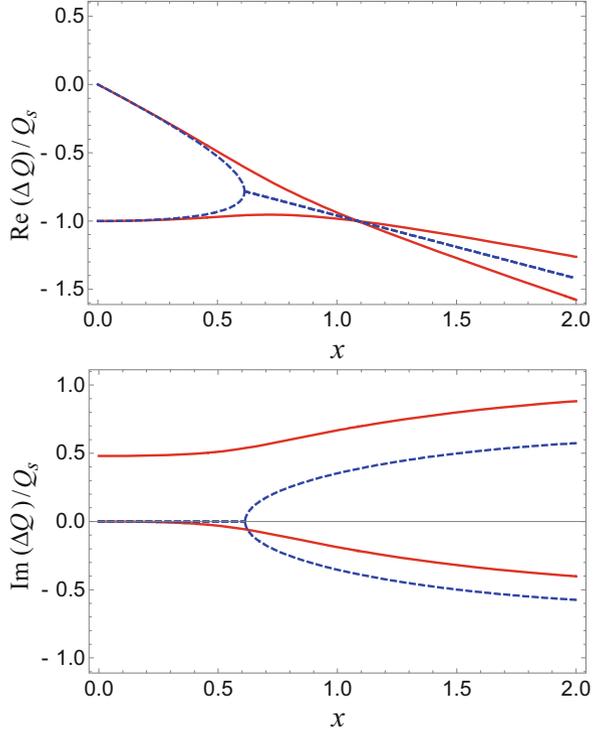


Fig. 4.13 Usual TMCI plots (for $f_r \tau_b = 0.8$, i.e. in the “short-bunch” regime) showing the real and imaginary parts of the normalised complex tune shift vs. the normalised parameter x (which is proportional to the bunch intensity [63]) without (in blue) and with (in red) a transverse damper: (left) reactive and (right) resistive [86]

As can be seen from Fig. 4.13(right) the resistive transverse damper exhibits a destabilising effect below the TMCI intensity threshold. This destabilising effect of (perfect) resistive transverse dampers was analysed in detail for the case of a single bunch with zero chromaticity [86]: in the presence of a resistive transverse damper the instability mechanism is completely modified as can be seen from Fig. 4.14. Due to the features, which are discussed in Ref. [86], the name “ISR (for Imaginary tune Split and Repulsion) instability” was suggested for this new kind of single-bunch instability with zero chromaticity.

It is also worth mentioning that in the case of hadrons (compared to leptons), another ingredient which should be taken into account while studying the transverse instabilities is space charge. This has been a subject of discussion for the last two decades as space charge was believed initially to be mainly beneficial as e.g. for the previous case of the CERN SPS TMCI predicted in the absence of space charge. It was recently found that space charge is actually destabilising in such a case (“long-bunch” regime) [88–90], while it is beneficial in the “short-bunch” regime [88, 89]. This is clearly revealed in Figs. 4.15 and 4.16, but still some work is needed to fully understand what happens.

Fig. 4.14 Usual TMCI plots showing the real and imaginary parts of the normalised complex tune shift vs. the normalised parameter x (which is proportional to the bunch intensity [63]) without (in blue) and with (in red) a resistive transverse damper [86]



4.4 Landau Damping

E. Metral

Several stabilising mechanisms exist which can prevent the previous instabilities from developing. One of them is Landau damping, which is a general process that arises when one considers a whole collection of particles or other systems, which have a spectrum of resonant frequencies, and interact in some way. In accelerators we are usually concerned with an interaction of a kind that may make the beam unstable (wake fields), and we want to find out whether or not (and how) the spread of resonant frequencies will stabilise it. If the particles have a spread in their natural frequencies, the motion of the particles can lose its coherency. In order to understand the physical origin of this effect, let us first consider a simple harmonic oscillator, which oscillates in the x -direction with its natural frequency [8]. Let this oscillator be driven, starting at time $t = 0$, by a sinusoidal force. The equation of motion is

$$\ddot{x} + \omega_x^2 x = f \cos(\omega_c t), \quad (4.30)$$

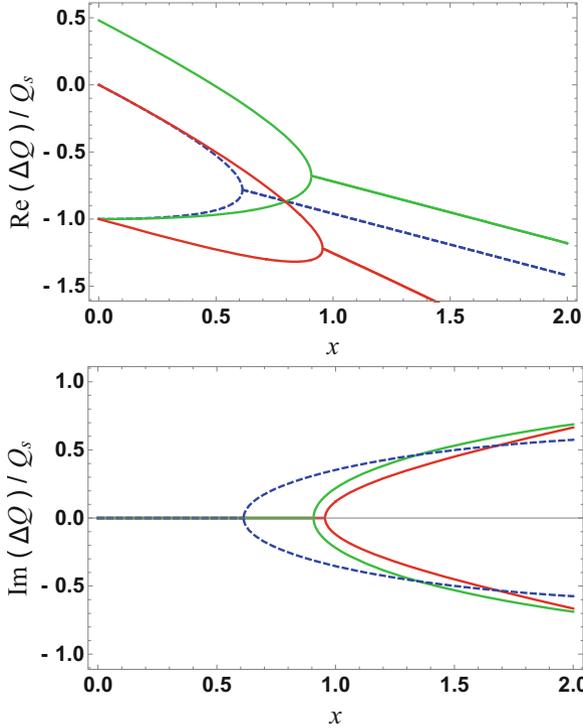


Fig. 4.15 Simplified model/example of Ref. [86], describing the mode-coupling in the “short-bunch” regime, i.e. the mode-coupling between modes 0 and -1 , extended here to take into account also space charge, using the parameters mentioned above: (dashed blue) with impedance only, (green) with impedance and a reactive transverse damper, (red) with impedance and space charge [88]. The normalised parameter x is proportional to the bunch intensity [63]

where a dot stands for derivative with respect to time and with $x(0) = 0$ and $\dot{x}(0) = 0$. The solution is

$$x(t > 0) = -\frac{f}{\omega_c^2 - \omega_x^2} [\cos(\omega_c t) - \cos(\omega_x t)] = \frac{f}{2\omega_{x0}} \sin(\omega_{x0} t) \frac{\sin[(\omega_c - \omega_x)t/2]}{(\omega_c - \omega_x)/2}. \quad (4.31)$$

Consider now an ensemble of oscillators (each oscillator represents a single particle in the beam) which do not interact with each other and have a spectrum of natural frequency ω_x with a distribution $\rho_x(\omega_x)$ normalised to unity. Let’s assume first that the origin of the betatron frequency spread is not specified: an externally given beam frequency spectrum is supposed. Now starting at time $t = 0$, subject this ensemble of particles to the driving force $f \cos(\omega_c t)$ with all particles starting with initial conditions $x(0) = 0$ and $\dot{x}(0) = 0$. We are interested in the ensemble average

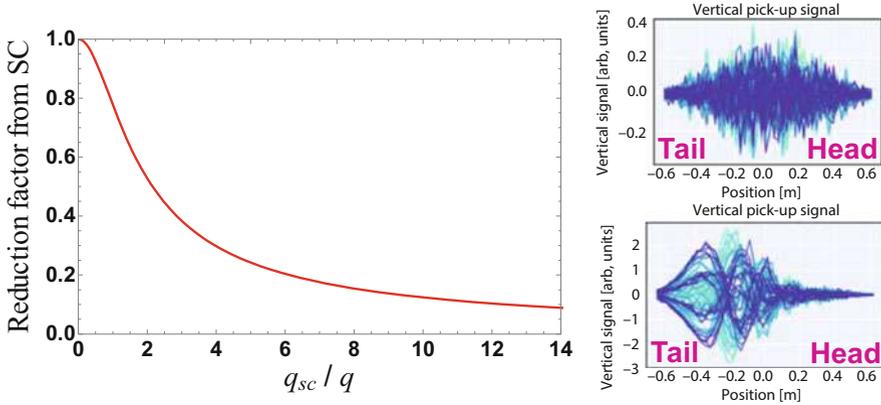


Fig. 4.16 (Left) reduction factor from a simplified model with space charge of the TMCI intensity threshold, as a function of the ratio between the space charge parameter q_{sc} and the radial mode number q , in the case of the “long-bunch” regime, as e.g. for the CERN SPS at injection [88]. (Right) simulated stable bunch without space charge (top) and unstable bunch with space charge (bottom) for the case $q_{sc}/q = 13.5$ with a bunch intensity a factor 3 lower than the TMCI intensity threshold without space charge [91] (Courtesy of A. Oeftiger)

of the response, which is given by superposition by

$$\bar{x}(t) = \frac{f}{2\omega_{x0}} \left[\cos(\omega_c t) \text{P.V.} \int_{-\infty}^{+\infty} \frac{\rho_x(\omega_x)}{\omega_x - \omega_c} d\omega_x + \pi \rho_x(\omega_c) \sin(\omega_c t) \right], \quad (4.32)$$

where P.V. stands for Principal Value. The sinus term has a definite sign relative to the driving force, because $\rho_x(\omega_c)$ is always positive. In particular, \bar{x} is always in phase with the force, indicating that work is being done on the system, which always reacts to the force “resistively”. The Landau damping effect is to be distinguished from a “decoherence (also called phase-mixing, or filamentation) effect” that occurs when the beam has nonzero initial conditions. Had we included an initial offset, we would have introduced two additional terms into the ensemble response, which do not participate in the dynamic interaction of the beam particles and are not interesting for our purposes here. In this decoherence effect, individual particles continue to execute oscillations of constant amplitude, but the total beam response \bar{x} decreases with time. As mentioned above, work is continuously being done on the system. However, the amplitude of \bar{x} , as given before, does not increase with time. Where did the energy go? The system absorbs energy from the driving force indefinitely while holding the ensemble beam response within bounds. The stored energy is incoherent in the sense that the energy is contained in the individual particles, but it is not to be regarded as heat in the system. This is because the stored energy is not distributed more or less uniformly in all particles, but is selectively stored in particles with continuously narrowing range of frequencies

around the driving frequency. If one observes two particles, one with the exciting frequency ω_c and one with a frequency slightly different, at the beginning, they oscillate “coherently” (same amplitude and same phase). However, after a while the particle with the exciting frequency, being resonantly driven, continues to increase in amplitude as time increases, whereas the other particle with a slightly different frequency realizes that its frequency is not the same as the driving one and the “beating” phenomenon is observed for this particle. If one considers the phenomenon for a time t , the number of particles which still oscillate coherently decreases with time as $1/t$, while their amplitude increases as t , the net contribution being constant with time.

The origins of the frequency spread that leads to Landau damping have not been taken into account till now. The case where the frequency spread comes from the longitudinal momentum spread of the beam is straightforward (for a coasting beam), because the longitudinal momentum is a constant, which just affects the coefficients in the equations of motion of the transverse oscillations, and hence their frequencies. It can be dealt with the same method as in the previous sections, i.e. it is the distribution function which is important. The same result applies also if one considers a tune spread that is due to a non-linearity (e.g. from octupole lenses) in the other plane. However, this result is no longer valid if the non-linearity is in the plane of coherent motion. In this case, the steady-state is more involved because the coherent motion is then a small addition to the large incoherent amplitudes that make the frequency spread, and it is inconsistent to assume that it can be treated as a linear superposition [92]. One needs to consider “second order” non-linear terms and the final result is that in this case it is not the distribution function which matters but its derivative. Using the Vlasov formalism, this result is recovered more straightforwardly.

4.4.1 Transverse

Considering the case of a beam having the same normalized rms beam size $\sigma = \sqrt{\varepsilon}$ in both transverse planes, the Landau damping mechanism from octupoles of coherent instabilities, e.g. in the horizontal plane, is discussed from the following dispersion relation [17, 93]

$$1 = -\Delta Q_{\text{coh}}^x \int_{J_x=0}^{+\infty} dJ_x \int_{J_y=0}^{+\infty} dJ_y \frac{J_x \frac{\partial f(J_x, J_y)}{\partial J_x}}{Q_c - Q_x(J_x, J_y) - mQ_s}, \quad (4.33)$$

with

$$Q_x(J_x, J_y) = Q_0 + a_0 J_x + b_0 J_y. \quad (4.34)$$

Here, Q_c is the coherent betatron tune to be determined, $J_{x,y}$ are the action variables in the horizontal and vertical plane respectively, with $f(J_x, J_y)$ the distribution function, ΔQ_{coh}^x is the horizontal coherent tune shift, $Q_x(J_x, J_y)$ is the horizontal tune in the presence of octupoles, m is the head–tail mode number, and Q_s is the small-amplitude synchrotron tune (the longitudinal spread is neglected).

The n th order distribution function is assumed to be

$$f(J_x, J_y) = a \left(1 - \frac{J_x + J_y}{b} \right)^n, \quad (4.35)$$

where a and b are constants to be determined by normalization, and which corresponds to a profile extending up to $\sqrt{2(n+3)}\sigma$. The dispersion relation of Eq. (4.33) can be re-written as

$$\Delta Q_{\text{coh}}^x = -\frac{a_0}{nab} I_n^{-1}(c, q), \quad (4.36)$$

with

$$I_n(c, q) = \int_{J_x=0}^1 dJ_x \int_{J_y=0}^{1-J_x} dJ_y \frac{J_x(1-J_x-J_y)^{n-1}}{q + J_x + cJ_y}, \quad (4.37)$$

$$q = \frac{Q_c - Q_0 - mQ_s}{-ba_0}, \text{ and } c = \frac{b_0}{a_0}. \quad (4.38)$$

It is convenient to write Eq. (4.36) in this way, with the left-hand-side (l.h.s) containing information about the beam intensity and the impedance and the right-hand-side (r.h.s) containing information about the beam frequency spectrum only. In the absence of frequency spread, the r.h.s. of Eq. (4.36) is equal to $Q_c - Q_0 - mQ_s$, which is thus given by ΔQ_{coh}^x (i.e. the l.h.s). Calculation of the l.h.s is now straightforward (following Sect. 4.3): for a given impedance (and transverse damper), one only needs to calculate the complex mode frequency shift, in the absence of Landau damping. Without frequency spread, the condition for the beam to be stable is thus simply $\text{Im}(\Delta Q_{\text{coh}}^x) \geq 0$ (oscillations of the form $e^{j\omega t}$ are considered). Once its l.h.s is obtained, Eq. (4.36) can be used to determine the coherent betatron tune Q_c in the presence of Landau damping when the beam is at the edge of instability (i.e. Q_c real). However, the exact value of Q_c is not a very useful piece of information. The more useful question to ask is under what conditions the beam becomes unstable regardless of the exact value of Q_c under these conditions, and Eq. (4.36) can be used in a reversed manner to address this question. To do so, one considers the real parameter $Q_c - Q_0 - mQ_s$ (stability limit) and observes the locus traced out in the complex plane by the r.h.s of Eq. (4.36), as $Q_c - Q_0 - mQ_s$ is scanned from $-\infty$ to $+\infty$. This locus defines a “stability boundary diagram”. The l.h.s of Eq. (4.36), a complex quantity, is then plotted in this plane as a single point. If this point lies on the locus, it means the solution of Q_c for Eq. (4.36) is real, and this $Q_c - Q_0 - mQ_s$

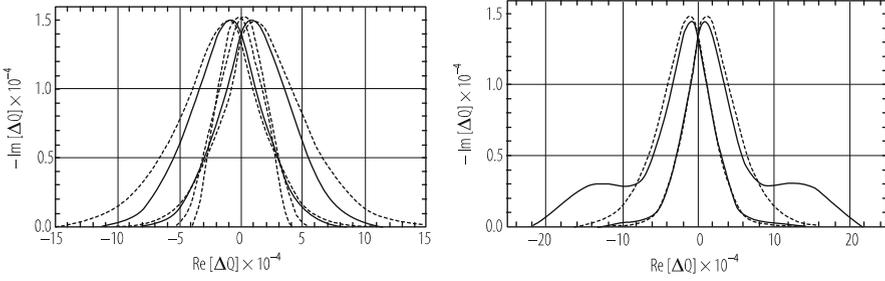


Fig. 4.17 Stability diagrams (for both positive and negative detunings a_0) for the LHC at top energy (7 TeV) with maximum available octupole strength: (Left) for the 2nd order (dashed curves), the 15th order (full curves), and the Gaussian (dotted curves) distribution; (Right) for the Gaussian distribution (dotted curve) and a distribution with more populated tails than the Gaussian (full curve)

is such that the beam is just at the edge of instability. If it lies on the inside of the locus (the side which contains the origin), the beam is stable. If it lies on the outside of the locus, the beam is unstable. The stability diagrams for the 2nd order, 15th order and Gaussian distribution functions are plotted in Fig. 4.17 for the case of the LHC at top energy (7 TeV) with maximum available octupole strength ($\varepsilon = 0.5$ nm, $|a_0| = 270440$ and $c = -0.65$).

The case of a distribution extending up to 6σ (as the 15th order distribution) but with more populated tails than the Gaussian distribution has also been considered and revealed a significant enhancement of the stable region compared to the Gaussian case ([93], see also Fig. 4.17(right)). This may be the case in reality in proton machines due to diffusive mechanisms.

It is worth reminding that Landau damping of coherent instabilities and maximization of the dynamic aperture are partly conflicting requirements. On the one hand, a spread of the betatron frequencies is needed for the stability of the beam coherent motion, which requires nonlinearities to be effective at small amplitude. On the other hand, the nonlinearities of the lattice must be minimized at large amplitude to guarantee the stability of the single-particle motion. A trade-off between Landau damping and dynamic aperture is therefore usually necessary [87].

Despite the destabilising effect of a resistive transverse damper in the case of a single bunch with zero chromaticity (as discussed in Sect. 4.3.2) below the TMCI intensity threshold (in the case of the “short-bunch” regime) without transverse damper, a transverse damper helps to reduce the amount of tune spread which would be needed to stabilise the bunch above the TMCI intensity threshold, as it can be seen in Fig. 4.18.

Note that linear coupling between the transverse planes can also influence the Landau damping mechanism [95], leading to a sharing of the Landau damping between the transverse planes, which can have a beneficial effect (i.e. stabilising the other plane, as it was used in the CERN PS for many years [96]) or a detrimental effect (i.e. destabilising one or two planes by loss of Landau damping, as it was

Fig. 4.18 Required normalised (to the synchrotron tune Q_s) tune spread Δq to stabilise the bunch in both cases of instabilities without and with Transverse Damper (TD), corresponding to the cases of Fig. 4.14 [94]. The normalised parameter x is proportional to the bunch intensity [63]

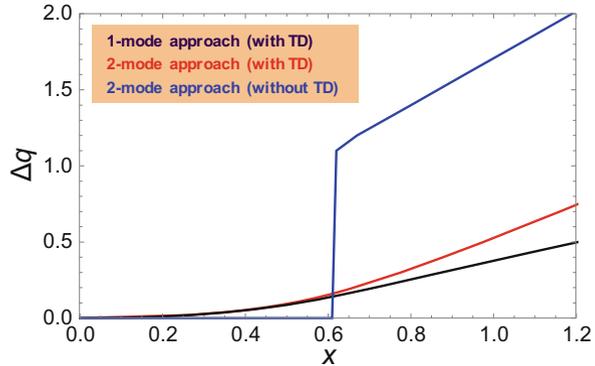
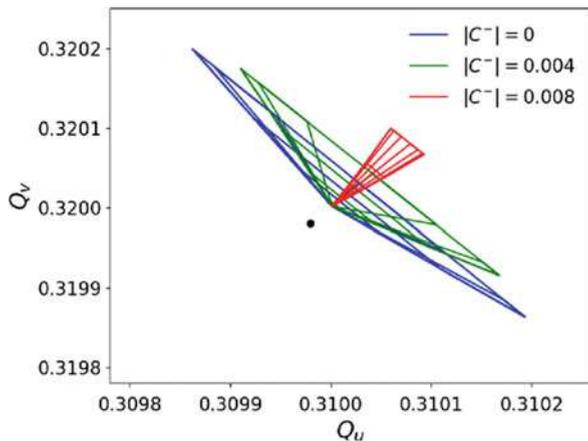


Fig. 4.19 Reduction of the tune footprint (and associated projections on the transverse tunes axes, responsible for Landau damping) vs. linear coupling (described here by the “closest tune approach” $|C^-|$) [98]. Courtesy of L.R. Carver



believed to be the case with the Batman instability of the HERA proton ring [97]: due to the features discussed in Ref. [97], the name “coupled head–tail instability” was suggested for this instability in the HERA proton ring). Recently, linear coupling was also observed to be detrimental in the CERN LHC [98], as revealed by both measurements and macroparticle simulations (see Fig. 4.19). This required a careful measurement and correction of linear coupling all along the LHC cycle to avoid to use much more Landau octupoles current than foreseen. One has also to remember that linear coupling modifies also the transverse emittances [99, 100].

In the case of additional space-charge nonlinearities, the stability diagram will be shifted and beam stability can be obtained or lost, depending on the coherent tune. The influence of space-charge nonlinearities on the Landau damping mechanism of transverse coherent instabilities has first been studied by Möhl and Schönauer for coasting and rigid bunched beams [101, 102]. It was studied in detail in the past years for higher-order head–tail modes from both theory [103–106] and numerical simulations [107].

The interplay between Landau octupoles and beam–beam long-range interactions can be either beneficial or detrimental depending on the sign of the Landau octupoles current (see Sect. 4.6) [108] and this effect has to be carefully taken into account in the CERN LHC to be able to push its performance.

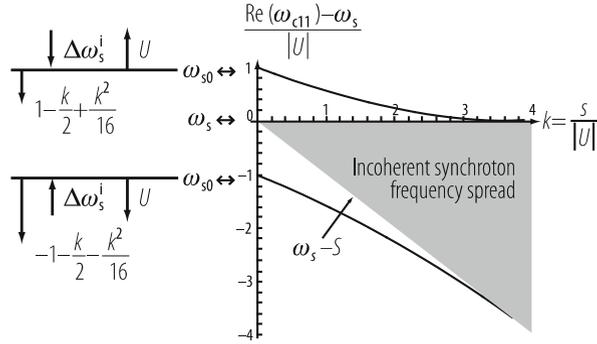
Another destabilising effect currently under investigation at the CERN LHC (and which could explain some long latencies observed in the past, of the order of few minutes or even tens of minutes) is the effect of noise, whose detrimental effect was predicted in 2012 [109] and confirmed experimentally in 2018 [110].

Some work is being done to try and use the nonlinear optics as a path to high intensity, providing “infinite (transverse) Landau damping” [111], or electron lenses [112] or Radio Frequency Quadrupoles (or similarly second order chromaticity) [113–115]. The latter two methods are believed to be more efficient than Landau octupoles at high energy due to the adiabatic damping and the associated significant reduction of the transverse beam sizes.

4.4.2 Longitudinal

When the bunch is very small inside the RF bucket, the motion of the particles is linear and all the particles have the same (unperturbed, maximum) synchrotron frequency ω_{s0} . By increasing the bunch length the incoherent synchrotron frequency spread S is increased (the maximum synchrotron frequency spread is obtained when the bunch length is equal to the RF bucket length as in this case the synchrotron frequency of the particles with the largest amplitude is equal to 0: the synchrotron frequency spread S is equal to ω_{s0} in this case). In the presence of an impedance, the coherent synchrotron frequency of the dipole mode ω_{c11} , which is equal to the low-intensity synchrotron frequency ω_{s0} without synchrotron frequency spread (due to the compensation between the incoherent and coherent tune shifts), moves closer and closer to the incoherent band (stable region). The two possible cases are represented in Fig. 4.20 (using the rigid-bunch approximation), which is similar to what was obtained by Besnier (who considered a parabolic distribution function, which introduces some pathologies in the stability diagram due to its sharp edge) ([116], and references therein): the case of a capacitive impedance below transition or inductive impedance above transition corresponds to $U > 0$ (the coherent synchrotron frequency shift of the dipole mode has been written $\Delta\omega_{c11} = U - jV$) and the incoherent synchrotron frequency shift (due to the potential-well distortion) is $\Delta\omega_s^i < 0$ (and thus $\omega_s < \omega_{s0}$), and the case of a capacitive impedance above transition or inductive impedance below transition corresponds to $U < 0$ and $\Delta\omega_s^i > 0$ (and thus $\omega_s > \omega_{s0}$). Motions $\propto e^{j\omega t}$ are considered, which means that the beam is unstable when $V > 0$ (V is called the instability growth rate). The usual case where the resistive part of the impedance is small compared to the imaginary part is assumed, i.e. $V \ll |U|$. Beam stability is obtained when ω_{c11} enters into the incoherent band. In both cases, the stability limit is reached for $k = 4$, i.e. $S = 4|U|$, which is Sacherer’s stability criterion for the dipole mode.

Fig. 4.20 Evolution of the coherent synchrotron frequency for the dipole mode with respect to the incoherent frequency spread (using the rigid-bunch approximation)



4.5 Two-Stream Effects (Electron Cloud and Ions)

G. Rumolo

4.5.1 Electron Cloud Build-Up in Positron/Hadron Machines

The term “electron cloud” is used for describing an accumulation of electrons inside the beam chamber of a circular accelerator, in which bunched beams of positively charged particles are accelerated or stored. The electron cloud can affect the accelerator operation by causing beam tune shift, emittance growth and coherent instabilities, as well as increase of the vacuum pressure and interference with beam diagnostics devices. Most of these effects eventually lead to beam quality degradation and loss. Electrons can be initially produced in the vacuum chamber by a number of processes. These electrons are called primary because, although some times their number can be sufficiently high to affect the circulating beam, they are usually only the seed for an avalanche process (see Fig. 4.21). In general, the primary electrons rapidly multiply via a beam-induced multipacting mechanism, which involves acceleration of the electrons in the beam field and secondary emission from their impact on the chamber wall. In the following, we first give an overview on the electron cloud formation (or build-up) process. We therefore list the main primary generation mechanisms and then discuss how the thus generated electrons can multipact in the presence of a train of bunches. The other important stages of the build-up of an electron cloud are its equilibrium and successive decay in the gap behind a bunch train. In the second part, we briefly discuss dynamics and consequences of the beam instabilities caused by the electron cloud that has formed in a beam pipe. In conclusion, we will try to give an up-to-date list of the possible techniques for electron cloud mitigation or suppression.

Primary electrons are the electrons generated during the passage of a bunch. They can be photo-electrons from synchrotron radiation in bending regions (mainly for positron beams) or secondary electrons desorbed from beam particles lost at the

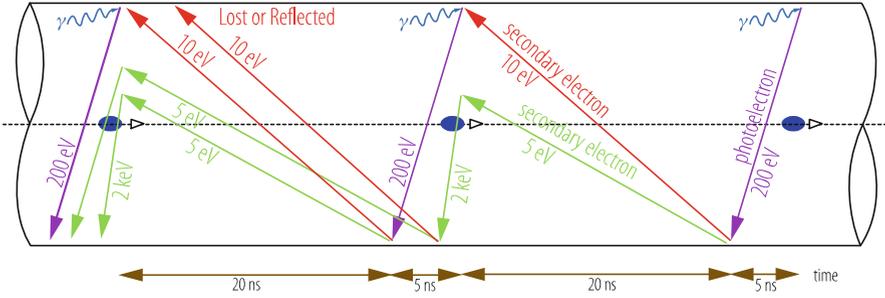


Fig. 4.21 Schematic of electron cloud build-up in the LHC beam pipe during multiple bunch passages, via photo-emission (due to synchrotron radiation) and secondary emission (Courtesy of F. Ruggiero). Note that the LHC is the 1st proton storage ring for which synchrotron radiation becomes a noticeable effect

walls (especially for ion beams). In this case they are emitted from the chamber wall. Primary electrons can also be created within the volume swept by the beam if the production mechanism is ionization of the residual gas. The location where the electrons are created can determine the energy gain of the primary electrons in the beam potential. The number of electrons created per unit length by synchrotron radiation or by beam loss during one bunch passage can be comparable to the average line density of beam particles, in which case these processes can alone give rise to amounts of electrons critical for the beam stability. The rate of photoemission (number of photoelectrons created per unit length) can be estimated as the product of the photo-electron yield Y_γ by the photoemission rate dN_γ/ds :

$$\frac{dN_{e\gamma}}{ds} = Y_\gamma \frac{dN_\gamma}{ds} = Y_\gamma \frac{5\alpha\gamma}{2\sqrt{3}\rho}, \quad (4.39)$$

where α denotes the fine structure constant and ρ the curvature radius of the beam in the dipole. For many materials, the photo-emission yield can be correctly approximated as being about 0.1 over a fairly large photon energy range, e.g. between a few eV and a few tens of keV. The azimuthal distribution of absorbed photons around the chamber wall and, thus, the launch positions of the emitted primary photo-electrons depend on the reflective properties of the chamber wall. The first simulation of an electron cloud build-up for short bunches was written by K. Ohmi. It served to explain coupled-bunch instabilities observed with positron beams at the KEK photon factory [117]. Ohmi's pioneering study considered only photo-emission at the chamber wall as a source of electrons, though a little later his initial code was extended to include secondary emission by electrons as well.

Ionization from scattering of individual charged beam particles against molecules of the residual gas occurs with typical cross sections of 1–2 Mbarn for most of the gas species that can be found in a beam chamber. However, a lower cross section of about 0.2 Mbarn applies to the lighter species, like H_2 [118]. These numbers

refer to singly charged particles at ultrarelativistic energies. For fully ionized atoms, the cross section scales roughly with the square of the atomic number, i.e. like Z^2 . Additionally, it increases by several orders of magnitude towards lower beam energies. If the beam density is sufficiently high, as it will be in certain sections of the next generation of linear colliders, ionization by the collective electric field of the bunch replaces single-particle scattering ionization as the dominant ionization process [119]. When this happens, the beam completely and instantly ionizes the residual gas in its neighbourhood.

Protons or ions impacting on the wall can be responsible for the generation of a large number of electrons. The secondary-electron yield from ion impact is approximately proportional to the projectile stopping power and inversely proportional to the cosine of the angle of incidence [120]. Since the stopping power is in turn proportional to the square of the charge number divided by the mass number, this value, usually very high because of the shallow angles at which losses typically occur, can be further amplified by one or two orders of magnitude for heavy ion beams. Production of this type of electrons also occurs with a high rate at the collimators, where significant beam loss routinely occurs by design.

The mechanism responsible for an exponential growth of the number of electrons is beam induced multipacting. The primary electrons are accelerated by the electric field of a passing bunch to such high energies that they produce, on average, more than one secondary electron when they again hit the wall of the vacuum chamber. The Secondary Emission Yield (SEY) of the chamber material is by definition the number of secondary electrons produced on average by an electron impact. It is obviously a function of the impinging electron energy, its angle of incidence, and the chamber history. For a round chamber of radius h and a short bunch, the resonance condition for beam-induced multipacting from electrons produced at the pipe walls takes the simple form [121]

$$N_b r_e L_b = h^2, \quad (4.40)$$

where r_e denotes the classical electron radius and L_b the bunch spacing in units of length. However, the condition of Eq. (4.40) is by far too stringent. Most secondary electrons have low energy and tend to stay in the vacuum chamber for a long time after a bunch passage. The survival of low energy electrons is made even more likely by the fact that their probability of being elastically backscattered at the chamber walls is almost one [122]. Moreover, a different multipacting regime, called “trailing edge multipacting”, also exists in presence of long bunches. The electrons produced on the falling edge of the bunch can gain energy as they cross the pipe section and cause multipacting just during the passage of the second part of the bunch. This is a special type of single-bunch multipacting, causing electron clouds that can be either significantly cleared before the arrival of the next long bunch, or accumulate further in a mixed single-bunch and multi-bunch process.

Many simulation codes have been developed over the years to study numerically the process of electron cloud build-up and explore different beam/machine parameter ranges (e.g. [123–125]). Due to the variety of possible regimes and processes

involved, simple considerations have usually turned out to be insufficient to predict the build-up thresholds and the change in electron cloud caused by some specific parameter change. It was a gratifying confirmation of the predictive power of the early simulations that the expected electron clouds were indeed observed at both B factories (PEP-II and KEKB), at the Large Hadron Collider injectors PS and SPS when operated with LHC-type beams, and finally in the LHC itself when operation with trains of closely spaced bunches started. In these rings, the electron cloud was seen to cause tune shift and emittance growth along the bunches of a train, both coupled and single-bunch instabilities and a degradation of certain beam diagnostics signals [126–128]. In the SPS and PS, significant beam loss could be observed at the end of a train if no countermeasures were put in place. As the electron cloud is potentially one of the main bottlenecks of the SPS after the upgrade of the LHC injector chain in the framework of the LHC Injectors Upgrade (LIU) project [129], dedicated electron diagnostics devices are installed in the machine to measure the electron flux and spatial electron distribution on surfaces with different coatings, as well as its variation with the time of exposure, i.e. what we call “beam-induced machine scrubbing” [129]. Since 2011, the electron cloud has been also observed routinely in the LHC, causing beam instability and emittance growth at the end of the multi-bunch trains but also additional heat load on the cold beam screens of the arcs as well as of the matching and final focusing quadrupoles [130]. While the effects linked to electron cloud have quickly disappeared for beams with 50 ns bunch spacing thanks to a relatively rapid beam induced machine scrubbing, running with 25 ns beams has proved to be rather challenging in this machine. With this type of beams, even after extensive machine scrubbing, the undesired effects of the electron cloud have remained visible on the beam and the machine equipment. In particular, the 25 ns beam needs to be stabilised with high values of chromaticity in both planes and large octupole settings. Besides, the large heat load on the cold beam screens still remains very close to the capacity of the cryogenic system in nominal operating conditions [131, 132]. While these effects have not prevented running LHC close to the nominal conditions from 2015 to 2018, they could still be a significant showstopper for future operation with double beam current in the High Luminosity LHC era [133]. Concerning lepton machines, the electron cloud is typically associated with a reduction of specific luminosity in e - p colliders and is expected to be one of the main limiting factors for the damping rings of future linear collider projects. The Cornell Electron Storage Ring (CESR) was reconfigured in 2008 as a Test Accelerator (CesrTA) for a program of electron cloud research with lepton beams. With its new local diagnostics for measurement of cloud density and improved instrumentation for the characterization of the beam dynamics of high intensity bunch trains interacting with the cloud, this test facility provided for many years both a benchmark case for the existing simulation codes and testing the effectiveness of several types of countermeasures [134]. Since the processes of secondary electron emission and elastic reflection at the walls play a fundamental role in causing beam induced multipacting, we now shortly describe their key parameters. The true secondary yield for perpendicular incidence, δ , can

be expressed by a universal function [135]

$$\delta(x) = \delta_{\max} \frac{sx}{s-1+x^2}, \quad (4.41)$$

where $x = E/E_{\max}$, with E the energy of the incident electron and E_{\max} the energy at which the yield assumes the maximum value, and s is a fit parameter that was measured to be about equal to 1.35 for LHC Cu samples [135]. The two variable parameters in Eq. (4.41) are δ_{\max} , the maximum yield, which typically assumes values between about 1.0 and 3.0 for conductive materials (but it can be higher for dielectrics), and E_{\max} . For non-normal incidence of the primary electron these two parameters are usually both increased by a factor depending on the cosine of the incidence angle [136]. Elastic reflection of electrons is mostly important at low energies, i.e. below about 20 eV. The measured electron reflection probability [122] can be parametrized as

$$\delta_{\text{el}}(E) = \left(\frac{\sqrt{E} - \sqrt{E + E_0}}{\sqrt{E} + \sqrt{E + E_0}} \right)^2, \quad (4.42)$$

with only one fit parameter, E_0 . Equation (4.42) implies that the reflection probability approaches one in the limit of vanishing electron energy, even if presently no general consensus has been reached around this point, which is still very controversial, as it is extremely difficult to measure the secondary emission yield at very low energy.

The electron cloud build-up saturates when the electron losses balance the electron generation rate. This can happen either at low bunch charges, when the average neutralization density is reached, or at high bunch currents, when the electrons rapidly accumulate until the kinetic energy of the newly emitted ones becomes too low to let them penetrate into the space charge field of the cloud. Simulations have demonstrated a complex behaviour of the electron cloud equilibrium, which strongly depends on the combination between bunch length, charge, spacing and on the chamber radius. First of all, the saturation phase is generally characterized by an oscillating behaviour of the electron cloud density over the bunch spacing and the amplitude of this oscillation can be very large. Furthermore, in some cases the steady-state value of the electron cloud density has been found not to be monotonically increasing with the bunch intensity. For instance, a beam with 50 ns spaced bunches in the SPS is predicted to hit its highest electron cloud equilibrium density for bunch populations of about 10^{11} p, while this value decreases both for lower and higher intensities.

The electron density decays after the passage of a bunch train (or in the gap between bunch trains) and two different regimes can be distinguished during this phase. In the first one, right after the train passage, the cloud decays quickly due to the space charge effects and the reminiscent energy distribution from the last bunch passage. In the second one, only low energy electrons will be left,

which move slowly and exhibit a dissipation rate depending on their probability of being elastically reflected at the chamber surface. Due to the second part of the decay evolution, the electron clearing time between bunch trains can become painfully long. Three effects are suspected to be responsible for the long memory and lifetime of the electron cloud. First, nonuniform fields, such as quadrupoles or sextupoles, may act as magnetic bottles and trap electrons for an indefinite time period [125, 137]. Second, if the probability of elastic reflection really approaches one in the limit of zero electron energy, as is suggested by measurements [122], low-energetic electrons could survive nearly forever, bouncing back and forth between the chamber walls, independently of the magnetic field. Third, slow ions produced by residual gas ionization have been also suspected to be long lived in the beam chamber and, therefore, possibly cause (or help) electron trapping and long survival time.

4.5.2 The Electron Cloud Instability

When a positron/hadron beam interacts electromagnetically with the electron cloud that has formed in the beam chamber, a coherent oscillation of both electrons and beam particles can grow from any small initial perturbation of the beam distribution, e.g. from the statistical fluctuations due to the finite number of beam particles. This instability can be considered as a two-stream instability of the same type as studied in plasma physics. Such instabilities can be very fast, since in the new generation of high intensity rings operating with many closely spaced bunches, the density of electrons can become quickly very large. Even machines operating with bunches spaced by hundreds of ns can actually suffer from electron cloud, because of the long survival time of low energy electrons in the beam pipe. Electron clouds can cause single-bunch instabilities as well as multi-bunch dipole mode instabilities. The multi-bunch instability appears when the electron cloud can carry a sufficiently long memory as to couple subsequent bunches. The single-bunch phenomenon, instead, is driven by a pinched electron cloud, which, over one single passage of the bunch through it, is able to transfer information from an offset bunch head to the bunch tail. Obviously, although this second type of instability is caused by a single-bunch mechanism, it can only occur in multi-bunch operation, since the electron cloud requires a train of several bunches to build up. For single-bunch instabilities caused by multi-bunch built electron clouds, electrons usually only perform a low number of oscillations while the bunch is passing (typically between fractions of unit and few units), and the bunch effectively interacts with a pre-existing cloud produced by the preceding bunches and filling almost uniformly the beam pipe prior to the bunch arrival. The number of electrons does not change appreciably during one bunch passage. In reality, another possible head tail effect resulting into a different type of two-stream instability was observed in some machines operating with long bunches. In this case, the instability is intimately related to an electron cloud from “trailing-edge multipacting”, described in the previous section. The electrons

multiplication happening over the falling edge of the bunch can reach levels as to render the beam unstable. Even coasting beams are not immune from electron cloud problems. The electrons produced from residual gas ionization remain trapped in the transverse potential of the uniform beam and tend to accumulate to very high central density values. The electrons created at the chamber walls (e.g. from beam loss) are accelerated and decelerated in the beam field and eventually hit the chamber with the same energy with which they were emitted. Multipacting can play here a role, since electrons can gain energy and create secondaries when hitting the wall, if the beam line density is perturbed. The interaction of the coasting beam with the electrons can make noise evolve into an unstable coupled oscillation, called e-p instability, which was widely studied already at the beginning of the 70s [138, 139]. While the single-bunch instability described earlier in this section can be treated separately from the build-up of the electron cloud that causes it, in all other cases the two processes are coupled together and need to be solved with a joint model.

Electron cloud instabilities for short bunches have been observed in form of emittance growth and beam loss at the KEKB LER, at the CERN PS, SPS and LHC, and at the PEP-II LER. At the KEKB LER a blow-up of the vertical beam size was already observed at the early commissioning time [140]. This blow-up was not accompanied by any coherent beam motion, which could be easily suppressed by transverse feedback and chromaticity, and the blow-up was only seen in multi-bunch operation with a narrow bunch spacing. The single-bunch two-stream instability provided a plausible explanation of the observed beam blow-up [141]. This explanation has since been reinforced by the simultaneous observation of a tune shift along the bunch train, which appears for the same bunches exhibiting vertical size blow-up. Also the experimental evidence that the installation of solenoids around the ring could increase the instability threshold in regular operation shows the relation between electron cloud and the instability. At the CERN SPS the electron cloud has been observed since the ring has been regularly operated with LHC-type bunch trains [142] and it has been held responsible for strong transverse instabilities. In the horizontal plane a low order coupled bunch instability develops within a few tens of turns after injection. In the vertical plane, a single-bunch head tail instability rises on a much shorter time. The reason of the different behavior in the two transverse planes is ascribed to the confinement of the electron cloud mostly in dipole regions, which can limit the intra-bunch electron pinching in the horizontal plane and therefore inhibit the single-bunch mechanism for instability. The horizontal (coupled-bunch) instability is cured by means of a transverse feedback system. Similar to the situation at the KEKB LER, running the SPS at high positive chromaticity can cure the vertical instability [143]. Upstream from the SPS, when the nominal LHC beam was generated by the PS machine, one of the standard signatures of the electron cloud was observed shortly before extraction: a baseline drift in electrostatic devices. However, the beam resided too shortly in the machine to become unstable, even if a dedicated experiment proved the onset of an electron cloud instability on the short bunches, if they are kept in the machine for a sufficiently long time. In the LHC, transverse beam instabilities affecting the last bunches of long 25 ns trains in both transverse planes have been

systematically observed at injection [144]. The reason why the instability appears in both planes is that the integrated central density of electrons causing the instability comes mainly from the electron cloud in quadrupole magnets and therefore affects equally both planes. This instability is controlled by means of high chromaticity and high octupole strength. Besides, electron cloud instabilities have been observed also at high energy (6.5 TeV) but mainly in the vertical plane and for values of bunch currents lower than nominal. This has been explained as due to the onset of a central stripe in all the dipoles (appearing when the bunch intensity decays), which leads to an integrated electron density capable of making a 6.5 TeV beam unstable [145]. Concerning long bunches, a great deal of evidence indicates that the primary instability limiting the performance of the LANL-PSR is an e-p instability [146]. Growth of the electron cloud results from multipacting of the electrons on the walls of the vacuum chamber during passage of the trailing edge of the proton beam, when the electrons can receive a net acceleration toward the wall. The instability was controlled by various measures to enhance Landau damping and transverse feedback. In coasting beams, an e-p instability was first observed in the LBNL-Bevatron [147] and CERN-ISR [139]. While in the Bevatron this instability was combated with active feedback and beam bunching, in the ISR additional pumping was installed to improve the vacuum from 0.1 to 0.01 nTorr and the number of clearing electrodes was increased to sweep away the electrons.

Several analytical approaches have been used to study the electron cloud instability, including few-particles models and an attempt to apply the TMCI theory to the electron cloud wake field, modeled as a broadband resonator. However the most widespread and comprehensive approach makes use of particle tracking simulations with localized electron cloud kicks. The numerical modeling of the interaction between an electron cloud and a particle bunch is discussed in Sect. 4.7. Simulation codes have had the merit to reveal interesting features of the electron cloud instability, which distinguish it from other types of conventional instabilities. For example, the electron cloud wake field is not only a function of the distance between source and probe particles, but it depends on the locations of the two separately. This translates into an impedance with a double frequency dependence [148]. Another interesting finding was that, for constant beam emittances (transverse and longitudinal) and constant bunch length, the electron cloud instability threshold decreases with the beam energy [149]. The reason of this anomalous behaviour is that, although the beam becomes stiffer at higher energies, its transverse sizes become smaller and the pinching effect on the electron cloud is amplified.

Concerning the multi-bunch instability, the usual approach is to calculate the bunch-to-bunch wake field with an electron cloud build-up code (which correctly models the electron cloud dynamics in the space between two bunches) and then apply the multi-bunch analytical formula to assess the threshold for the beam stability. Simulation codes with bunches modeled as single macroparticles, valid for machines operating with short bunches, have been also developed to study numerically the multi-bunch instabilities due to electron cloud in a more self-consistent manner.

A number of simulation tools have been developed over the years in order to study the electron cloud single-bunch instability for short bunches via direct particle tracking. The simulations of electron cloud build-up are generally treated separately, since they make use of a weak-strong approach, in which the beam is rigid and is approximated by bunches with static transverse and longitudinal Gaussian distributions, while the electrons are macroparticles. Build-up simulations need to be run prior to the instability simulations, because they provide the necessary input on the transverse distribution of the electron cloud density at saturation just before the arrival of a bunch. A variety of simulation codes are presently available for this purpose [150]. Fully self-consistent computations, in which the cloud generation over a bunch train around the ring as well as the resulting bunch instabilities, are treated by a single program are still under development. The existing simulation programs to study the electron cloud instabilities model the interaction of a single bunch with an electron cloud on successive turns. The cloud is always assumed to be generated by the preceding bunches, and can be considered initially uniform or the distribution is imported from a build-up code. The electrons give rise to a head-to-tail wake field, which amplifies any initial small deformation in the bunch offset, e.g. due to the finite number of macroparticles in the simulation. All simulation tools that have been developed for this study are essentially of the strong-strong type, since the purpose is to investigate how the bunch particles are affected by the electron cloud via the continuous interaction. In particular, electrons are always modeled as macroparticles either concentrated at one or several locations along the ring or uniformly smeared along the axis of the machine. The bunch consists of macroparticles or of microbunches with a fixed transverse size. The bunch is then subdivided into slices, which interact in sequence with the electrons of the cloud, creating the distortion of the initially uniform cloud distribution that can affect the body and tail of the bunch. The electric fields of the electrons and of the beam acting mutually on each other are calculated by means of a Particle-In-Cell (PIC) algorithm. The transformation of the 6D phase-space vectors of the beam particles between two kick points is achieved using the appropriate transport matrices or nonlinear tracking. The field of the electron cloud acting on itself can be optionally included, but in general does not seem to play a significant role in this type of mechanisms and hence it is neglected. For the purpose of studying the interplay of electron cloud instability with other mechanisms, the simulation codes contain synchrotron motion, chromaticity and usually additional options to model the action of an independent impedance source beside the electron cloud, as well as space charge and detuning with amplitude.

4.5.3 Mitigation and Suppression

There are at least three possible actions to reduce, or even suppress, the electron cloud: (i) reducing the production rate of primary electrons or confining their motion to a region where they are not likely to do any harm; (ii) eliminating the possibility

of multipacting by lowering the SEY via surface treatment; (iii) alleviating the effect on the beam or on the diagnostics. In most cases a combined approach is desirable, that's why most machines affected by electron cloud problems have usually chosen to implement more than one of these mitigating techniques. The primary production of photoelectrons needs to be reduced, because it may be so high that the electron cloud could reach saturation within a few bunch passages even without any multipacting. This is the case at KEKB, the photon factory, if no countermeasures were taken. The obvious solution is an antechamber to intercept most of the synchrotron radiation, or also photon absorbers (as those designed for the CLIC Damping Rings). For dipole fields, a saw-tooth pattern impressed on the chamber wall, as was implemented for the LHC (actually on the beam screen that forms the inner part of the chamber and serves to protect the cold bore of the magnets from synchrotron radiation), is used for effectively reducing the photon reflectivity thanks to the perpendicular impact. Weak solenoids of the order of 50 G are a possibility in field-free regions, which was successfully implemented in the straight sections of KEKB and in RHIC. The solenoids do not really affect the photoemission process, but they keep the photoelectrons close to the wall and, thus, strongly mitigate the subsequent beam–electron interaction. Since the gas ionization rate is linearly proportional to the vacuum pressure in the beam chamber, the number of electrons created by gas ionization can only be reduced by significant factors improving the vacuum. If field ionization is important, however, a possible cure would be lengthening the bunches, though this will mainly be a concern for future projects such as linear colliders or X-ray FELs operating with positrons. Electrons generated by beam loss can be controlled if the localization of the losses is known with good precision. For example, electrons produced by the beam losses at a collimator can be controlled by solenoids or clearing electrodes. A large number of electrons is also generated at the injection stripping foils, for accelerators employing charge-exchange injection. At the SNS, a 10-kV clearing voltage is applied to channel the electrons liberated at the stripping foil onto a collector plate that is monitored by a TV camera, while solenoids are used along the collimator straights.

The reduction of the SEY of the inner wall of the beam chamber can be achieved in different manners. First of all, a serendipitous feature of the electron cloud build up in an accelerator's chamber is that, while the electrons hit the beam chamber with high energy and multiply, they also 'scrub' the surface by first removing layers of impurities responsible for high SEY values and eventually graphitising the surface with a further reduction of the SEY from that of the pure metal [151]. This means that, if a method is found to run an accelerator with electron cloud and stable beam, e.g. by stabilising the beam against the electron cloud through appropriate machine settings, it will be the electron cloud itself to gradually lower the SEY of the inner wall of the chamber and eventually turn itself off. This obviously relies on that the final SEY reachable through scrubbing is below the value that sets off the electron cloud build up in the operational configuration. Besides, it may take a significant amount of time to reach this condition, because the electron flux is decreasing while scrubbing and the electron doses required to perform further SEY reduction steps are

also exponentially increasing when moving to SEYs below 1.3–1.4. The technique described here is what we call ‘beam induced machine scrubbing’ and machine like the SPS and LHC fully rely on it to run successfully with 25 ns spaced beams. While beam induced scrubbing is an important option for already built machines, coatings with intrinsically low SEY materials can be envisaged at the design stage to limit the creation of an electron cloud in future machines. A well established method to reduce multipacting is coating with TiN, a material whose secondary emission yield becomes quickly low after some conditioning (through illumination under synchrotron light). The thickness of the coating must be of the order of a μm , such as not to alter the resistive impedance seen by the beam. A more favorable getter material made from TiZrV, called Non-Evaporable Getter (NEG), was developed at CERN. This getter material is characterized by its greater structural stability than TiN, its pumping capability and its low activation temperature. The warm sections of the LHC, about 10% of the circumference, have been coated with NEG. The NEG coating was also tested at several light source insertion devices, where circumstantial evidence suggests an increase in the effective impedance, presumably due to a larger surface roughness and low conductivity. The additional contributions to the ring impedance from the surface roughness and low conductivity impedance of the coating layer is of no concern for the longer proton bunches in the LHC, but could significantly affect the stability of the short positron beams in the Damping Rings of a future linear collider. From 2007 on, new efforts have been put on the search for coating materials that do not require high temperature activation and do not suffer from aging. In particular, amorphous carbon (a-C) thin films, deposited with d.c. magnetron sputtering, have shown to possess all these qualities. Besides, their maximum secondary emission yields, measured in the lab, reach values even below one and the films are also stable against mechanical stress. Testing of a-C coating in accelerator environments (SPS and Cern-TA) has demonstrated all these features. Another method to reduce the secondary emission yield of a surface is to use a naturally rough material. Here the SEY reduction is a geometrical effect due to the high probability of quick re-absorption of the electrons emitted with low energy.

Multipacting can also be suppressed by solenoids, though one should pay attention to the possibility of exciting undesired cyclotron resonances. Electric clearing fields are an efficient cure, as shown both in simulations and measurements of electron cloud in the CERN PS. They were already used to cure electron-proton instabilities for the coasting proton beams in the CERN ISR during the early 70s. At the SNS operating with long proton bunches all BPMs can be biased with a clearing voltage of 1 kV. To be effective for the multipacting experienced by short bunches with close spacing, the clearing electrodes must be mounted all around the ring, in distances of a few tens of cm and voltages of the order 1 kV are probably required. The impedance introduced by many such devices could be prohibitive, as it appeared to be the case in the DAΦNE positron ring with the very first clearing electrode design. Other options for a practical implementation of electric clearing fields may be splitting the beam pipe into a top and bottom half, isolated from each other and held at different potential. Biasing the two jaws of a collimator

against each other is a similar idea. Recently, there is a growing interest towards the suppression of multipacting by means of grooves on the chamber wall. This technique, first tested in simulations, has proven to be efficient in KEKB and Cesr-TA. Similarly to a rough surface, but in a controlled way and on a macroscopic scale, these grooves essentially act as electron traps. Angle and depth of the grooves are key parameters and specifications are different in dipole or field-free regions. Proper tailoring of the bunch filling patterns (bunch spacing, bunch trains and bunch charges) is yet another way of achieving an acceptable electron density. Examples include the actual bunch spacing chosen for PEP-II and KEKB operation, which are twice or three times the design spacing, and satellite bunches proposed for the LHC [152]. Gaps within or between trains can lower the density and reset the cloud at least to some extent. Extensive studies of the electron cloud formation as a function of the bunch filling patterns were also carried out at RHIC, in which the optimization could be found using the maps approach to quickly scan the build-up for different configurations.

The electron cloud causes a large variety of undesired effects. Common stabilising measures can be taken against the resulting instabilities, which include transverse bunch-to-bunch feedback, increased chromaticity, Landau-damping octupoles, intra-bunch head-tail feedback, and linear coupling. All these measures are anyway necessary when a machine is operating in beam-induced scrubbing mode. Degradation of diagnostics signals due to impacting electrons can be also overcome with local solenoid windings.

4.6 Beam-Beam Effects

W. Herr

4.6.1 Introduction

The problem of the beam-beam interaction is the subject of many studies since the introduction of the first particle colliders [153]. It has been and will be one of the most important limits to the performance and therefore attracts the interest at the design stage of a new colliding beams facility. A particle beam is a collection of a large number of charges and represents an electromagnetic potential for other charges. It will therefore exert forces on itself and other beams. The forces are most important for high density beams, i.e. high intensity and small beam sizes, which are the key to high luminosity.

The electromagnetic forces from particle beams are very non-linear and result in a wide spectrum of consequences for the beam dynamics. Furthermore, as a result of the interaction, the charge distribution creating the disturbing fields can change as well. This has to be taken into account in the evaluation of beam-

beam effects and in general a self-consistent treatment is required. Although we now have a good qualitative understanding of the various phenomena, a complete theory does not exist and exact predictions are still difficult. Numerical techniques such as computer simulations have been used with great success to improve the picture on some aspects of the beam–beam interaction while for other problems the available models are not fully satisfactory in their predictive power [154].

4.6.2 Beam–Beam Force

In the rest frame of a beam we have only electrostatic fields and to find the forces on other moving charges, we have to transform the fields into the moving frame and to calculate the Lorentz forces (see [153, 155–160] and references therein).

The fields are obtained by integrating over the charge distributions. The forces can be defocusing or focusing since the test particle can have the same or opposite charge with respect to the beam producing the forces.

The distribution of particles producing the fields can follow various functions, leading to different fields and forces. It is not always possible to integrate the distribution to arrive at an analytical expression for the forces in which case either an approximation or numerical methods have to be used. This is in particular true for hadron beams, which usually do not experience significant synchrotron radiation and damping. For $e^- e^+$ colliders the distribution functions are most likely Gaussian with truncated tails.

In the two-dimensional case of a beam with bi-Gaussian beam density distributions in the transverse planes, i.e. $\rho(x, y) = \rho_x(x) \rho_y(y)$ with r.m.s. of σ_x and σ_y

$$\rho_u(u) = \frac{1}{\sigma_u \sqrt{2\pi}} \exp\left(-\frac{u^2}{2\sigma_u^2}\right) \text{ where } u = x, y \quad (4.43)$$

one can give the two-dimensional potential $U(x, y, \sigma_x, \sigma_y)$ as a closed expression

$$U(x, y, \sigma_x, \sigma_y) = \frac{ne}{4\pi\epsilon_0} \int_0^\infty \frac{\exp\left(-\frac{x^2}{2\sigma_x^2+q} - \frac{y^2}{2\sigma_y^2+q}\right)}{\sqrt{(2\sigma_x^2+q)(2\sigma_y^2+q)}} dq \quad (4.44)$$

where n is the line density of particles in the beam, e is the elementary charge and ϵ_0 the permittivity of free space [159]. From the potential one can derive the transverse fields \vec{E} by taking the gradient $\vec{E} = -\nabla U(x, y, \sigma_x, \sigma_y)$.

4.6.2.1 Elliptical Beams

For the above case of bi-Gaussian distributions (i.e. elliptical beams with $\sigma_x \neq \sigma_y$) the fields can be derived and for the case of $\sigma_x > \sigma_y$ we have [160]

$$E_x = \frac{ne}{2\epsilon_0 \sqrt{2\pi(\sigma_x^2 - \sigma_y^2)}} \text{Im} \times \left[\text{erf} \left(\frac{x + iy}{\sqrt{2(\sigma_x^2 - \sigma_y^2)}} \right) - \exp \left(-\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2} \right) \text{erf} \left(\frac{x \frac{\sigma_y}{\sigma_x} + iy \frac{\sigma_x}{\sigma_y}}{\sqrt{2(\sigma_x^2 - \sigma_y^2)}} \right) \right] \quad (4.45)$$

$$E_y = \frac{ne}{2\epsilon_0 \sqrt{2\pi(\sigma_x^2 - \sigma_y^2)}} \text{Re} \times \left[\text{erf} \left(\frac{x + iy}{\sqrt{2(\sigma_x^2 - \sigma_y^2)}} \right) - \exp \left(-\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2} \right) \text{erf} \left(\frac{x \frac{\sigma_y}{\sigma_x} + iy \frac{\sigma_x}{\sigma_y}}{\sqrt{2(\sigma_x^2 - \sigma_y^2)}} \right) \right] \quad (4.46)$$

The function $\text{erf}(t)$ is the complex error function

$$\text{erf}(t) = \exp(-t^2) \left[1 + \frac{2i}{\sqrt{\pi}} \int_0^t \exp(z^2) dz \right] \quad (4.47)$$

The magnetic field components follow from

$$B_y = -\beta \frac{E_x}{c} \text{ and } B_x = \beta \frac{E_y}{c} \quad (4.48)$$

The Lorentz force acting on a particle with charge q is finally

$$\vec{F} = q \left(\vec{E} + \vec{v} \times \vec{B} \right) \quad (4.49)$$

4.6.2.2 Round Beams

With the simplifying assumption of round beams ($\sigma_x = \sigma_y = \sigma$), one can re-write (4.49) in cylindrical coordinates

$$\vec{F} = q \left(E_r + \beta c B_\phi \right) \times \vec{r} \quad (4.50)$$

From (4.44) and with $r^2 = x^2 + y^2$ one can immediately write the fields from (4.50) as

$$E_r = -\frac{ne}{4\pi\epsilon_0} \frac{\delta}{\delta_r} \int_0^\infty \frac{\exp\left(-\frac{r^2}{2\sigma^2+q}\right)}{2\sigma^2+q} dq \quad (4.51)$$

and

$$B_\phi = -\frac{ne\beta c\mu_0}{4\pi} \frac{\delta}{\delta_r} \int_0^\infty \frac{\exp\left(-\frac{r^2}{2\sigma^2+q}\right)}{2\sigma^2+q} dq \quad (4.52)$$

δ stands for derivative (Eqs. 4.51 and 4.52) and μ_0 is the permeability of free space (Eq. 4.52).

We find from (4.51) and (4.52) that the force (4.50) has only a radial component. The expressions (4.51) and (4.52) can easily be evaluated when the derivative is done first and $1/(2\sigma^2 + q)$ is used as integration variable. We can now express the radial force in a closed form (using $\epsilon_0\mu_0 = c^{-2}$)

$$F_r(r) = -\frac{ne^2(1+\beta^2)}{2\pi\epsilon_0} \frac{1}{r} \left[1 - \exp\left(-\frac{r^2}{2\sigma^2}\right) \right] \quad (4.53)$$

and for the Cartesian components in the two transverse planes we get

$$F_x(r) = -\frac{ne^2(1+\beta^2)}{2\pi\epsilon_0} \frac{x}{r^2} \left[1 - \exp\left(-\frac{r^2}{2\sigma^2}\right) \right] \quad (4.54)$$

and

$$F_y(r) = -\frac{ne^2(1+\beta^2)}{2\pi\epsilon_0} \frac{y}{r^2} \left[1 - \exp\left(-\frac{r^2}{2\sigma^2}\right) \right] \quad (4.55)$$

The forces (4.54) and (4.55) are computed when the charges of the test particle and the opposing beam have opposite signs. For equally charged beams the forces change signs. For small amplitudes the force is approximately linear and a particle crossing a beam at small amplitudes will experience a linear field. This results in a change of the tune like in a quadrupole. At larger amplitudes (i.e. above $\sim 1\sigma$) the force deviates strongly from this linear behaviour. Particles at larger amplitudes will also experience a tune change, however this tune change will depend on the amplitude. Already from the analytical form (4.55) one can see that the beam-beam force includes higher multipoles.

4.6.3 Incoherent Effects: Single Particle Effects

The force we have derived is the force of a beam on a single test particle. It can be used to study single particle or incoherent effects. For that we treat a particle crossing a beam like it was moving through a static electromagnetic lens. We have to expect all effects that are known from resonance and non-linear theory such as

- Unstable and/or irregular motion
- Beam blow up or bad lifetime

4.6.3.1 Beam–Beam Parameter

We can derive the linear tune shift of a small amplitude particle crossing a round beam of a finite length. We use the force to calculate the kick it receives from the opposing beam, i.e. the change of the slope of the particle trajectory. Starting from the two-dimensional force and multiplying with the longitudinal distribution which depends on both position s and time t , and assuming a Gaussian shape with a width of σ_s

$$F_r(r, s, t) = -\frac{Ne^2(1+\beta^2)}{\sqrt{(2\pi)^3}\epsilon_0\sigma_s} \frac{1}{r} \left[1 - \exp\left(-\frac{r^2}{2\sigma^2}\right) \right] \exp\left[-\frac{(s+vt)^2}{2\sigma_s^2}\right]$$

Now N is the total number of particles. We make use of Newton's law and integrate over the collision to get the radial deflection

$$\Delta r' = \frac{1}{mc\beta\gamma} \int_{-\infty}^{\infty} F_r(r, s, t) dt$$

The radial kick $\Delta r'$ a particle with a radial distance r from the opposing beam centre receives is then

$$\Delta r' = -\frac{2Nr_0}{\gamma} \frac{1}{r} \left[1 - \exp\left(-\frac{r^2}{2\sigma^2}\right) \right] \quad (4.56)$$

where I have re-written the constants and use the classical particle radius

$$r_0 = \frac{e^2}{4\pi\epsilon_0 mc^2} \quad (4.57)$$

where m is the mass of the particle. After the integration along the bunch length, N is the total number of particles. For small amplitudes r one can derive the asymptotic

limit

$$\Delta r' \Big|_{r \rightarrow 0} = -\frac{N r_0 r}{\gamma \sigma^2} = -r f \quad (4.58)$$

This limit is the slope of the force at $r = 0$ and the force becomes linear with a focal length as the proportionality factor.

It is well known how the focal length relates to a tune change and one can derive a quantity ξ which is known as the linear beam–beam parameter

$$\xi = \frac{N r_0 \beta^*}{4\pi \gamma \sigma^2} \quad (4.59)$$

r_0 is the classical particle radius, (e.g.: r_e , r_p) and β^* is the optical amplitude function (β -function) at the interaction point.

For small values of ξ and a tune far enough away from linear resonances this parameter is equal to the linear tune shift ΔQ .

The beam–beam parameter can be generalized for the case of non-round beams and becomes

$$\xi_{x,y} = \frac{N r_0 \beta_{x,y}^*}{2\pi \gamma \sigma_{x,y} (\sigma_x + \sigma_y)} \quad (4.60)$$

The beam–beam parameter is often used to quantify the strength of the beam–beam interaction, however it does not reflect the non-linear nature.

4.6.3.2 Non-linear Effects

Since the beam–beam forces are strongly non-linear, the study of beam–beam effects encompasses the entire field of non-linear dynamics (see earlier chapter) as well as collective effects. First, we briefly discuss the immediate effect of the non-linearity of the beam–beam force on a single particle. It manifests as an amplitude dependent tune shift and for a beam with many particles as a tune spread. The instantaneous tune shift of a particle when it crosses the other beam is related to the derivative of the force with respect to the amplitude $\delta F/\delta x$. For a particle performing an oscillation with a given amplitude the tune shift is calculated by averaging the slopes of the force over the range (i.e. the phases) of the particle's oscillation amplitudes. An elegant calculation can be done using the Hamiltonian formalism [156] developed for non-linear dynamics and as demonstrated in the chapter on non-linear dynamics using the Lie formalism. We get the formula for the non-linear detuning with the amplitude J

$$\Delta Q(J) = \xi \frac{2}{J} \left[1 - I_0 \left(\frac{J}{2} \right) e^{-\frac{J}{2}} \right] \quad (4.61)$$

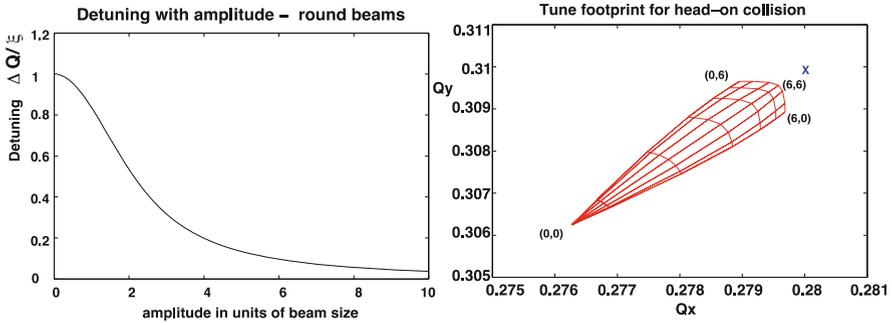


Fig. 4.22 Tune shift (non-linear detuning) as a function of the amplitude (left) and 2-dimensional tune footprint (right)

where $I_0(x)$ is the modified Bessel function and $J = \varepsilon\beta/2\sigma^2$ in the usual units. Here ε is the particle “emittance” and not the beam emittance.

In the 2-dimensional case, the tune shifts (ΔQ_x , ΔQ_y) of a particle with amplitudes x and y depend on both, horizontal and vertical amplitudes. The detuning must be computed and presented in a 2-dimensional form, i.e. the amplitude (x, y) is mapped into the tune space (Q_x, Q_y) or alternatively to the 2-dimensional tune change $(\Delta Q_x, \Delta Q_y)$. Such a presentation is usually called a “tune footprint” and an example is shown in Fig. 4.22(right) and it maps the amplitudes into the tune space and each “knot” of the mesh corresponds to a pair of amplitudes. Amplitudes between 0 and 6σ in both planes are used. The cross indicates the original, unperturbed tunes without the beam–beam interaction.

The maximum tune spread for a single head-on collision is equal to the tune shift of a particle with small amplitudes and for small tune shifts equal to the beam–beam parameter ξ . In the simple case of a single head-on collision the parameter ξ is therefore a measure for the tune spread in the beam.

4.6.3.3 Beam Stability

When the beam–beam interaction becomes too strong, the beam can become unstable or the beam dynamics is strongly distorted. One can distinguish different types of distortions and a few examples are

- Non-linear motion can become stochastic and can result in a reduction of the dynamic aperture and particle loss and bad lifetime.
- Distortion of beam optics: dynamic beta (LEP) [161].
- Vertical blow-up above the so-called beam–beam limit.

Since the beam–beam force is very non-linear, the motion can become “chaotic”. This often leads to a reduction of the available dynamic aperture. The dynamic aperture is the maximum amplitude where the beam remains stable. Particles outside

the dynamic aperture are eventually lost. The dynamic aperture is usually evaluated by tracking particles with a computer program through the machine where they experience the fields from the machine elements and other effects such as wake fields or the beam–beam interaction.

Since the beam–beam interaction is basically a very non-linear lens in the machine, it distorts the optical properties and it may create a noticeable beating of the β -function around the whole machine and at the location of the beam–beam interaction itself. This can be approximated by inserting a quadrupole which produces the same tune shift at the position of the beam–beam interaction. The r.m.s. beam size at the collision point is now proportional to $\sqrt{\beta_p^*}$, where β_p^* is the perturbed β -function which can be significantly different from the unperturbed β -function β^* . This in turn changes the strength of the beam–beam interaction and the parameters have to be found in a self-consistent form. This is called the dynamic beta effect. This is a first deviation from our assumption that the beams are static non-linear lenses. A strong dynamic beta effect was found in LEP [161] due to its very large tune shift parameters.

Another effect that can be observed in particular in $e^+ e^-$ colliders is the blow up of the emittance which naturally limits the reachable beam–beam tune shifts.

4.6.3.4 Beam–Beam Limit

In $e^+ e^-$ colliders the beam sizes are usually an equilibrium between the damping due to the synchrotron radiation and heating mechanisms such as quantum excitation, intra-beam scattering and very importantly, the beam–beam effect. This leads to a behaviour that is not observed in a hadron collider. When the luminosity is plotted as a function of the beam intensity, it should increase approximately as the current squared [162], in agreement with

$$\mathcal{L} = \frac{N^2 k f}{4\pi \sigma_x \sigma_y} \quad (4.62)$$

Here k is the number of bunches per beam and f the revolution frequency [162]. At the same time the beam–beam parameter ξ should increase linearly with the beam intensity according to (4.60)

$$\xi_y = \frac{N r_e \beta_y}{2\pi \gamma \sigma_y (\sigma_x + \sigma_y)} \quad (4.63)$$

In all $e^+ e^-$ colliders the observation can be made that above a certain current, the luminosity increases approximately proportional to the current, or at least much less than with the second power. Another observation is that at the same value of the intensity the beam–beam parameter ξ saturates. This limiting value of ξ is commonly known as the beam–beam limit.

When we re-write the luminosity as

$$\mathcal{L} = \frac{N^2 k f}{4\pi\sigma_x\sigma_y} = \frac{N k f}{4\pi\sigma_x} \frac{N}{\sigma_y} \quad (4.64)$$

we get an idea of what is happening. In $e^+ e^-$ colliders the horizontal beam size σ_x is usually much larger than the vertical beam size σ_y and changes very little. In order for the luminosity to increase proportionally to the intensity N , the factor N/σ_y must be constant. This implies that with increasing current the vertical beam size increases in proportion above the beam–beam limit. This has been observed in all $e^+ e^-$ colliders and since the vertical beam size is usually small, this emittance growth can be very substantial before the life time of the beam is affected or beam losses are observed [163].

The dynamics of machines with high synchrotron radiation is dominated by the damping properties and the beam–beam limit is not a universal constant nor can it be predicted. Simulation of beams with many particles can provide an idea of the order of magnitude [164, 165].

4.6.4 Studies of Head-on Collisions at the LHC

The layout of experimental regions in the LHC is shown in Fig. 4.23. The beams travel in separate vacuum chambers and cross in the experimental areas where they share a common beam pipe. In these common regions the beams experience head-on collisions as well as a large number of long range beam–beam encounters [166]. This arrangement together with the bunch filling scheme of the LHC as shown in Fig. 4.24 [166, 167] leads to very different collision pattern for different bunches, often referred to as “PACMAN” bunches. The number of both, head-on as well as long range encounters, can be very different for different bunches in the bunch trains and lead to a different integrated beam–beam effect [167]. This was always a worry in the LHC design and the effects have been observed in an early stage of the commissioning. Strategies have been provided to minimize these effects, e.g. different planes for the crossing angles [166, 167].

4.6.4.1 PACMAN Bunches

The bunches in the LHC do not form a continuous train of equidistant bunches spaced by 25 ns, but some empty space must be provided to allow for the rise time of kickers (Fig. 4.24). These gaps and the number of bunches per train are determined by requirements from the LHC injectors. The whole LHC bunch pattern is composed of 39 smaller trains (each with 72 bunches) separated by gaps of various length followed by a large abort gap for the dump kicker. Due to the symmetry,

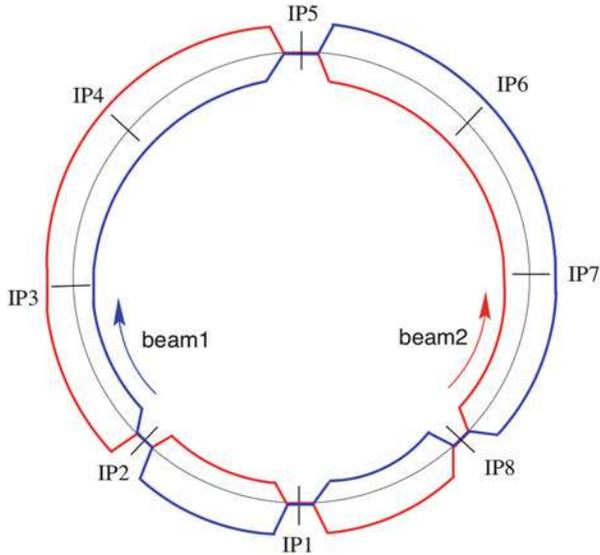


Fig. 4.23 Layout of the experimental collision points in the LHC [166]

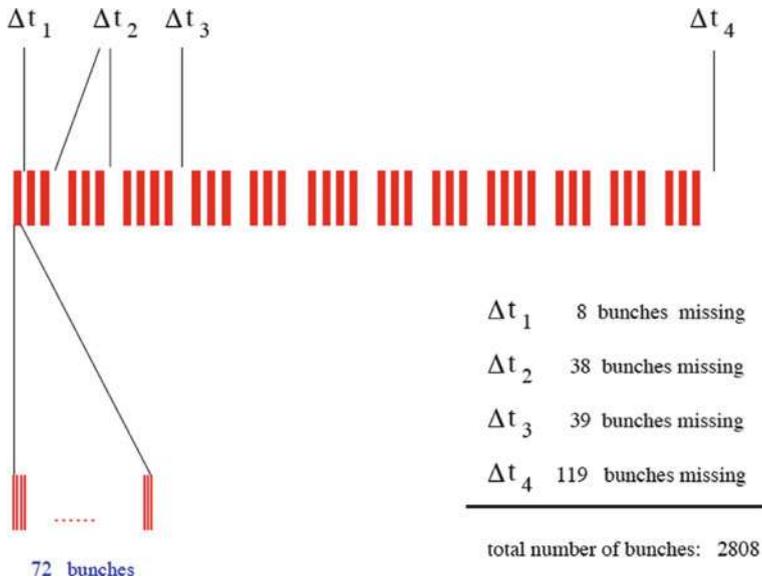


Fig. 4.24 Bunch filling scheme of the nominal LHC

bunches normally meet other bunches at the head-on collision point. For the long-range interactions this is no longer the case. Bunches at the beginning and at the end of a small train will encounter a hole and as a result experience fewer long-range

interactions than bunches from the middle of a train [168]. Bunches with fewer long-range interactions have a very different integrated beam–beam effect and a different dynamics must be expected. In particular they will have a different tune and occupy a different area in the working diagram, therefore may be susceptible to resonances which can be avoided for nominal bunches. The overall space needed in the working diagram is therefore largely increased [166, 168].

4.6.5 Head-on Beam–Beam Tune Shift

The nominal LHC parameters have been chosen to reach the design luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ [166]. The main parameters relevant for beam–beam effects are summarized in Table 4.1. At a very early stage of the LHC operation it was tested whether the nominal beam–beam parameters can be achieved. After this has been successfully demonstrated, we have performed a dedicated experiment to test the achievable beam–beam tune shift. To that purpose we have filled the LHC with single bunches per beam, colliding in IP1 and IP5 (see Fig. 4.23). We have used bunch intensities of $\sim 1.9 \times 10^{11}$ p/b, i.e. well above the nominal and the emittances have been reduced below $1.20 \mu\text{m}$ in both planes. It was shown that such bunches can be collided in both interaction points without significant losses or emittance increase [169] and we have demonstrated that a beam–beam tune shift of 0.017 for a single interaction and an integrated tune shift of 0.034 for both collision was possible. These tune shifts have been obtained in the absence of any long range encounters and it should be expected that the operationally possible tune shifts are lower.

4.6.6 Effect of Number of Head-on Collisions

Due to the filling pattern in the LHC, different bunches experience different numbers of head-on as well as long range interactions. Details are given in another contribution [170]. In Fig. 4.25 we show as illustration the losses of bunches with very different (0–3) numbers of head-on collisions. The data was taken during a

Table 4.1 LHC nominal parameters and achieved during operation and experiments in 2010/2011

Parameter	Nominal	Achieved
Intensity (p/bunch)	1.15×10^{11}	2.3×10^{11}
Emittance	$3.75 \mu\text{m}$	$\leq 2.00 \mu\text{m}$
β_*	0.55 m	1.5 m
ξ/IP	0.0035	0.0170
Bunch spacing	25 ns	50 ns
Bunches/beam	2808	1380

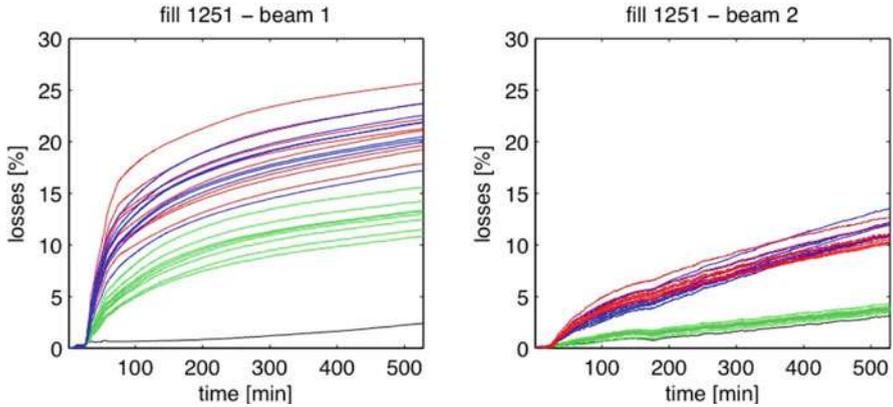


Fig. 4.25 Losses of bunches with different number of head-on collisions [170]. Numerology: blue (3 collisions), red (2 collisions), green (1 collision), black (no collision)

regular operational fill of 10 h duration. The correlation between losses and number of head-on collisions is apparent and a more detailed analysis is found in [170]. The transverse emittances during normal operation are larger ($\sim 2.5 \mu\text{m}$) than in the head-on test. In a second experiment we increased the bunch intensity further to $\sim 2.3 \times 10^{11}$ p/b with emittances of $\sim 1.80 \mu\text{m}$. Although the tune shift was slightly lower than in the previous experiment (0.015), the lifetime was worse. We interpret these results as losses of particles at large amplitudes. This is supported by the observation that the strongest losses occur at the very beginning of a fill (Fig. 4.25).

4.6.7 Crossing Angle and Long Range Interactions

To reach the highest luminosity, it is desirable to operate a collider with as many bunches as possible since the luminosity is proportional to their number (4.62) [162].

In a single ring collider such as the SPS, Tevatron or LEP, the operation with k bunches leads to $2k$ collision points. When k is a large number, most of them are unwanted and must be avoided to reduce the perturbation due to the beam-beam effects. Various schemes have been used to avoid these unwanted “parasitic” interactions. In the SPS, Tevatron and in LEP so-called Pretzel schemes were used. When the bunches are equidistant, this is the most promising method. When two beams of opposite charge travel in the same beam pipe, they can be moved onto separate orbits using electrostatic separators. In a well-defined configuration the two beams cross when the beams are separated. To avoid a separation around the whole machine, the bunches can be arranged in so-called trains of bunches following each other closely. In that case a separation with electrostatic separators is only needed

around the interaction regions. Such a scheme was used in LEP in the second phase [171].

Contrary to the majority of the colliders, the LHC collides particles of the same type which therefore must travel in separate beam pipes. At the collision points of the LHC the two beams are brought together and into collision. During that process it is unavoidable that the beams travel in a common vacuum chamber for more than 120 m. In the LHC the distance between the bunches is only 25 ns and therefore the bunches will meet in this region. In order to avoid the collisions, the bunches collide at a small crossing angle of $285 \mu\text{rad}$. While two bunches collide at a small angle (quasi head-on) at the centre, the other bunches are kept separated by the crossing angle. However, since they travel in a common beam pipe, the bunches still feel the electromagnetic forces from the bunches of the opposite beam. When the separation is large enough, these so-called long-range interactions should be weak.

4.6.7.1 Long-Range Beam–Beam Effects

Although the long-range interactions distort the beams much less than a head-on interaction, their large number and some particular properties require careful studies:

- They break the symmetry between planes.
- While the effect of head-on collisions is strongest for small amplitude particles, they mostly affect particles at large amplitudes.
- The tune shift caused by long-range interactions has opposite sign in the plane of separation compared to the head-on tune shift.
- They cause changes of the closed orbit [153].
- They largely enhance so-called PACMAN effects [168].

4.6.7.2 Opposite Sign Tune Shift

The opposite sign of the tune shift can easily be understood when we average the oscillation of a small amplitude particle as it samples the focusing force of the beam–beam interaction. When the separation is larger than $\sim 1.5\sigma$, the focusing (slope of the force as a function of the amplitude) changes the sign and the resulting tune shift assumes the opposite sign.

To some extent this property could be used to partially compensate long-range interactions when a configuration is used where the beams are separated in the horizontal plane in one interaction region and in the vertical plane in another one.

4.6.7.3 Strength of Long-Range Interactions

Assuming a separation d in the horizontal plane, the kicks in the two planes can be written as

$$\Delta x' = -\frac{2Nr_0}{\gamma} \frac{(x+d)}{r^2} \left(1 - e^{-\frac{r^2}{2\sigma^2}}\right) \quad (4.65)$$

with $r^2 = (x+d)^2 + y^2$. The equivalent formula for the plane orthogonal to the separation is

$$\Delta y' = -\frac{2Nr_0}{\gamma} \frac{y}{r^2} \left(1 - e^{-\frac{r^2}{2\sigma^2}}\right) \quad (4.66)$$

The effect of long-range interactions must strongly depend on the separation. The calculation shows that the tune spread ΔQ_{lr} from long-range interactions alone follows an approximate scaling (for large enough separation, i.e. above $\sim 6\sigma$)

$$\Delta Q_{lr} \propto -\frac{N}{d^2} \quad (4.67)$$

where N is the bunch intensity and d the separation. Small changes in the separation can therefore result in significant differences. Since the symmetry between the two planes is broken, the resulting footprint shows no symmetry. In fact, the tune shifts have different signs for x and y , as expected.

4.6.7.4 Footprint for Long-Range Interactions

Contrary to the head-on interaction where the small amplitude particles are mostly affected, now the large amplitude particles experience the strongest long-range beam–beam perturbations. This is rather intuitive since the large amplitude particles are the ones which can come closest to the opposing beam as they perform their oscillations. We must therefore expect a totally different tune footprint. Such a footprint for only long-range interactions is shown in Fig. 4.26.

4.6.8 Studies of Long Range Interactions in the LHC

To study the effect of long range beam–beam interactions we have performed a dedicated experiment [172]. The LHC was set up with single trains of 36 bunches per beam, spaced by 50 ns. The bunch intensities were $\sim 1.2 \times 10^{11}$ p/b and the normalized emittances around $2.5 \mu\text{m}$. The trains collided in IP1 and IP5, leading to a maximum of 16 long range encounters per interaction point for nominal bunches. First, the crossing angle (vertical plane) in IP1 was decreased in small steps and the

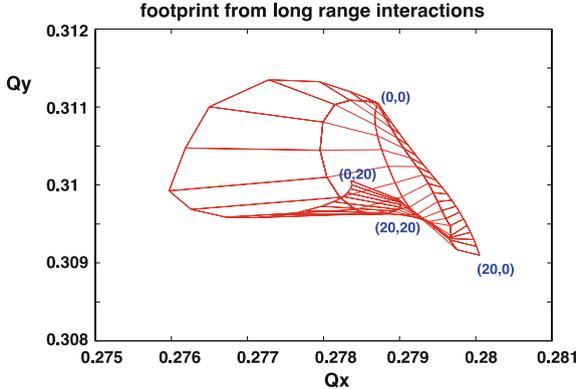


Fig. 4.26 Tune footprint for long-range interactions only. Vertical separation and amplitudes between 0 and 20σ

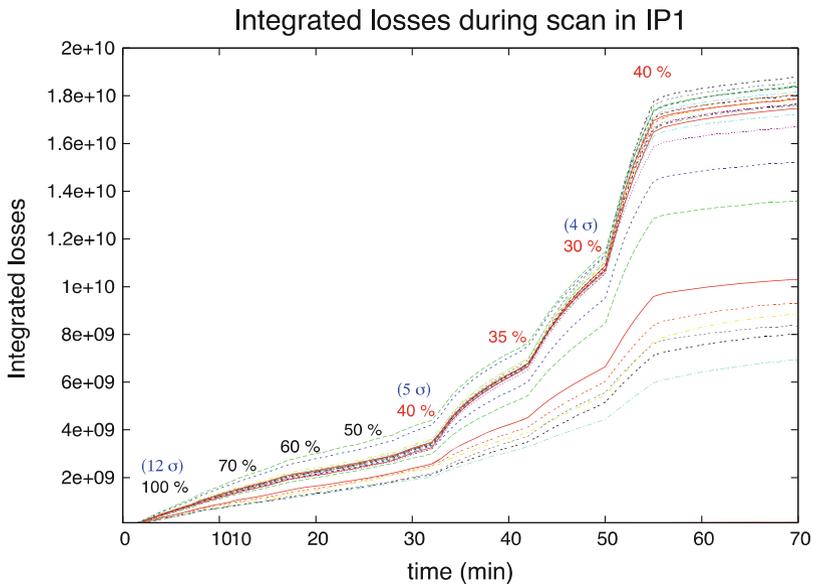


Fig. 4.27 Integrated losses of all bunches as a function of time during scan of beam separation in IP1. Numbers show percentage of full crossing angle

losses of each bunch recorded. The details of this procedure are described in [173] and the results are shown in Fig. 4.27 where the integrated losses for the 36 bunches in beam 1 are shown as a function of time and the relative change of the crossing angle is given in percentage of the nominal ($100\% \equiv 240 \mu\text{rad}$). The nominal value corresponds to a separation of approximately 12σ at the parasitic encounters. From Fig. 4.27 we observe significantly increased losses for some bunches when the separation is reduced to about 40%, i.e. around 5σ . Not all bunches are equally

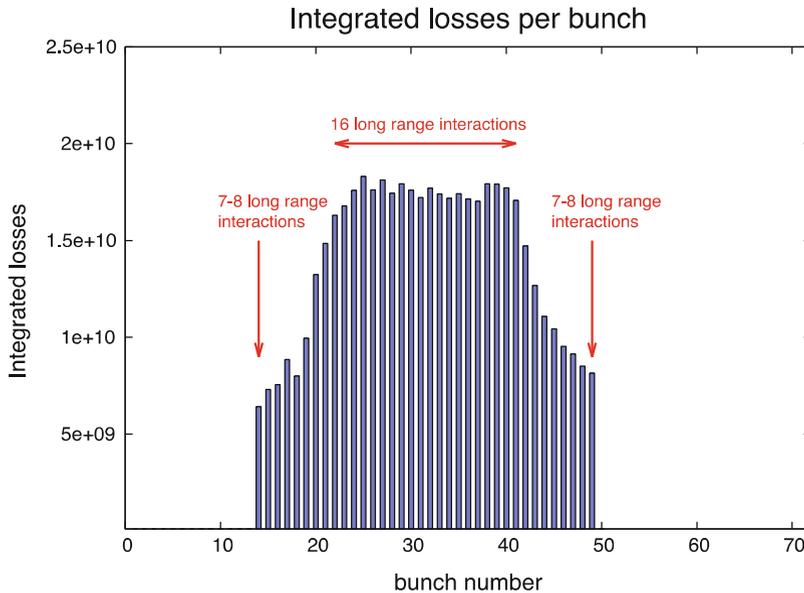


Fig. 4.28 Integrated losses of all bunches along a train of 36 bunches, after reducing the crossing angle in IP1

affected. At a smaller separation of 30% all bunches experience significant losses ($\sim 4\sigma$). Returning to a separation of 40% reduces the losses significantly, suggesting that mainly particles at large amplitudes have been lost during the scan due to a reduced dynamic aperture. Such a behaviour is expected [174]. The different behaviour is interpreted as a “PACMAN” effect and should depend on the number of long range encounters, which varies along the train. This is demonstrated in Fig. 4.28 where we show the integrated losses for the 36 bunches in the train at the end of the experiment. The maximum loss is clearly observed for the bunches in the centre of the train with the maximum number of long range interactions (16) and the losses decrease as the number of parasitic encounters decrease. The smallest loss is found for bunches with the minimum number of interactions, i.e. bunches at the beginning and end of the train [166, 167]. This is a very clear demonstration of the expected different behaviour, depending on the number of interactions.

In the second part of the experiment we kept the separation at 40% in IP1 and started to reduce the crossing angle in the collision point IP5, opposite in azimuth to IP1. Due to this geometry, the same pairs of bunches meet at the interaction points, but the long range separation is in the orthogonal plane. This alternating crossing scheme was designed to compensate first order effects from long range interactions [166]. The Fig. 4.29 shows the evolution of the luminosity in IP1 as we performed the scan in IP5. The numbers indicate again the relative change of separation, this time the horizontal crossing angle in IP5. The luminosity seems to show that the lifetime is best when the separation and crossing angles are equal for

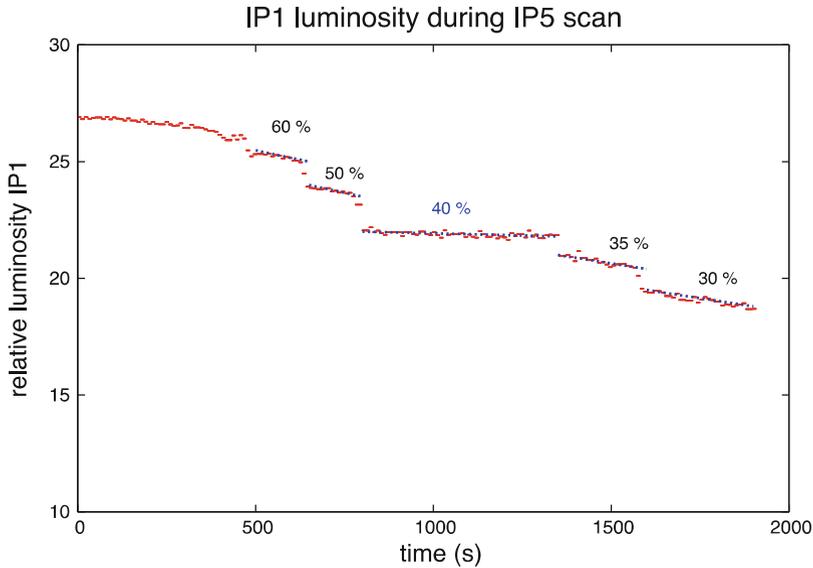


Fig. 4.29 Luminosity in IP1 as a function of time during scan of beam separation in IP5

the two collision points. It is worse for smaller as well as for larger separation. This is the expected behaviour for a passive compensation due to alternating crossing planes, although further studies are required to conclude.

4.6.8.1 Dynamic Aperture Reduction Due to Long-Range Interactions

For too small separation, the tune spread induced by long-range interactions can become very large and resonances cannot be avoided any more. The motion can become irregular and as a result particles at large amplitudes may be lost.

To evaluate the dynamic aperture in the presence of beam–beam interactions, a simulation of the complete machine is necessary and the interplay between the beam–beam perturbation and possible machine imperfections is important [174].

For the present LHC parameters we consider the minimum crossing angle to be $285 \mu\text{rad}$.

4.6.8.2 Beam–Beam Induced Orbit Effects

When two beams do not collide exactly head-on, the force has a constant contribution which can easily be seen when the kick $\Delta x'$ from (4.65), for sufficiently large

separation, is developed in a series

$$\Delta x' = \frac{\text{const}}{d} \left[1 - \frac{x}{d} + O\left(\frac{x^2}{d^2}\right) \right] \quad (4.68)$$

A constant contribution, i.e. more precisely an amplitude independent contribution, changes the orbit of the bunch as a whole. When the beam–beam effect is strong enough, i.e. for high intensity and/or small separation, the orbit effects are large enough to be observed.

When the orbit of a beam changes, the separation between the beams will change as well, which in turn will lead to a slightly different beam–beam effect and so on. The orbit effects must therefore be computed in a self-consistent way [175], in particular when the effects are sizeable. The closed orbit of an accelerator can usually be corrected, however an additional effect which is present in some form in many colliders, sets a limit to the correction possibilities. A particularly important example is the LHC and therefore it will be used to illustrate this feature.

We have to expect a slightly different orbit from bunch to bunch. The bunches in the middle of a train have all interactions and therefore the same orbit while the bunches at the beginning and end of a train show a structure which exhibits the decreasing number of long-range interactions. The orbit spread is approximately 10–15% of the beam size. Since the orbits of the two beams are not the same, it is impossible to make all bunches collide exactly head-on. A significant fraction will collide with an offset. Although the immediate effect on the luminosity is small [162], collisions at an offset can potentially affect the dynamics and are undesirable. The LHC design should try to minimize these offsets [168, 176]. A further consequence of the LHC filling and collision scheme is that not all bunches experience all head-on collisions [176]. Some of the bunches will collide only in 2 instead of the 4 nominal interaction points, leading to further bunch-to-bunch differences. In Fig. 4.30 we show a prediction for the vertical offsets in IP1 [166, 167]. The offsets should vary along the bunch train. Although the orbit measurement in the LHC is not able to resolve these effects, the vertex centroid can be measured bunch by bunch in the experiment (Fig. 4.31).

4.6.9 Coherent Beam–Beam Effects

So far, we have mainly studied how the beam–beam interaction affects the single particle behaviour and treated the beam–beam interaction as a static lens. In the literature, this is often called a “weak–strong” model: a “weak” beam (a single particle) is perturbed by a “strong” beam (not affected by the weak beam). When the beam–beam perturbation is important, the model of an unperturbed, strong beam is not valid anymore since its parameters change under the influence of the other beam and vice versa. When this is the case, we talk about so-called “strong–strong” conditions. The first example of such a “strong–strong” situation was the orbit effect

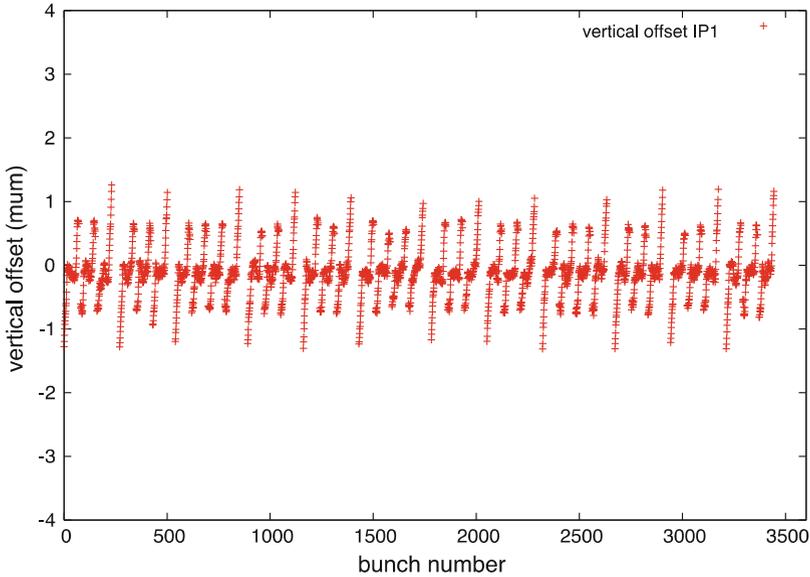


Fig. 4.30 Computed orbit offsets in IP1 along the bunch train [166, 167]

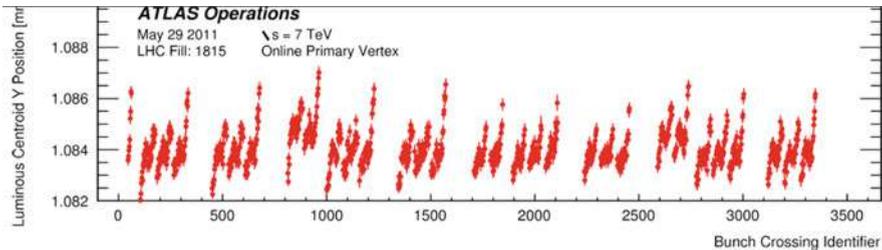


Fig. 4.31 Measured orbit offsets in IP1 along the bunch train [177, 178]

where the beams mutually changed their closed orbits. These closed orbits had to be found in a self-consistent way. This represents a static strong–strong effect.

In the next step, we investigate dynamic effects under the strong–strong condition [179].

When we consider the coherent motion of bunches, the collective behaviour of all particles in a bunch is studied. A coherent motion requires an organized behaviour of all particles in a bunch. A typical example are oscillations of the centre of mass of the bunches, so-called dipole oscillations. Such oscillations can be driven by external forces such as impedances and may be unstable. At the collision of two counter-rotating bunches not only the individual particles receive a kick from the opposing beam, but the bunch as an entity gets a coherent kick. This coherent kick of separated beams can excite coherent dipole oscillations. Its strength depends on the distance between the bunch centres at the collision point. It can be computed

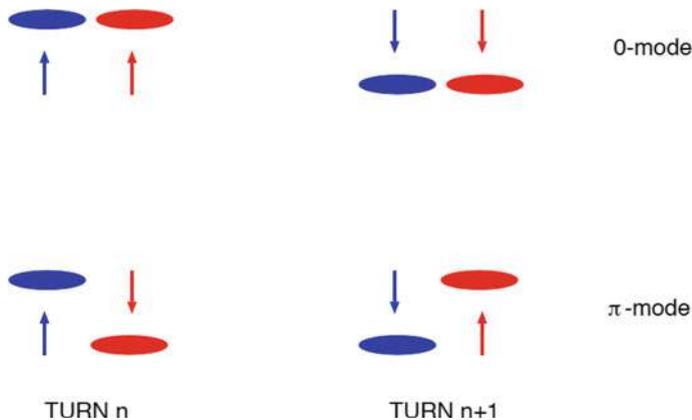


Fig. 4.32 Basic dipole modes of two bunches. Relative position of the bunches at the interaction point at two consecutive turns

by adding the individual contributions of all particles. For small distances, it can be shown [153, 180] that it is just one half of the incoherent kick a single particle would receive at the same distance. For distances large enough the incoherent and coherent kicks become the same.

4.6.9.1 Coherent Beam–Beam Modes

To understand the dynamics of dipole oscillations we first study the simplest case with one bunch in each beam. When the bunches meet turn after turn at the collision point, their oscillation can either be exactly in phase (0 degree phase difference) or out of phase (180 degrees or π phase difference). Any other oscillation can be constructed from these basic modes. The modes are sketched very schematically in Fig. 4.32. The relative positions of the bunches as observed at the interaction point are shown for two consecutive turns n and $n + 1$. The first mode is called the 0-mode (or sometimes called σ -mode) and the second the π -mode. In the first mode, the distance between the bunches does not change turn by turn and therefore there is no net force driving an oscillation. This mode must oscillate with the unperturbed frequency (tune) Q_0 . For the second mode, the net force difference between two turns is a maximum and the tune becomes $Q_0 + \Delta Q_{coh}$. The sign of ΔQ_{coh} depends whether the two beams have equal charge (defocusing case) or opposite charge (focusing case). The calculation of ΔQ_{coh} is non-trivial: when the bunches are considered as rigid objects, the tune shift can be computed easily using the coherent kick but is underestimated [181]. The correct calculation must allow for changes of the density distribution during the collision and moreover, must allow a deviation from a Gaussian function. The computation requires to solve the Vlasov-equation of two coupled beams [182–185].

Self-consistent multi-particle simulations form a complementary study, but require large computing resources. Furthermore, the fields produced by the beams must be computed in a self-consistent form before every collision [186].

The 0-mode is found at the unperturbed tune as expected. The π -mode is shifted by $1.2\text{--}1.3\xi$. The precise value depends on the ratio of the horizontal and vertical beam sizes [183]. We have seen before the incoherent tune spread (footprint) the individual particles occupy and we know that it spans the interval $[0.0, 1.0]\xi$, starting at the 0-mode.

Here one can make an important observation: under the strong–strong condition the π -mode is a discrete mode outside the incoherent spectrum [184, 185]. This has dramatic consequences for the stability of the beams. A coherent mode that is outside an incoherent frequency spectrum cannot be stabilized by Landau damping. Under these conditions the coherent beam–beam effect could drive the dipole oscillation to large amplitudes and may result in the loss of the beam. Observations of the coherent beam–beam effects have been made at PETRA [181]. Beam–beam modes have been observed with high intensity coasting beams in the ISR [187], and recently in a bunched hadron collider at RHIC [188].

Coherent beam–beam modes can be driven by head-on collisions with a small offset or by long-range interactions. In the first case and for small oscillations, the problem can be linearized and the theoretical treatment is simplified. The forces from long-range interactions are very nonlinear but the numerical evaluation is feasible. Since the coherent shift must have the opposite sign for long-range interactions, the situation is very different. In particular the π -mode from long-range interactions alone would appear on the opposite side of the 0-mode in the frequency spectrum [185, 186]. Both, the incoherent and the coherent spectra include both types of interactions.

4.6.10 Compensation of Beam–Beam Effects

For the case the beam–beam effects limit the performance of a collider, several schemes have been proposed to compensate all or part of the detrimental effects. The basic principle is to design correction devices which act as non-linear “lenses” to counteract the distortions from the non-linear beam–beam “lens”. For both head-on and long-range effects schemes have been proposed

- Head on effects:
 - Electron lenses
 - Linear lens to shift tunes
 - Non-linear lens to decrease tune spread
- Long-range effects:
 - At large distance: beam–beam force changes like $1/r$
 - Same force as a wire!

4.6.10.1 Electron Lenses

The basic principle of a compensation of proton–proton (or antiproton) collisions with an “electron lens” implies that the proton (antiproton) beam travels through a counter-rotating high current electron beam (“electron lens”) [189, 190]. The negative electron space charge can reduce the effect from the collision with the other proton beam.

An electron beam with a size much larger than the proton beam can be used to shift the tune of the proton beam (“linear lens”). When the current in the electron bunches can be varied fast enough, the tune shift can be different for the different proton bunches, thus correcting PACMAN tune shifts.

When the electron charge distribution is chosen to be the same as the counter-rotating proton beam, the non-linear focusing of this proton beam can be compensated (“non-linear lens”). When it is correctly applied, the tune spread in the beam can be strongly reduced.

Such lenses have been constructed at the Tevatron at Fermilab [190] and experiments are in progress.

4.6.10.2 Electrostatic Wire

To compensate the tune spread from long-range interactions, one needs a non-linear lens that resembles a separated beam. At large enough separation, the long-range force changes approximately with $1/r$ and this can be simulated by a wire parallel to the beam [191].

In order to compensate PACMAN effects, the wires have to be pulsed according to the bunch filling scheme. Tests are in progress at the SPS to study the feasibility of such a compensation for the LHC.

4.6.10.3 Möbius Scheme

The beam profiles of $e^+ e^-$ colliders are usually flat, i.e. the vertical beam size is much smaller than the horizontal beam size. Some studies indicate that the collision of round beams, even for $e^+ e^-$ colliders, show more promise for higher luminosity since larger beam–beam parameters can be achieved. Round beams can always be produced by strong coupling between horizontal and vertical planes. A more elegant way is the so-called Möbius lattice [192, 193]. In this lattice, the horizontal and vertical betatron oscillations are exchanged by an insertion. A horizontal oscillation in one turn becomes a vertical oscillation in the next turn and vice versa. Tests with such a scheme have been done at CESR at Cornell [193].

4.7 Numerical Modelling

G. Rumolo

Collective effects can be studied analytically, either through the perturbation formalism applied to the Vlasov equation or by means of few (typically two or three) particles models with the basic ingredients such as to reproduce the essential features of the phenomenon under study. Both ways are usually based on simplified approaches, in which some assumptions are necessary to make the models analytically solvable and lead to limited sets of equations relatively easy to interpret and handy to use.

The analytically solvable two or three particles models can be refined further to more realistic models, in which more than just few particles are assumed to represent the full particle beam. However, since the number of coupled differential equations to be solved grows proportionally with the number of particles used in the model, the resulting set of equations will rapidly become unmanageable as we increase the number of macroparticles (which are used to approximate the beam with a reduced number of particles), unless it is fed into a numerical simulation to be run on a computer. By using computers, the number of macroparticles necessary to model a beam can be pushed up to several millions, which is very useful to study the details of all possible internal oscillation modes of a bunch (or train of bunches), and also incoherent effects like emittance growth. Although ideally we would like to develop simulation programs that take into account the highest possible number of effects, in practice the existing codes narrow down their models to one or few effects, whose consequences in the beam dynamics are interesting to single out. For example, to study the effects of electron clouds, the beam will be made to interact with a given electron cloud at some locations around the accelerator ring, but in general other possible interactions with impedances, or the concurrent effects of space charge, beam–beam, Intra Beam Scattering, will be neglected. Although the study of two or more effects simultaneously is technically possible in most cases, at the present state of art of the simulations, it is generally preferred to limit the study of such interplays, because the combined models are difficult to control and tend to break down. Pushing further on this line, not only different effects can be decoupled in simulations, but also different regimes can be studied separately in the beam dynamics. For instance, for some problems only a partial description of the beam will be sufficient, so that transverse problems can be treated separately from longitudinal problems as well as single-bunch/multi-turn effects can be studied ignoring that these bunches are parts of long trains. In some cases, single-bunch/single- or multi-turn effects can also be modelled to generate driving terms to be used in reduced studies extending over longer time scales.

To perform a simulation, we will therefore have to define our beam as an ensemble of macroparticles, identified through arrays containing the phase space coordinates of each macro-particle (2–6-dimensional). This beam is first initialized and then transported across selected points of the accelerator ring using the

appropriate transformation matrices. At each of these points, the interaction with the desired collective effect will be applied (e.g. the beam's own space charge field, an electron cloud, a wake field). It is clear, therefore, that a numerical simulation requires in the first place the knowledge of the driving term to be applied at each interaction point. That is why in the following we separate the general simulation into the solution of an electromagnetic problem, in which the collective interaction is modelled and the resulting excitation on the beam is calculated (at least, its non-self-consistent part), and the beam tracking part, in which the evolution of a beam is studied under the effect of this excitation. Note that most of numerical simulations including collective effects are based on time domain models, as these are best suited to describe the usually non-stationary beam evolution under the effect of collective interactions.

4.7.1 The Electromagnetic Problem

The first step to set up a numerical simulation including a collective effect consists of identifying the source of the self-induced perturbation acting on the beam and modelling it in a way that can be subsequently used. We usually distinguish three different types of collective interactions, which can take place with: (i) space charge (see Sect. 4.1); (ii) wake fields from an accelerator component or part of the resistive beam pipe (see Sect. 4.2); (iii) another "beam" of charged particles. This secondary beam can be either a counter-rotating beam in a collider (see Sect. 4.6), or a static cloud formed by the accumulation of particles, usually of opposite charge, around the primary beam (see Sect. 4.5).

If the source of the perturbation is space charge, then two different approaches are possible to compute its effect. Analytical formulae are available for the electromagnetic fields of coasting beams with ellipsoidal or Gaussian transverse sizes, as well as of ellipsoidal or Gaussian bunches (in all dimensions). The additional kicks given by these electromagnetic fields can be therefore calculated and applied to the beam macroparticles in a finite number of locations along the ring (even if the space charge interaction is in reality continuous). When doing that, self-consistency requires that the sizes of the bunch are updated at every kick point. Another possible approach consists of using the macroparticle distributions at each selected kick point to calculate self-consistently the electric field with a Poisson solver and use it to calculate the electromagnetic kicks on the macroparticles. It is worth noting that the same approach can be used for beam–beam problems, because the shape of the required electric field is the same, even if the coefficients need to be adapted (electric and magnetic forces tend to cancel at ultra-relativistic energies for space charge, while they add up for beam–beam).

If the source of the perturbation is a wake field from an accelerator component or resistive wall, the shape of the relative wake function has to be calculated beforehand. This is done analytically for some specific cases (e.g. resistive wall, step or tapered transitions), but in general dedicated electromagnetic codes can be

used for this purpose. Some of them work in time domain, and can provide the wake potentials for given source bunches (usually chosen to be short enough as to simulate ideal pulse excitations and thus provide directly the wake functions). Other work in frequency domain and output impedances, which need then to be back-transformed into time domain to obtain the wake functions.

If the source of the perturbation is for example an electron cloud, then the electron distribution of the cloud is usually calculated beforehand by means of an electron cloud build-up code, and then its interaction with a coming bunch is calculated. Programs tracking simultaneously electrons and beam particles are presently under development or test, due to the massive memory and CPU requirements to solve this type of problems. On the other hand, generation and tracking of the ions can be included in multi-bunch beam tracking programs to calculate the effect of ions on bunched electron beams in a fully self-consistent manner. This is due to the fact that, while ions do not move significantly during the passage of an electron beam and allow modelling the bunches as charged disks, electrons can even perform several oscillations during the passage of a bunch, which requires a much more detailed modelling of the bunch.

4.7.2 *Beam Dynamics*

Beam dynamics tracking codes simulate the motion of beam particles inside an accelerator by transporting them across a number of discrete points by means of transformation matrices. In each of these points, additional kicks can be added, modelling either nonlinear components and errors of the external fields or the collective interactions. In the previous subsection, we have outlined the procedure to calculate the excitation to be applied to the beam to compute its evolution when it feels one or more collective interactions. To model the effects of space charge, wake fields and electron clouds, it is certainly necessary to describe the beam as an ensemble of macroparticles but beside that, its longitudinal structure needs also to be detailed. In particular, to model coupled bunch instabilities the relative positions of the macroparticles across the different bunches are necessary to determine the total effect of the wake acting on each of them. For single-bunch effects, a possible technique is to subdivide the bunch into several slices, so that the macroparticles of each slice can feel the integrated effect of the wakes left behind by the preceding slices (or the space charge from its own and the neighbouring slices, or the electron cloud as was deformed by the previous slices). A possible scheme of numerical simulation of a single bunch under the effect of a longitudinal wake field is illustrated in Fig. 4.33.

The bunch is first divided into N slices and a kick must be applied to each macroparticle within a given slice at a certain kick point. The kick depends on the longitudinal wake function and the charge distribution of the preceding slices. In the longitudinal plane, particles within a slice feel also the effect of the same slice to which they belong, because the bunch suffers a net energy loss. After all

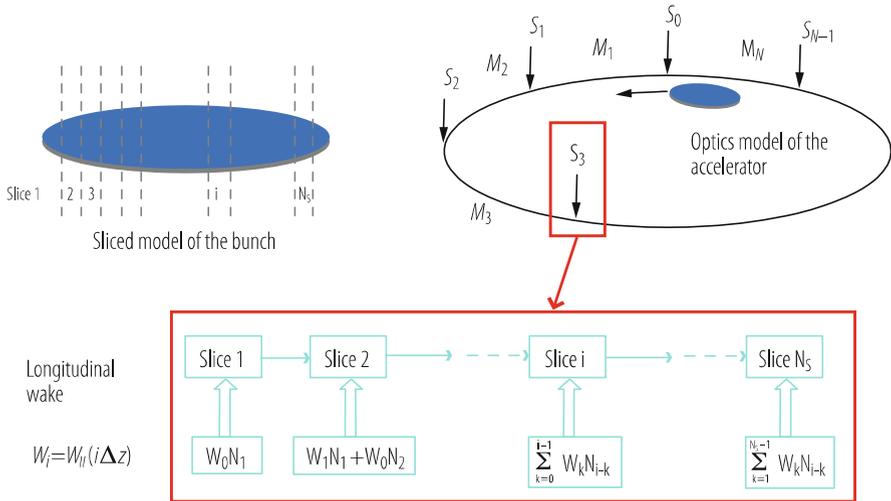


Fig. 4.33 Example of numerical simulation with collective effects: schematic view of the interaction of a single bunch with a longitudinal wake field at several locations in an accelerator ring

the particles in the bunch feel the effect of the wake at the kick point, they are subsequently transported to the next kick point in the accelerator optical model. Since synchrotron motion plays a key role in most of the effects under study, it is essential that particles are made to execute their synchrotron oscillations and move across slices from turn to turn. This means that collective effects dealing with single-bunch problems need to have at least one model of synchrotron motion built in, and that the bunch binning has to be regularly updated. Since the synchrotron motion is slow enough, and in reality the RF cavities do really kick the beam particles once or few times per turn, the longitudinal coordinates and the bunch slicing are usually not updated more frequently than once per turn. However, the update of the longitudinal coordinates from kick to kick point within one single turn, based on the only drift from momentum spread, could become significant especially in space charge simulations.

The simulation scheme with transverse wakes is basically the same as the one displayed in Fig. 4.33, except that particles inside one slice do not feel the effect of the same slice (as the transverse wakes are zero in the origin, for ultra-relativistic particles) and dipolar and quadrupolar contributions can be separated, making the wake kicks depending not only on the position of the source slice but also on the position of the witness macroparticles. The simulation scheme with the electron cloud is again similar to the one shown in Fig. 4.33, but the fundamental difference is that there is a mutual action between beam and electron cloud, so that, while macroparticles within a slice feel the effect of the electron cloud, the electron cloud itself is also deformed by the action of the passing slice.

The modelling described in the previous subsections has been frequently applied to explain collective instabilities observed in running machines, as well as to predict instability thresholds (both in existing and future machines) and develop strategies to circumvent limitations from collective effects. For instance, a detailed impedance model of the SPS comprises the contributions from several accelerator components and is used for deriving single-bunch wake fields, which are the driving terms for HEADTAIL simulations. The kicks given to the beam particles by the different wake fields can be then either applied at the real locations in which the sources are situated, or weighted by the beta functions, summed up and applied in a single location using a one-kick approximation. These simulations can be used for predicting at which intensity transverse mode coupling occurs and the effects of chromaticity on this threshold value [194]. This is very important to extrapolate the beam stability limits in different conditions of operation, e.g. with a different optics or to the upgraded machine, which will be in principle enabled to receive higher intensity bunches. The mode shift can be plotted as a function of the bunch intensity, because the main modes are detectable from the Fourier analysis of the centroid motion. A typical plot of mode shift provided by simulations is displayed in Fig. 4.10.

References

1. E. Métral et al., *Beam Instabilities in Hadron Synchrotrons*, IEEE Transactions on Nuclear Science, Vol. 63, No. 2, 50 p, April 2016 (invitation for the 50th anniversary of the PAC conference).
2. E. Métral (Issue Editor), ICFA Beam Dynamics Newsletter No. 69 devoted to the Collective Effects in Particle Accelerators, 310 p, December 2016.
3. E. Métral and V.G. Vaccaro (chairs), ICFA Mini-Workshop on “Electromagnetic Wake Fields and Impedances in Particle Accelerators”, Erice (Sicily, Italy), 2014: <https://indico.cern.ch/event/287930/>
4. M.R. Masullo, S. Petracca and G. Rumolo (chairs), ICFA Mini-Workshop on “Impedances and Beam Instabilities in Particle Accelerators”, Benevento (Italy), 2017: <https://agenda.infn.it/event/12603/>
5. E. Métral, G. Rumolo and T. Pieloni (chairs), ICFA Mini-Workshop on “Mitigation of Coherent Beam Instabilities in Particle Accelerators”, Zermatt (Switzerland), 2019: <https://indico.cern.ch/event/775147/>
6. B. Zotter: *Betatron Frequency Shifts due to Image and Self Fields*, CERN Accelerator School: General Accelerator Physics, CERN 85-19, Vol. I, p. 253, 1985.
7. K. Schindl: *Space Charge*, CERN-PS-99-012-DI, 1999.
8. A.W. Chao: *Physics of Collective Beam Instabilities in High Energy Accelerators*, New York: Wiley, 371 p, 1993.
9. K.Y. Ng: *Physics of Intensity Dependent Beam Instabilities*, World Scientific, 776 p, 2006.
10. S. Andriamonje, et al.: *Neutron TOF facility (PS213): Technical Design Report*, CERN-INTC-2000-004, 2000.
11. E. Métral, *Some effects near transition*, Proceedings of the CAS-CERN Accelerator School on Intensity Limitations in Particle Beams on November 2015, CERN Yellow Reports: School Proceedings, CERN-2017-006-SP.
12. E. Métral and G. Rumolo, USPAS course on “Collective Effects in Beam Dynamics” in Albuquerque, New Mexico, USA, June 22–26, 2009: <http://emetral.web.cern.ch/emetral/>

13. W. Hardt, D. Möhl: *Q-jump at Transition*, CERN ISR-300/GS/69-16, 1969.
14. D. Möhl: *Compensation of Space Charge Effects at Transition by a Asymmetric Q-jump – A Theoretical Study*, CERN ISR-300/GS/69-62, 1969.
15. W. Hardt: Proc. 9th Intern. Conf. High Energy Accelerators, Washington DC, 1974.
16. E. Métral and D. Möhl, *Transition Crossing*, Fifty years of the CERN Proton Synchrotron - Volume I (editors: S. Gilardoni and D. Manglunki), CERN-2011-004, 16 June 2011.
17. J.S. Berg, F. Ruggiero: *Landau Damping with Two-Dimensional Betatron Tune Spread*, CERN SL-AP-96-71 (AP), December 1996.
18. E. Métral, F. Ruggiero: *Stability Diagrams for Landau Damping with Two-Dimensional Betatron Tune Spread from both Octupoles and Nonlinear Space Charge Applied to the LHC at Injection*, Proc. EPAC2004, Lucerne, Switzerland, 5–9 July 2004.
19. G. Franchetti, et al.: *Space Charge and Octupole Driven Resonance Trapping Observed at the CERN Proton Synchrotron*, Phys. Rev. ST Accel. Beams 6, 124201 (2003).
20. G. Franchetti: *Simulations of the Space Charge and Beam Loss in the SIS18 Experiment*, Proc. ICFA HB2010, Morschach, Switzerland, Sept. 27–Oct. 1, 2010.
21. B.W. Montague: *Fourth-Order Coupling Resonance Excited by Space-Charge Forces in a Synchrotron*, CERN 68-38, 1968.
22. E. Métral: *Space Charge Experiment at the CERN Proton Synchrotron*, Proc. ICFA HB2004, Bensheim, Germany, 18–22/10/2004.
23. L.J. Laslett: *On Intensity Limitations Imposed by Space Charge Effects in Circular Particle Accelerators*, in: Proc. Summer Study on Storage Rings, Accelerators and Instrumentation at Super High Energies, BNL Report 7534, p. 324, 1963.
24. K. Yokoya: *Resistive Wall Impedance of Beam Pipes of General Cross Section*, Part. Accel., Vol. 41, 3–4, p. 221, 1993.
25. V.G. Vaccaro, *Longitudinal Instability of a Coasting Beam Above Transition, due to the Action of Lumped Discontinuities*, CERN Rep. ISRRF/66-35, Tech. Rep., 1966.
26. A. Sessler and V. Vaccaro, CERN Report ISR-RF/67-2 (1967).
27. H. Tsutsui: *On single Wire Technique for Transverse Coupling Impedance Measurement*, CERN-SL-Note-2002-034 AP, 2002.
28. A. Burov, V. Danilov: *Suppression of Transverse Bunch Instabilities by Asymmetries in the Chamber Geometry*, Phys. Rev. Lett. 82, 2286–2289 (1999).
29. N. Mounet, E. Métral: *Impedances of Two Dimensional Multilayer Cylindrical and Flat Chambers in the Non-Ultrarelativistic case*, Proc. ICFA HB2010, Morschach, Switzerland, Sept. 27–Oct. 1, 2010.
30. CST Particle Studio. <http://www.cst.com/Content/Products/PS/Overview.aspx>
31. F. Caspers: *Bench Measurements*, in: Handbook of Accelerator Physics and Engineering, 2nd printing, edited by A.W. Chao and M. Tigner, p. 574.
32. Ansoft – HFSS. <http://www.ansoft.com/products/hf/hfss/>
33. E. Métral, et al.: *Kicker Impedance Measurements for the Future Multiturn Extraction of the CERN Proton Synchrotron*, Proc. EPAC2006, Edinburgh, UK, 26–30 June 2006.
34. H.A. Day: *Measurements and Simulations of Impedance Reduction Techniques in Particle Accelerators*, PHD Thesis, University of Manchester (England), 2013.
35. L.J. Laslett, V.K. Neil, A.M. Sessler: *Transverse Resistive Instabilities of Intense Coasting Beams in Particle Accelerators*, R.S.I 36(4), 436–448, 1965.
36. B. Zotter: *New Results on the Impedance of Resistive Metal Walls of Finite Thickness*, CERN-AB-2005-043, 2005.
37. E. Métral: *Transverse Resistive-Wall Impedance from Very Low to Very High Frequencies*, CERN-AB-2005-084, 08 August 2005.
38. A.M. Al-Khateeb, et al.: *Transverse Resistive Wall Impedances and Shielding Effectiveness for Beam Pipes of Arbitrary Wall Thickness*, Phys. Rev. ST Accel. Beams 10, 064401 (2007).
39. F. Roncarolo, et al.: *Comparison Between Laboratory Measurements, Simulations, and Analytical Predictions of the Transverse Wall Impedance at Low Frequencies*, Phys. Rev. ST Accel. Beams 12, 084401 (2009).

40. N. Mounet, E. Métral: *Impedances of an Infinitely Long and Axisymmetric Multilayer Beam Pipe: Matrix Formalism and Multimode Analysis*, Proc. 1st Intern. Particle Accelerator Conf., Kyoto, Japan, 23–28 May 2010.
41. N. Mounet, E. Métral: *Generalized Form Factors for the Beam Coupling Impedances in a Flat Chamber*, Proc. 1st Intern. Particle Accelerator Conf., Kyoto, Japan, 23–28 May 2010.
42. B. Salvant, et al.: *Quadrupolar Transverse Impedance of Simple Models of Kickers*, Proc. 1st Intern. Particle Accelerator Conf., Kyoto, Japan, 23–28 May 2010.
43. C. Zannini, et al.: *Electromagnetic Simulations of Simple Models of Ferrite Loaded Kickers*, Proc. 1st Intern. Particle Accelerator Conf., Kyoto, Japan, 23–28 May 2010.
44. E. Métral: *Beam Screen Issues (with 20 T dipole magnets instead of 8.3 T)*, Proc. High-Energy LHC workshop, Malta, 14–16 October 2010.
45. A. Hofmann: *Beam Instabilities*, CERN Accelerator School, Rhodes, Greece, 20 Sept.–1 Oct., 1993, p. 307.
46. N. Biancacci, B. Salvant: *Dependence of Impedance vs. Phase Advance in HEADTAIL*, https://indico.cern.ch/event/164172/contributions/1420072/attachments/196173/275289/19-12-2011_Impedance-Vs-PhaseAdvance.pdf
47. N. Biancacci, *HEADTAIL + MAD-X*, https://indico.cern.ch/event/178920/contributions/1446485/attachments/235706/329825/HDTL_lattice_def.pdf
48. E.D. Courant, A.M. Sessler: *Transverse Coherent Resistive Instabilities of Azimuthally Bunched Beams in Particle Accelerators*, R.S.I., 37(11), 1579–1588, 1966.
49. C. Pellegrini, *Nuovo Cimento*, 64°, 477, 1969.
50. M. Sands: *The Head-Tail Effect: an Instability Mechanism in Storage Rings*, SLAC-TN-69-8, 1969.
51. M. Sands: *Head-Tail Effect II: From a Resistive-Wall Wake*, SLAC-TN-69-10, 1969.
52. F. Sacherer: *Single beam collective phenomena-transverse (bunched beams)*, Proc. First Course Intern. School on Accelerators, Erice, 1976: Theoretical Aspects of the Behaviour of Beams in Accelerators and Storage Rings, CERN 77-13, p. 210, 1977.
53. J.L. Laclare: *Bunched Beam Coherent Instabilities*, CERN Accelerator School, Oxford, CERN 87-03, p. 264, 1987.
54. J.P. Gourber, H.G. Hereward, S. Myers: *Overlap Knock-Out Effects in the CERN Intersecting Storage Rings (ISR)*, IEEE Trans. Nucl. Sci. NS-24(3), June 1977.
55. S. Myers: private communication (2011).
56. K.W. Robinson: Cambridge Electron Accel. Report CEAL-1010 (1964).
57. J.L. Laclare: *Introduction to Coherent Instabilities – Coasting Beam Case*, CERN Accelerator School: General Accelerator Physics, CERN 85-19, Vol. II, p. 377, 1985.
58. E. Keil, W. Schnell: CERN Report ISR-TH-RF/69-48, 1969.
59. D. Boussard: *Observation of Microwave Longitudinal Instabilities in the CPS*, CERN Report LabII/RF/Int./75-2, 1975.
60. F.J. Sacherer: *Bunch Lengthening and Microwave Instability*, IEEE Trans. Nucl. Sci. NS-24(3), June 1977.
61. K. Ng: *Potential-Well Distortion and Mode-Mixing Instability in Proton Machines*, FERMILAB-FN-630, 1995.
62. E. Métral: *Stability Criterion for the Longitudinal Mode-Coupling Instability in the Presence of Both Space-Charge and Resonator Impedances*, CERN/PS 2001-063 (AE), 2001.
63. E. Métral, *GALACTIC and GALACLIC: Two Vlasov Solvers for the Transverse and Longitudinal Planes*, Proceedings of IPAC2019, Melbourne, Australia, May 19–24, 2019.
64. E. Métral and M. Migliorati, *Longitudinal Mode-Coupling Instability: GALACLIC Vlasov Solver vs. Macroparticle Tracking Simulations*, Proceedings of IPAC2019, Melbourne, Australia, May 19–24, 2019.
65. E. Shaposhnikova: *Signatures of Microwave Instability*, CERN-SL-99-008 HRF, 1999.
66. J.S. Berg, R.D. Ruth: *Transverse Instabilities for Multiple Nonrigid Bunches in a Storage Ring*, Phys. Rev. E 52(3), September 1995.
67. H. Hereward: Rutherford Lab. Reports RL-74-062, EPIC/MC/48 (1974), and RL-75-021, EPIC/MC/70 (1975). See also B. Chen, A. W. Chao: SSCL Report 606 (1992).

68. D. Boussard, T. Linnecar: Proc. 2nd Euro. Part. Accel. Conf, Nice, 1990, p. 1560.
69. R.D. Kohaupt: *Head Tail Turbulence and the Transverse PETRA Instability*, DESY 80/22, 1980.
70. E. Métral: *Stability Criteria for High-Intensity Single-Bunch Beams in Synchrotrons*, Proc. 8th EPAC, Paris, France, 3–7 June 2002.
71. A. Burov and T. Zolkin, *TMCI with Resonator Wakes*, arXiv:1806.07521v2 [physics.acc-ph].
72. G. Rumolo et al., *Simulation Study on the Energy Dependence of the TMCI Threshold in the CERN-SPS*, Proceedings of EPAC 2006, Edinburgh, Scotland.
73. Y.H. Chin: *User's guide for new MOSES version 2.0: Mode-coupling Single bunch instability in an Electron Storage ring*, CERN/LEP-TH/88-05, 1988.
74. G. Rumolo, F. Zimmermann: *Practical user guide for HEADTAIL*, CERN-SL-Note-2002-036-AP, 2002.
75. B. Salvant: *Impedance Model of the CERN SPS and Aspects of LHC Single-Bunch Stability*, PHD Thesis, Ecole Polytechnique Fédérale de Lausanne (Switzerland), 2010.
76. A. Burov, *Nested Head-Tail Vlasov Solver*, Phys. Rev. ST Accel. Beams 17, 021007 (2014).
77. N. Mounet, *DELPHI: An Analytic Vlasov Solver for Impedance-Driven Modes*, CERN internal HSC meeting, May 07, 2014.
78. E. Métral, G. Rumolo: *Simulation Study on the Beneficial Effect of Linear Coupling for the Transverse Mode Coupling Instability in the CERN Super Proton Synchrotron*, Proc. EPAC 2006, Edinburgh, UK, 26–30 June 2006.
79. S. Myers: LEP Note 436, 1983.
80. R. Ruth: CERN LEP-TH/83-22, 1983.
81. R. Ruth: Proc. 12th Intern. Conf. High Energy Accelerators, Fermilab, Batavia, August 11–16, 1983, p. 389.
82. S. Myers: J. Vancraeynest: CERN LEP-RF/84-13, 1984.
83. S. Myers: CERN LEP-RF/85-22, 1985.
84. S. Myers: Proc. IEEE PAC, Washington D.C., March 16, 1987, p. 503–507.
85. M. Karliner, K. Popov: *Theory of a Feedback to Cure Transverse Mode Coupling Instability*, Nucl. Instrum. Meth. A 537 (2005), p. 481–500.
86. E. Métral et al., *Destabilising Effect of the LHC Transverse Damper*, in Proc. IPAC'18, Vancouver, BC, Canada, Apr–May 2018, paper THPAF048, pp. 3076-3079. <https://doi.org/10.18429/JACoW-IPAC2018-THPAF048>
87. E. Métral et al., *Impedance Models, Operational Experience and Expected Limitations*, International Review of the HL-LHC Collimation System, CERN, 11-12/02/2019.
88. E. Métral, *A Two-Mode Model to Study the Effect of Space Charge on TMCI in the “Long-Bunch” regime*, Proceedings of IPAC2019, Melbourne, Australia, May 19–24, 2019.
89. E. Métral et al., *Space Charge and Transverse Instabilities at the CERN SPS and LHC*, Proceedings of ICAP'18, Key West, Florida, USA, October 20–24, 2018.
90. A. Burov, “Convective Instabilities of Bunched Beams with Space Charge”, arXiv:1807.04887v4 [physics.acc-ph].
91. E. Métral, *Summary of my 3 IPAC19 papers*, CERN LIU-SPS BD WG meeting, 17/01/2019.
92. H.G. Hereward: *Landau Damping by Non-Linearity*, CERN/MPS/DL 69-11, 1969.
93. E. Métral, A. Verdier: *Stability Diagrams for Landau Damping with a Beam Collimated at an Arbitrary Number of Signas*, CERN-AB-2004-019-ABP.
94. E. Métral, *Landau Damping for TMCI: Without vs. With Transverse Damper (TD)*, CERN HSC internal meeting, 25/03/2019, https://indico.cern.ch/event/807899/contributions/3362767/attachments/1816203/2971974/LDforTMCI_EM_25-03-2019_2.pdf
95. E. Métral: *Theory of Coupled Landau Damping*, Part. Accelerators 62(3–4), p. 259, 1999.
96. E. Métral et al., *Simulation Study of the Horizontal Head-Tail Instability Observed at Injection of the CERN Proton Synchrotron*, Proceedings of PAC07, Albuquerque, New Mexico, USA, June 25–29, 2007.
97. E. Métral, et al.: *Destabilising Effect of Linear Coupling in the HERA Proton Ring*, Proc. 8th EPAC, Paris, France, 3–7 June 2002.

98. L.R. Carver et al., Transverse Beam Instabilities in the Presence of Linear Coupling in the Large Hadron Collider, PRAB 21, 044401 (2018).
99. E. Métral: Simple Theory of Emittance Sharing and Exchange due to Linear Betatron Coupling, CERN/PS 2001-066 (AE), December 2001.
100. A. Franchi, E. Métral, R. Tomás: *Emittance Sharing and Exchange Driven by Linear Betatron Coupling in Circular Accelerators*, Phys. Rev. ST Accel. Beams 10, 064003 (2007).
101. D. Möhl, H. Schönauer: Landau Damping by Non-Linear Space-Charge Forces and Octupoles, Proc. IX Intern. Conf. High Energy Acc., Stanford, 1974 (AEC, Washington, D.C., 1974), [CONF 740522] p. 380.
102. D. Möhl: On Landau Damping of Dipole Modes by Non-Linear Space Charge and Octupoles, CERN/PS 95-08 (DI), 1995.
103. M. Blaskiewicz: Phys. Rev. ST Accel. Beams 1, 044201 (1998).
104. E. Métral, F. Ruggiero: Stability Diagrams for Landau Damping with Two-Dimensional Betatron Tune Spread from both Octupoles and Nonlinear Space Charge, CERN-AB-2004-025 (ABP), 2004.
105. A. Burov: Phys. Rev. ST Accel. Beams 12, 044202 (2009).
106. V. Balbekov: Phys. Rev. ST Accel. Beams 12, 124402 (2009).
107. V. Kornilov: Head-Tail Bunch Dynamics with Space Charge, Proc. ICFA HB2010, Morschach, Switzerland, Sept. 27–Oct. 1, 2010.
108. X. Buffat et al., *Transverse Mode Coupling Instability of Colliding Beams*, Phys. Rev. ST Accel. Beams 17, 041002 (2014).
109. X. Buffat, *Transverse Beams Stability Studies at the Large Hadron Collider*, Ph.D. dissertation, EPFL, 2015.
110. S.V. Furusest et al., MD3288: Instability latency with controlled noise, CERN internal note, under publication.
111. S. Nagaitsev, et al.: Nonlinear Optics as a Path to High-Intensity Circular Machines, Proc. ICFA HB2010, Morschach, Switzerland, Sept. 27–Oct. 1, 2010.
112. V. Shiltsev et al., *Landau Damping of Beam Instabilities by Electron Lenses*, Phys. Rev. Lett. 119, 134802 – Published 27 September 2017.
113. M. Schenk et al., *Analysis of Transverse Beam Stabilization with Radio Frequency Quadrupoles*, PRAB 20, 104402 (2017).
114. M. Schenk et al., *Experimental Stabilization of Transverse Collective Instabilities in the LHC with Second Order Chromaticity*, PRAB 21, 084401 (2018).
115. M. Schenk et al., *Vlasov Description of the Effects of Nonlinear Chromaticity on Transverse Coherent Beam Instabilities*, PRAB 21, 084402 (2018).
116. E. Métral: *Stability of the Longitudinal Bunched-Beam Coherent Dipole Mode*, Proc. APAC'04, Gyeongju, Korea, 22–26 March 2004.
117. K. Ohmi: Phys. Rev. Lett. 75, 1526 (1995).
118. *Tables and Graphs of electron-Interaction Cross Sections from 10 eV to 100 GeV Derived from the LLNL Evaluated Electron Data Library (EDLL), Z=1-100*, UCRL-50400, 31 (1991).
119. T. Raubenheimer, P. Chen: *Ions in the Linacs of Future Linear Colliders*, Proc. LINAC'92, Ottawa, Canada (1992).
120. P. Thieberger, et al: Phys. Rev. A 61, 042901 (2000).
121. O. Grobner: *Bunch Induced Multipacting*, 10th Intern. Conf. High Energy Accelerators, Protvino, July (1977).
122. R. Cimino, et al.: Phys. Rev. Lett. 93, 014801 (2004).
123. M. Furman, G. Lambertson: *The Electron-Cloud Instability in PEP-II*, Proc. EPAC'96, Sitges (1996).
124. F. Zimmermann: LHC Project Report 95 (1997).
125. L. Wang, et al.: PRST-AB 5, 124402 (2002).
126. H. Fukuma, et al.: *Observation of Vertical Beam Blow-up in KEKB Low Energy Ring*, Proc. EPAC'00, Vienna (2000).

127. G. Arduini, et al.: *Transverse Behaviour of the LHC Proton Beam in the SPS: An Update*, PAC2001, Chicago (2001).
128. K. Cornelis: *The Electron Cloud Instability in the SPS*, ELOUD'02, Geneva (2002).
129. C. Yin Valgren, et al.: *Amorphous Carbon Coatings for Mitigation of Electron Cloud in the CERN SPS*, Proc. 1st Intern. Particle Accelerator Conf., Kyoto, Japan, 23–28 May 2010.
130. G. Rumolo et al., *Electron Cloud observation in the LHC*, Proc. IPAC'11, San Sebastian, Spain (2011).
131. G. Iadarola, *Electron cloud studies for CERN particle accelerators and simulation code development*, PhD thesis, CERN-THESIS-2014-047 (2014).
132. L. Methier et al., *Electron cloud in 2016: cloudy or clear?*, Proc. 7th Evian Workshop, Evian, France, 2016, CERN-ACC-2017-094 (2017).
133. G. Arduini et al., *How to Maximize the HL-LHC performance*, Proc. of the Review of the LHC and Injector Upgrade Plans Workshop (RLIUP), 29–31 October 2013 Archamps, France, CERN-2014-006 (2013)
134. M. Palmer, et al.: *Electron Cloud at Low Emittance in CEsrTA*, Proc. 1st Intern. Particle Accelerator Conf., Kyoto, Japan, 23–28 May 2010.
135. V. Baglin, et al.: LHC Project Report 472, (2002).
136. M. Furman, G. Lambertson: *The Electron-Cloud Instability in the Arcs of the PEP-II Positron Ring*, Proc. MBI97 Tsukuba (1997).
137. P. Channell, A. Jason: LA AOT-3 TN 93-11, private communication (1994).
138. P.R. Zenkevich, D.G. Koskarev: *Coupling Resonances of the Transverse Oscillations of Two Circular Beams*, ITEF-841, (1970), translated as UCRL-Trans. 1451 (1971).
139. E. Keil, B. Zotter: CERN-ISR-TH/71-58 (1971).
140. K. Oide, et al., in: Proc. International Workshop on Performance Improvement of Electron-Positron Collider Particle Factories, KEKB Report No. 99-24 (2000).
141. K. Ohmi, F. Zimmermann: Phys. Rev. Lett. 85, 3821 (2000)
142. G. Arduini, et al., in: Proc. Workshop LEP-SPS Performance Chamonix X (2000), pp. 119–122.
143. G. Arduini, et al., in: Proc. PAC 2003, 12–16 May, Portland, USA (2003).
144. K. Li et al., *Beams during the cycle: Quality and behaviour*, LHC Performance Workshop 2017, Chamonix, France (2017).
145. A. Romano et al., *Electron cloud instabilities triggered by low bunch intensity at the Large Hadron Collider*, presently submitted to PRAB.
146. D. Neuffer, et al: *Observations of a fast transverse instability in the PSR*, Nucl. Instrum. Meth. A 321, pp. 1–12 (1992).
147. A. Grunder, G.R. Lambertson: Report UCRL 20691 (1971).
148. G. Rumolo, in: Proc. CARE-HHH-APD Workshop, CERN (2004).
149. G. Rumolo, et al.: Phys. Rev. Lett. 100, 144801 (2008).
150. Proc. ELOUD02, CERN, Geneva, 2002, edited by G. Rumolo and F. Zimmermann as CERN Yellow Report No. CERN-2002-001 (2002).
151. R. Cimino et al., Phys. Rev. Lett., 109, 064801 (2012).
152. F. Ruggiero, X.L. Zhu: *Collective Instabilities in the LHC: Electron Cloud and Satellite Bunches*, Proc. Workshop Instabilities of High Intensity Hadron Beams in Rings, Upton, NY, USA, 28 Jun–1 Jul 1999, pp. 40–48.
153. W. Herr: *Beam-Beam Interactions*, Proceedings of CERN Accelerator School, Zeuthen 2003, CERN-2006-002 (2006).
154. Proceedings of the ICFA-Mini-Workshop on Beam-Beam effects in Hadron Colliders, CERN, Geneva, Switzerland, 18–22 April 2013, edited by W. Herr and G. Papotti, CERN-2014-004, <https://doi.org/10.5170/CERN-2014-004>
155. A. Chao: *The Beam-Beam Instability*, SLAC-PUB-3179 (1983).
156. L. Evans, J. Gareyte: *Beam-Beam Effects*, CERN Accelerator School, Oxford 1985, CERN 87-03 (1987).
157. A. Zholents: *Beam-Beam Effects in Electron-Positron Storage Rings*, Joint US-CERN School on Particle Accelerators, in Springer, Lecture Notes in Physics, Vol. 400 (1992).

158. E. Keil: *Beam-Beam Dynamics*, CERN Accelerator School, Rhodes 1993, CERN 95-06 (1995).
159. S. Kheifets, PETRA-Kurzmitteilung 119 (DESY 1976).
160. M. Basetti and G.A. Erskine, *Closed Expression for the Electrical Field of a 2-Dimensional Gaussian Charge*, CERN-ISR-TH/80-06 (1980).
161. D. Brandt, et al.: *Is LEP Beam-Beam Limited at its Highest Energy?* In Proceedings of the Particle Accelerator Conference 1999, New York, (1999) p. 3005.
162. W. Herr: *Concept of Luminosity*, Proceedings of CERN Accelerator School, Zeuthen 2003, CERN-2006-002 (2006).
163. J.T. Seeman: *Observations of the Beam-Beam Interaction*, Lecture Notes in Physics, Vol. 247, Springer, (1986).
164. S. Myers: IEEE Trans. Nucl. Sci. NS-28, 2503 (1981).
165. S. Myers: *Review of beam-Beam Simulations*, Lecture Notes in Physics, Vol. 247, Springer, (1986).
166. W. Herr, *Features and Implications of Different LHC Crossing Schemes*, CERN LHC Project Report 628 (2003).
167. W. Herr, *Dynamic Behaviour of Nominal and PACMAN Bunches for Different LHC Crossing Schemes*, CERN LHC Project Report 856 (2005).
168. W. Herr, *Effects of PACMAN Bunches in the LHC*, CERN LHC Project Report 39 (1996).
169. W. Herr et al., *Head-On Beam-Beam Tune Shifts with High Brightness Beams in the LHC*, CERN-ATS-Note-2011-029 (2011).
170. G. Papotti et al., *Observations of Bunch to Bunch Differences due to Beam-Beam Effects*, Proceedings of IPAC2011, San Sebastián, Spain.
171. B. Goddard, et al.: *Bunch Trains for LEP*, Particle Accelerators 57, 237 (1998).
172. W. Herr et al., *Observations of beam-beam effects at high intensities in the LHC*, In Proceedings of the Particle Accelerator Conference 2011, San Sebastian (2011).
173. W. Herr et al., *Head-on beam-beam interactions with high intensities and long range beam-beam studies in the LHC*, CERN-ATS-Note-2011-058 (2011).
174. W. Herr, D. Kaltchev et al., *Large Scale Beam-beam Simulations for the CERN LHC using distributed computing*, Proc. EPAC 2006, Edinburgh (2006).
175. H. Grote, W. Herr: *Self-Consistent Orbits with Beam-Beam Effects in the LHC*, In Proceedings of the 2001 Workshop Beam-Beam Effects, FNAL, 25-27/6/2001.
176. W. Herr, *Consequences of Periodicity and Symmetry for the Beam-Beam Effects in the LHC*, CERN LHC Project Report 49 (1996).
177. W. Kozanecki and J. Cogan, private communication (2011).
178. R. Bartoldus, *Online determination of the LHC Luminous Region with the ATLAS High Level trigger*, TIPP 2011, Int. Conf. on Tech. and Instr. in Particle Physics, Chicago (2011).
179. Y. Alexahin, et al., *Coherent Beam-Beam Effects in the LHC*, Proc. HEACC 2001, Tsukuba, Japan, CERN LHC Project Report 466 (2001).
180. K. Hirata: Nucl. Instrum. Meth. A 269, 7 (1988).
181. A. Piwinski: *Observation of Beam-Beam Effects in PETRA*, IEEE Trans. Nucl. Sci. NS-26, 4268 (1979).
182. R.E. Meller, R.H. Siemann: IEEE Trans. Nucl. Sci. NS-28, 2431 (1981).
183. K. Yokoya, et al., *Tune Shift of Coherent Beam-Beam Oscillations*, Part. Acc. 27, 181 (1990).
184. Y. Alexahin, *On the Landau Damping and Decoherence of Transverse Dipole Oscillations in Colliding Beams*, Part. Acc. 59, 43 (1996).
185. Y. Alexahin, *A Study of the Coherent Beam-Beam Effect in the Framework of the Vlasov Perturbation Theory*, Nucl. Instrum. Meth. A 380, 253 (2002).
186. W. Herr et al., *A Hybrid Fast Multipole Method Applied to Beam-Beam Collisions in the Strong-Strong Regime*, Phys. Rev. ST Accel. Beams 4, 054402 (2001).
187. J.P. Koutchouk, *ISR Performance Report - A Numerical Estimate of the Coherent Beam-Beam Effect in the ISR*, CERN ISR-OP/JPK-bm (1982).
188. W. Fischer, et al., *Observation of Coherent Beam-Beam Modes in RHIC*, BNL C-AD/AP/75 (2002).

189. V.D. Shiltsev, et al., *Compensation of Beam-Beam Effects in the Tevatron Collider with Electron Beams*, In Proceedings of the Particle Accelerator Conference 1999, New York, p. 3728.
190. V.D. Shiltsev, et al., *Considerations on Compensation of Beam-Beam Effects in the Tevatron with Electron Beams*, Phys. Rev. ST Accel. Beams 2, 071001 (1999).
191. J.P. Koutchouk, et al., *Correction of the Long-Range Beam-Beam Effect in LHC using Electro-Magnetic Lenses*, In Proceedings of the Particle Accelerator Conference 1999, New York, p. 1681.
192. R. Talman, *A Proposed Möbius Accelerator*, Phys. Rev. Lett. 74, 1590 (1995).
193. S. Henderson et al., *Investigation of the Möbius Accelerator at CESR*, CBN 99-5 (1999).
194. LHC Injectors Upgrade Technical Design Report – Volume I: Protons, edited by J. Coupard et al., CERN-ACC-2014-0337

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 5

Interactions of Beams with Surroundings



M. Brugger, H. Burkhardt, B. Goddard, F. Cerutti, and R. G. Alia

With the exceptions of Synchrotron Radiation sources, beams of accelerated particles are generally designed to interact either with one another (in the case of colliders) or with a specific target (for the operation of Fixed Target experiments, the production of secondary beams and for medical applications). However, in addition to the desired interactions there are unwanted interactions of the high energy particles which can produce undesirable side effects. These interactions can arise from the unavoidable presence of residual gas in the accelerator vacuum chamber, or from the impact of particles lost from the beam on aperture limits around the accelerator, as well as the final beam dump. The wanted collisions of the beams in a collider to produce potentially interesting High Energy Physics events also reduces the density of the circulating beam and can produce high fluxes of secondary particles.

All of these unwanted interactions affect the performance of an accelerator, in desorption of gas from the vacuum system, in reduced lifetime of the circulating beam, in reduction of the collider luminosity and in background for the detectors. In this chapter the basic physical phenomena of particle interactions with matter are described, together with the techniques used to simulate the interaction with matter. The different types of particle interactions with the surroundings are elaborated in the context of their adverse effects on the accelerator performance and the mitigation measures. A full description of the effects and mitigation measures associated with the vacuum systems of accelerators is given separately in Chap. 8.

M. Brugger (✉) · H. Burkhardt (✉) · B. Goddard (✉) · F. Cerutti · R. G. Alia
CERN (European Organization for Nuclear Research), Meyrin, Genève, Switzerland
e-mail: Markus.Brugger@cern.ch; Helmut.Burkhardt@cern.ch; Brennan.Goddard@cern.ch

5.1 The Interactions of High Energy Particles with Matter

Modern accelerators use leptons or hadrons (including ions) and have beam energies spanning the MeV to TeV range. Therefore, the capability of modelling particle interactions and showers from these energy ranges down to thermal energies is essential during all stages of the lifetime cycle of an accelerator, from the accelerator design through operation to its final decommissioning.

Before briefly discussing general aspects of hadronic and electromagnetic showers, a short overview is given of important ingredients for accelerator applications:

- energy deposition for the design of accelerator components and elements (e.g., collimators, magnets);
- particle fluences as a function of energy, angle and position (e.g., detector or radiation damage and radiation to electronics studies);
- distribution of particle interactions, inelastic interaction density (e.g., for tracking and loss pattern studies);
- residual nuclei production and generation of radioactive isotopes by beam interactions (e.g., radiation protection aspects like air activation or equipment handling).

To allow for calculations of related quantities, the underlying physical processes must not only be well understood and described in models, but also included in calculation codes able to yield allow reliable estimates within a reasonable time.

5.1.1 *Basic Physical Processes in Radiation Transport Through Matter*

Hadron and electromagnetic showers are very complex phenomena, whose description in terms of basic physical interactions requires a detailed and complex modelling.

As soon as the energy of a primary hadron beam exceeds a few tens of MeV, inelastic interactions start playing a major role and generate secondary particles that will have enough energy to trigger further interactions giving rise to hadronic showers. Furthermore, whenever the beam energy is high enough that significant pion production can occur, an increasing fraction of the energy will be transferred from the hadronic to the electromagnetic part due to meson decay (e.g., π^0 decaying into a gamma pair). The pion production threshold for nucleons interacting with stationary nucleons is around 290 MeV.

Therefore, energetic hadronic showers are always accompanied by significant electromagnetic showers, where the latter ones tend to develop independently without further hadronic particle production (with the exception of electro- and photo-nuclear interactions of lower importance for hadron accelerators, which however have to be considered for lepton accelerators).

While electromagnetic interactions can be described in one coherent theory (QED), the same does not apply to hadronic nuclear interactions.

The development of hadron initiated showers is determined both by atomic processes (e.g. ionization, multiple Coulomb Scattering, etc.), which take place very frequently, as well as relatively rare nuclear interactions (both elastic and nonelastic). Electromagnetic showers are determined by the same atomic processes plus additional ones (e.g., Bremsstrahlung, pair production, Compton scattering, etc.) which are specific for electrons, positrons or photons. Nuclear interactions coming from the electromagnetic component usually play a minor role, and whenever the interest is not in the small fraction of hadrons produced by electromagnetic particles they can be safely neglected.

Concerning particle production, energetic (shower) particles are concentrated mainly around the primary beam axis, regardless of their identity. Their ionization as well as the electromagnetic cascades define the core of the energy deposition distribution. At the same time, neutrons (since these are the only neutral hadrons with a long enough lifetime) will dominate at energies where charged particle ranges become shorter than the respective interaction length. In this sense, the energy deposition associated with low energy neutron interactions constitute the peripheral tails of the energy deposition distribution. Most of the interactions are due to particles (mainly neutrons) of moderate energy.

Pions can be only produced by shower particle interactions, so that they are often considered as the real indicator of a high energy cascade. Neutrons, and to a lesser extent protons, are copiously produced also in the final stages of nuclear interactions (e.g., evaporation) down to projectile energies that are comparable to their nuclear binding energy.

As previously mentioned, to focus on the relevant processes one usually distinguishes between continuous and discrete (or explicit) processes. This distinction reflects a real physical distinction, between processes which occur very frequently with mean free paths much shorter than particle ranges in matter, and others that, however, are often the dominating ones in determining the shower development.

The most important discrete processes are:

- inelastic nuclear interaction;
- decay;
- elastic nuclear interaction;
- delta-ray production;
- bremsstrahlung;
- annihilation;
- photoelectric effect;
- Compton scattering;
- pair production;
- coherent (Rayleigh) scattering.

In addition to these processes, nuclear interactions with a much lower rate can occur also for photons (as well as with a reduced rate of about 1/137 for electrons

and positrons). Bremsstrahlung radiation can also be produced by “heavy” charged particles even though it is significantly suppressed. Furthermore, charged particles (light and heavy ones) can produce electron-positron pairs.

5.1.2 *Simulation Tools*

In all life cycle stages of an accelerator the use of simulation, notably Monte-Carlo codes, became fundamental. Thanks to the variety of such codes over different particle physics applications and the associated extensive benchmarking with experimental data, the modelling has reached an unprecedented accuracy. Furthermore, most codes allow the user to simulate all aspects of a high energy particle cascade in one and the same run: from the first interaction over the transport and re-interactions (hadronic and electromagnetic) of the produced secondaries, to detailed nuclear fragmentation, the calculation of radioactive decays and even of the electromagnetic shower from such delayed decays.

In the following we give a brief overview of the most used multi-purpose codes around accelerator applications.

5.1.2.1 FLUKA

FLUKA is a general-purpose particle interaction and transport code with roots in radiation protection and respective design and detector studies for high energy accelerators [1, 2]. It comprises all features needed in this area of application, such as detailed nuclear interaction models, full coupling between hadronic and electromagnetic processes and numerous variance reduction options.

The module for hadronic interactions is called PEANUT and consists of a phenomenological description (Dual Parton Model-based Glauber Gribov cascade) of high energy interactions, through a generalized intranuclear cascade, pre-equilibrium emissions as well as evaporation, fragmentation, fission and de-excitation by gamma emission. Interactions of ions are simulated through interfaces with different codes depending on the energy range (DPMJET3 above 5 GeV/n and rQMD-2.4 between 0.125 and 5 GeV/n, while the embedded Boltzmann Master Equation model is applied below 0.125 GeV/n).

The transport of neutrons with energies below 20 MeV is performed by a multi-group algorithm based on evaluated cross section data (ENDF/B, JEFF, JENDL etc.) binned into 260 energy groups, 31 of which in the thermal region. For a few isotopes point-wise cross sections can be optionally used. The detailed implementation of electromagnetic processes in the energy range between 1 keV and 1 PeV is fully coupled with the models for hadronic interactions.

5.1.2.2 GEANT4

GEANT4 is an object-oriented toolkit originally designed to simulate detector responses of modern particle and nuclear physics experiments [3, 4]. It consists of a kernel which provides the framework for particle transport, including tracking, geometry description, material specifications, management of events and interfaces to external graphics systems.

The kernel also provides interfaces to physics processes. In this regard the flexibility of GEANT4 is unique as it allows the user to select freely the physics models which serve best the particular application needs. This freedom comes with high responsibility as the user must ensure that the most adequate models are used for a given problem. Implementations of interaction models exist over an extended range of energies, from optical photons and thermal neutrons to high energy interactions as required for the simulation of accelerator and cosmic ray experiments. In many cases complementary or alternative modelling approaches are offered from which the user can choose.

Descriptions of intranuclear cascades include implementations of the Binary and the Bertini cascade models (the latter significantly reworked and not at all linked to the original Bertini model). Both are valid for interactions of nucleons and charged mesons, the former for energies below 3 GeV and the latter below 10 GeV. The Intranuclear Cascade of Liège (INCL) is also an usable option. At higher energies (up to 10 TeV), three models are available: a high-energy parameterized model (using fits to experimental data), a quark-gluon string model and the Fritiof fragmentation model, both based on string excitations and decay into hadrons. Nuclear de-excitation models include abrasion-ablation and Fermi-breakup models. Furthermore, heavy ion interactions can also be simulated if the appropriate packages are linked.

The package for electromagnetic physics comprises the standard physics processes as well as extensions to energies below 1 keV, including emissions of X-rays, optical photon transport, etc.

5.1.2.3 MARS15

The MARS15 code system [5, 6] is a set of Monte Carlo programs for the simulation of hadronic and electromagnetic cascades which is used for shielding, accelerator design and detector studies. Correspondingly, it covers a wide energy range: 100 keV–100 TeV for muons, charged hadrons and heavy ions, 1 keV–100 TeV for electromagnetic showers and down to 0.00215 eV for neutrons.

Hadronic interactions above 5 GeV can be simulated with either an inclusive or an exclusive event generator. While the former is CPU-efficient (especially at high energy) and based on a wealth of experimental data on inclusive interaction spectra, the latter provides final states on a single interaction level and preserves correlations. In the exclusive mode, the cascade-exciton model code CEM03, the Quark-Gluon String Model code LAQGSM03 and the DPMJET3 code are

implemented, including models for a detailed calculation of nuclide production via evaporation, fission and fragmentation processes.

Interfaced to MARS, the MCNP4C code handles all interactions of neutrons with energies below 14 MeV. Produced secondaries other than neutrons are directed back to the MARS15 modules for further transport.

5.1.2.4 MCNP

MCNP6 [7] originates from the Monte Carlo N-Particle transport (MCNP)-family of neutron interaction and transport codes and, therefore, features one of the most comprehensive and detailed description of the related physical processes. The extension to other particle types, including ions and electromagnetic particles, allowed an expansion of the areas of application from purely neutronics to, among others, accelerator shielding design, medical physics and space radiation.

The neutron interaction and transport modules use standard evaluated data libraries mixed with physics models where such libraries are not available. The transport is continuous in energy and includes all features necessary for reactor simulations, including burn-up, depletion and transmutation. Different generalized intranuclear cascade codes can be linked to explore different physics implementations, such as CEM03, INCL4 and ISABEL. They either contain fission-evaporation models or can be coupled to such models (i.e., ABLA) allowing detailed predictions for radio-nuclide production. While the intranuclear cascade codes are limited to interaction energies below a few GeV, a link to the Quark-Gluon String Model code LAQGSM03 extends this energy range to about 800 GeV. The latter code also allows the simulation of ion interactions.

5.1.2.5 PHITS

The Particle and Heavy-Ion Transport code System PHITS (see [8] and [9] and references therein) was among the first general-purpose codes to simulate the transport and interactions of heavy ions in a wide energy range, from 10 MeV/nucleon to 100 GeV/nucleon. It is based on the high-energy hadron transport code NMTC/JAM which was extended to heavy ions by incorporating the JAERI Quantum Molecular Dynamics code JQMD.

Below energies of a few GeV hadron-nucleus interactions in PHITS are described through the production and decay of resonances while at higher energies (up to 200 GeV) inelastic hadron-nucleus collisions proceed via the formation and decay of so-called strings which eventually hadronize through the creation of (di)quark-anti(di)quark pairs. Both are embedded into an intranuclear cascade calculation. Nucleus-nucleus interactions are simulated, within a molecular dynamics framework, based on effective interactions between the two self-binding system of nucleons.

The generalized evaporation model GEM treats the fragmentation and de-excitation of the spectator nuclei and includes 66 different ejectiles (up to Mg) as well as fission processes. The production of radioactive nuclides, both from projectile and target nuclei, thus follows directly from the microscopic interaction models.

The transport of low energy neutrons employs cross sections from evaluated nuclear data libraries such as JENDL below 20 MeV. Electromagnetic interactions are simulated based on the EGS5 code in the energy range between 1 keV and 1 TeV.

Due to its capability to transport nuclei PHITS is frequently applied in ion-therapy and space radiation studies. The code is also used for general radiation transport simulations, such as in the design of spallation neutron sources.

5.1.2.6 Simulation Uncertainties

Depending on the complexity of Monte-Carlo simulations, the size of geometries and the energy range involved, calculations can carry considerable uncertainties of various sources which are in many cases difficult to evaluate. While statistical uncertainties are generally below a few percent thanks to the available computing power, systematic errors are in most cases very difficult, or even impossible, to predict in an accurate way.

The main sources of error are:

- Error due to the physics modelling: e.g., in the uncertainty in cross sections, especially at modern accelerators operating at very high energies or applications sensitive to other uncertainties in the modelling used in the simulation code. One can usually expect up to few 10% uncertainty on integral quantities, while for multi differential quantities the uncertainty can be much worse.
- Further uncertainties due to the assumptions used in the description of the geometry and of the materials under study. Usually it is difficult to quantify this uncertainty and experience shows that a factor of 2 can be taken as a safe limit for general calculations. For special cases, even in case of rather complex geometries, but when the design is implemented to a very detailed level, the latter can be reduced to about 10–20%, but rarely significantly below.
- Typical for accelerator applications, additional errors appear when having beams grazing at small angles to surfaces, where either the surface roughness is not taken into account, small misalignments can have large effects, or where one is interested in scattering effects over large distances. Therefore, especially for the latter, a safety factor of 2–3 has to be considered.

Only a detailed case-by-case analysis and careful evaluation with monitoring data can reduce the above to very low levels. Such studies are continuously done by code developers, as well as core code users and nicely show the possible high-accuracy reached with modern Monte-Carlo codes (see for instance [10]).

5.1.3 Practical Shielding Considerations

In most of the cases around high-energy hadron accelerator operation, the particle cascades originate from beam interactions due to four basic source terms:

- beam–beam interactions (at and close to the experiments);
- beam–residual gas interactions (all along the accelerator);
- direct losses: beam cleaning at collimator locations or beam dump;
- spurious losses (random locations around the accelerator).

The emerging secondary particle cascade is then again defining a multi particle-type and energy spectrum, interacting with the various materials around the accelerator. Therefore, in order to reduce the radiation levels and the corresponding fluences, shielding material can be employed to initiate and absorb showers. If the shielding material is thick, the cascades continue until most of the charged particles and photons have been absorbed, except for neutrons and secondary photons.

The following aspects have to be considered in the shielding conceptual design:

5.1.3.1 Radiation Attenuation

The propagation of high-energy hadrons features an exponential decrease due to nuclear reactions, while their energy loss, in case they are charged, is mostly due to ionization. The latter accounts for the majority of the energy loss in electromagnetic showers and for about 2/3 of the energy deposited in hadronic showers. Most of the other 1/3 of the hadronic energy is carried away by neutrons. Being neutral these are not affected by ionization losses, are decoupled from the rest of the shower, are subject to elastic and inelastic collisions and are considerably more penetrating than the charged component. Hadronic showers above 100 MeV progress through a variety of hadronic and nuclear processes that result in secondaries that are predominantly pions (π^+ , π^- , π^0), followed in importance by nucleons (protons, neutrons), strange mesons/baryons and photons. The π^0 component, appearing whenever hadron energies are above the pion production threshold, is particularly important, since it decays immediately to two photons producing electromagnetic showers and shifting the shower energy from the hadronic to the electromagnetic sector. Hadrons above few tens of MeV undergo nuclear reactions that increase the neutron multiplicity. The spallation process breaks the nucleus into a few large fragments. Additional neutrons may also be produced just during this process, as well as by the subsequent evaporation from these excited fragments.

Since the fragments have large mass (M) and total charge (Z) and ionize, they will be stopped quickly in dense shield materials.

5.1.3.2 Shielding of Electromagnetic Showers

As mentioned earlier, electromagnetic showers in the multi-GeV range develop through successive bremsstrahlung and pair production. The particles involved are photons, electrons and positrons that in dense materials have typical radiation lengths in the centimeter range. As a result, high-energy electromagnetic showers are halted by a typical concrete shielding wall having thickness of the order of 1 m or by a tungsten layer of a few centimeters. It has to be noted that secondary neutrons are also generated through photo-production, thus associated shielding issues have to be correctly considered.

5.1.3.3 Shielding of Neutrons

The elastic scattering cross section of neutrons on nuclei is large at all energies and has a role in attenuating them. The scattered neutron will lose energy on every elastic collision, especially if the target nucleus is light and so carries away a larger fraction of the energy as it recoils. Therefore, the presence of hydrogen is most effective in reducing the neutron kinetic energy. Special care must be taken with respect to:

- neutrons streaming through ventilation channels or other penetrations, where intermediate and low-energy neutrons readily scatter from the wall. Propagation down the channel is thus possible via a series of ‘reflections’ even if the channel is not in the original direction of the neutron;
- the thermalization of neutrons, in the presence of certain elements with high capture cross-sections (e.g., Boron or Cobalt) that effectively clean the neutron field but can induce relevant secondary effects (e.g., electronic damage or activation issues).

Monte Carlo calculations aimed at shielding design optimization may require dedicated biasing techniques to overcome the associated statistical convergence challenges.

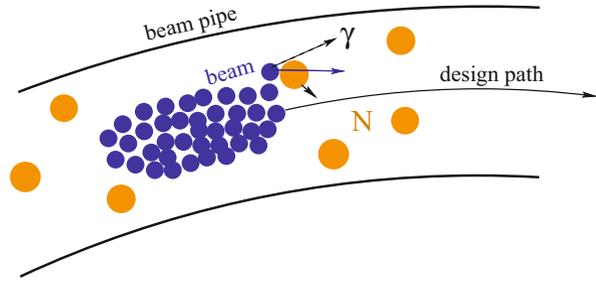
5.2 Lifetimes, Intensity and Luminosity

Particles which circulate on stable orbits in an accelerator within the geometrical and energy acceptance can get lost as a result of collisions with other particles. This leads to a decrease of the number of particles N circulating with time t .

The beam intensity lifetime τ is the inverse of the total loss rate dN/dt , normalized to the actual intensity N (at the time t)

$$\frac{1}{\tau} = -\frac{1}{N} \frac{dN}{dt}. \quad (5.1)$$

Fig. 5.1 Schematic view of beam collisions with the nuclei (N) of the rest-gas



A constant lifetime corresponds to an exponential decay of the intensity with time t proportional to $\exp(-t/\tau)$.

The intensity and beam lifetime can be monitored by measuring the beam current I which is directly proportional to the intensity, $I = Nef$, where e is the charge of the particle and f the revolution frequency in a ring [11].

We can distinguish between several types of collision processes.

- Beam-gas scattering: collisions of beam particles with particles from the residual gas in the evacuated beam pipe, sketched in Fig. 5.1.
- Thermal photon scattering: inverse Compton scattering of electrons or positrons with photons from black-body radiation in the beam pipe; relevant for beam energies > 10 GeV
- Intra-beam scattering: collisions or close encounters with other particles in the same beam; most relevant at lower energies and for very dense beams
- Quantum lifetime: particles leaving the acceptance by quantum fluctuations in synchrotron radiation
- Colliding beams: particle collisions or close encounters when beams cross in colliders

Estimates for the first two types which involve collisions with particles outside the beam will be given below.

The losses from the different loss mechanism have to be added. This implies, that the corresponding lifetime contributions add reciprocally

$$\frac{1}{\tau} = \sum_i \frac{1}{\tau_i}$$

An example of observed lifetimes in LEP is listed in Table 5.1, where losses from intrabeam scattering were negligible.

The single beam lifetime is observed before beams are brought into collisions. In LEP this was dominated by the thermal photon scattering. LEP was generally operated with sufficient aperture and over-voltage from the RF-acceleration system, so that losses from the quantum lifetime were negligible. More details on lifetimes observed in LEP with a discussion of quantum lifetime and small, but non-negligible effects which made lifetimes longer than originally anticipated can be found in [12].

Table 5.1 Example of the beam lifetime in LEP, fill 4163 from 14-Sep-1997 at Eb = 91.5 GeV

Component	Lifetime τ in hours
Thermal Compton	50
Beam Gas, 0.3 nTorr CO	160
Combined, single beam	38
e^+e^- collisions, $\sigma = 0.21$ barn	8.6
Total	7

5.2.1 Beam-Gas

From ideal gas theory, the density ρ_m in terms of molecules or atoms per unit volume is

$$\rho_m = \frac{p}{kT}. \quad (5.2)$$

Multiplying ρ_m with the cross section σ for beam-gas collisions, gives us the collision probability per unit length

$$P_{coll} = \sigma \rho_m.$$

Further multiplication with the velocity of the beam particles $v = \beta c$ gives us the collision rate per unit time

$$\beta c \sigma \rho_m = \frac{1}{\tau} \quad (5.3)$$

which corresponds to the inverse lifetime, if σ is the cross section for collisions which lead to a loss of the beam particles. This will be the case for inelastic scattering processes and for elastic scattering in which the scattering angle is larger than the angular acceptance.

For numerical estimates in this section we take $\rho_m = 3.26 \times 10^{13}$ molecules/m³ which corresponds to a pressure of $p = 1$ Torr = 1.33×10^{-7} Pa at room temperature ($T = 296.15$ K = 23°C) and can be considered as typical number for good vacuum conditions. At high energy we have $\beta \approx 1$. For a cross section of $\sigma = 1$ b (one barn, where $1\text{b} = 10^{-28}$ m²), we obtain a beam-gas lifetime $\tau = 284$ h.

The main beam-gas scattering processes are shown in Fig. 5.2.

eN Scattering Relevant for Electron Rings

The elastic cross section for eN scattering scales strongly with energy (with $1/\gamma^2$) and scattering angle $1/\theta^4$. Elastic scattering is mostly relevant as a halo production process for lower energy rings and becomes negligible for lifetime estimates for high energy electron rings.

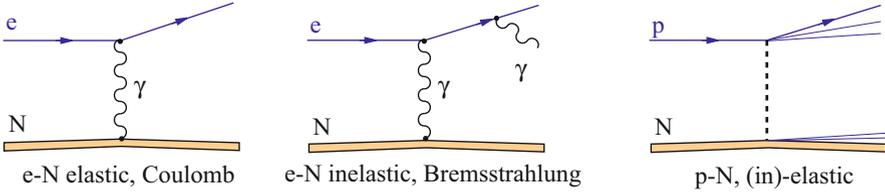


Fig. 5.2 Beam-gas scattering processes; elastic and inelastic eN scattering relevant for e^+ , e^- machines is shown on the left and pN scattering relevant for proton machines on the right

At high energy, the dominating beam-gas process for electron rings is the inelastic scattering or bremsstrahlung in which the incident electron interacts with the field of the residual gas nucleus and radiates a photon.

The high energy cross section for eN scattering can be written in good approximation in dependently of the electron energy as [13]

$$\sigma_{eN} = 4\alpha r_e^2 Z(Z+1) \log(287/\sqrt{Z}) \left(-\frac{4}{3} \log k_{\min} - \frac{5}{6} + \frac{4}{3} k_{\min} - \frac{k_{\min}^2}{2} \right), \quad (5.4)$$

where k_{\min} is the fractional energy loss or minimum photon energy in units of the electron energy, α the fine-structure constant ($1/137$) and Z the atomic number (or number of protons). We can see that the cross section scales with $Z(Z+1)$. Numerical values obtained from Eq. 5.4 for $k_{\min} = 0.01$ are shown in Table 5.4.

pN Scattering Relevant for Electron Rings

At high energies ($p_{\text{lab}} > 10$ GeV), the pN cross section is mostly inelastic ($> 80\%$). It depends only weakly on the proton energy and scales approximately with the atomic mass $\propto A^{2/3}$ [14], as can be expected for the cross-section of a sphere. Numerical values are listed in Table 5.2. We can see that pN cross sections are much smaller ($\sim 10\times$) than eN cross sections. Good beam-gas lifetimes in proton machines can be several 10^3 h compared to typically 10^2 h in high energy electron machines.

5.2.2 Thermal Photons

Even a perfectly evacuated beam pipe remains “filled” with photons from black body radiation which can be relevant as source of backgrounds and reduction of beam lifetime as first pointed out by V. Telnov in 1987 [15]. The photon density from black-body radiation is

Table 5.2 Numerical values for σ_{eN} for an energy loss of at least 1% and for σ_{pN} , the pN cross section at high energy ($p_{lab} = 0.01-10$ TeV)

Gas	σ_{eN} b	σ_{pN} b
H ₂	0.28	0.08
He	0.39	0.19
CH ₄	3.02	0.43
H ₂ O	4.38	0.40
N ₂	6.47	0.56
CO	6.56	0.56
CO ₂	10.7	0.87
Ar	17.8	0.60

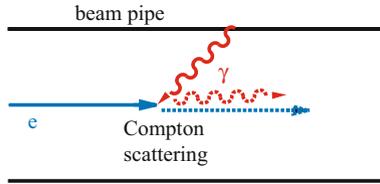


Fig. 5.3 Schematic view of the inverse Compton scattering with thermal photons. A high energy beam particle collides and loses energy to a low energy photon radiated from the beam pipe by black body radiation

$$\rho_\gamma = 8\pi \left(\frac{kT}{hc}\right)^3 \underbrace{\int_0^\infty \frac{x^2}{e^x - 1} dx}_{= 2.404} \tag{5.5}$$

where T is the absolute temperature, and k, h, c the Boltzmann, Planck constants and the speed of light. For a beam pipe at room temperature (23 °C), we get a photon density of $\rho_\gamma = 5.3 \times 10^{14} \text{ m}^{-3}$ which is an order of magnitude higher than the typical residual gas molecular densities ρ_m , considered in this chapter for beam gas estimates (Fig. 5.3).

The lifetime from thermal photon scattering is

$$\tau_t = \frac{1}{\underbrace{\rho_\gamma c \sigma_C}_{\sim 26 \text{ h}}} \frac{1}{f_{loss}} \tag{5.6}$$

where ρ_γ is the photon density, σ_C the Compton cross section (at high energy ~ 0.665 barn) and f_{loss} the fraction of the e^\pm lost after collision. Numerical values calculated using the program described in [16] are given in Table 5.3. We can see that thermal photon scattering at room temperature becomes only relevant for electron beam energies above 10 GeV.

Table 5.3 Thermal photon scattering

E_b (GeV)	f_{loss} (%)	τ_l (h)
10	0.3	9000
45.6	19	144
100	39	72
250	61	49

Fraction of beam particles lost and lifetime, for various electron beam energies E_b . At room temperature and for an energy acceptance of 2%

5.2.3 Luminosity Lifetime

In analogy to Eq. 5.1, the luminosity L lifetime is defined as

$$\frac{1}{\tau_L} = -\frac{1}{L} \frac{dL}{dt}. \quad (5.7)$$

The luminosity for colliding beams depends on the product of the colliding beam intensities, or N^2 in case of equal beam intensities. At constant beam sizes, the luminosity decreases as $dN^2/dt = 2dN/dt$, so that the luminosity lifetime is half of the intensity lifetime $\tau_L = \tau/2$. For operation at the beam-beam limit as is typical for high luminosity e^+e^- storage rings, beam sizes increase with intensity such that the beam-beam parameter is constant, and $L \propto N$ and therefore $\tau_L = \tau$. In proton machines, beam sizes tend to increase with time due to intrabeam scattering, noise and vibrations, resulting in luminosity lifetimes $\tau_L < \tau/2$.

5.3 Experimental Conditions

The most important performance parameters for colliders for particle physics are

- the beam energy; higher beam energies allow to study smaller distances and to produce new heavier particles;
- high luminosity to obtain sufficient collisions rates to observe new processes or to improve the measurement precisions.

It is also essential to provide good experimental conditions for the particle detectors installed around the collision regions. Criteria for good experimental conditions are

- low backgrounds;
- good knowledge and stability of beam parameters;

- minimize the risk of damage of the detectors by beam loss and irradiation;
- minimize the size of the beam pipe in the detector region to allow for the installation of vertex detectors as close as possible to the interaction region;
- maximize the space available for the detector and the solid angle coverage down to low angles close to the beam.

These are rather conflicting requirements. Minimizing the size of the beam pipe for installation of sensitive vertex detectors close to the beam pipe will increase the background rates hitting the detector region and increase the risk of damage by beam loss. The requirements for maximum space and solid angle coverage for the detectors are also in conflict with the requirements of the accelerator to maximize luminosity by

- allowing for space close to the interaction point for final focus quadrupoles;
- reducing the β -function at the interaction point, which increases the beam size in the final focus quadrupoles with the risk to create local aperture limits and losses and which limits the space and low angle coverage available for particle detection;
- installation of beam-separators close to the interaction region to allow to fill the machine with many bunches.

It is essential to consider the accelerator and detector requirements together, both during the design stage and also later in the optimization of the running parameters.

Experience shows that it typically takes several years to commission and optimize the performance of a new accelerator. During these first years, it will often be possible to increase the luminosity without compromising on the experimental conditions for the detectors. Detailed simulation and continuous monitoring of the background conditions are important to identify potential limitations, to guide further optimization and to identify the potential for upgrades towards higher luminosities or smaller beam pipes.

Different types of backgrounds and their mitigation with examples from LEP and LHC are now discussed.

5.3.1 Sources of Detector Background and Detector Performance

It is possible to distinguish between two main types of machine induced backgrounds

- backgrounds induced by losses of beam particles;
- background from synchrotron radiation, relevant for high energy e^+ , e^- beams.

Table 5.4 Lifetimes τ and beam loss rates dN/dt in LEP2 and the (nominal) LHC, compared to the bunch crossing rates f_c

	N_{tot}	τ (h)	$-dN/dt$ (Hz)	f_c (Hz)	$-\frac{dN}{dt}/f_c$ (Hz)
LEP2	3.2e12	5	1.8×10^8	4.5×10^3	4×10^4
LHC	6.5e14	10	1.8×10^{10}	3.2×10^7	6×10^2

Machine induced backgrounds by particle loss are relevant for all (circular and linear) colliders. Even under good conditions, millions of particles will be lost per second, exceeding by several orders of magnitude the beam crossing rates, see Table 5.4. A minimum requirement is that only a very small fraction of these particles gets lost close to the detector, such that the background rate in the interaction region is small compared to the bunch crossing rates.

To achieve this one has to assure

- good vacuum conditions in the region around the detectors, in order to minimize local losses by beam-gas scattering in the detector region;
- that there is no aperture limitation which would concentrate losses close to the detectors.

the latter imposes limits on the minimum β in the interaction region and hence the maximum luminosity. A standard method to reduce backgrounds from particle losses is to use aperture limiting collimators to remove high amplitude halo particles. These should be placed far from the experiments, to minimize the probability that secondary particles scattered off the collimators reach the experiments.

For beam energies above about 50 GeV, the production of secondary muons in electromagnetic showers has to be taken into account. High energy muons are hard to shield. Muon production and shielding is taken into account in the design studies for high energy linear colliders [17, 18].

The final focus quadrupoles placed around the interaction regions of colliders generate a high local chromaticity which can lead to a concentration of losses of off-momentum particles into the detectors. LEP2 was equipped with momentum collimators in the dispersion suppressors around all experimental sections to reduce the flux of off-momentum particle generated by e^+e^- collisions, bremsstrahlung in the residual gas and thermal photon scattering.

5.3.2 Synchrotron Radiation Background

The energy spectrum of the synchrotron radiation photons radiated by a high energy electron (or positron or proton) travelling on a circular orbit of radius ρ

is characterised by the critical energy

$$E_c = \frac{3}{2} \hbar c \frac{\gamma^3}{\rho}. \quad (5.8)$$

The number of photons radiated in a bending magnet of length L and bending radius ρ is ($\alpha = e^2 / 4\pi\epsilon_0\hbar c$ is the fine-structure constant and γ the Lorentz factor of the particle) and the energy loss U_0 per beam particle and turn are

$$N_{\gamma,L} = \frac{5\alpha\gamma L}{2\sqrt{3}\rho}, \quad U_0 = \frac{4\pi\alpha\hbar c \gamma^4}{3\rho} \quad (5.9)$$

The photon energy increases with γ^3 and the energy loss in synchrotron radiation over one turn as γ^4 . A practical limit was reached with electrons at LEP at beam energies around 100 GeV, corresponding to a Lorentz factor $\gamma \approx 2 \times 10^5$ when 3% of the particle energy was lost on a single turn, while for the 1836 times heavier protons synchrotron radiation only becomes noticeable at TeV energies. Numerical values for the main synchrotron radiation parameters for several e^+e^- colliders, the LHC with protons at 7 TeV and the proposed FCC colliders[23, 24] are shown in Table 5.5.

The normalised quadrupole power spectrum (for flat beams as typical for e^+ , e^- colliders) is

$$s_d(k) = \frac{9\sqrt{3}}{8\pi} k \int_k^\infty K_{5/3}(s) ds, \quad k = \frac{E}{E_{\text{cr}}}.$$

Table 5.5 Synchrotron radiation parameters; p is the beam-momentum, ρ the bending radius of the main dipole magnets

Machine	p (GeV/c)	γ	ρ (m)	E_c	N_γ	U_0	N_{tot}	P_{tot}
DaΦne	0.51	998	97.7	457 eV	66	9.3 keV	2.1×10^{13}	98 kW
PEP-II HER	9.0	17,613	163	9.9 keV	1166	3.57 MeV	4.5×10^{13}	3.4 MW
PETRA	23.4	45,792	192	148 keV	3031	138 MeV	1.0×10^{12}	3.0 MW
TRISTAN	32.	62,622	244	298 keV	4144	380 MeV	8.8×10^{11}	5.3 MW
LEP1	45.6	89,237	3026	70 keV	5906	126 MeV	1.9×10^{12}	0.4 MW
LEP2	94.5	184,932	3026	619 keV	12,239	2.3 GeV	3.2×10^{12}	13 MW
FCC-ee,Z	45.6	89,237	10,190	22 keV	5973	36 MeV	2.8×10^{15}	50 MW
FCC-ee,tt	182.5	357,144	10,190	1323 keV	23,636	9.19 GeV	1.1×10^{13}	50 MW
LHC	7000	7460.5	2784	44.1 eV	494	6.7 keV	6.5×10^{14}	7.8 kW
FCC-hh	50,000	53,289	10,572	4.24 keV	3733	4.6 MeV	1.1×10^{15}	2.5 MW

N_γ , U_0 are the number of photons radiated and the energy loss per beam particle and turn. N_{tot} , P_{tot} are the number of beam particles and the power radiated in the accelerator

Half of the power is radiated below the critical photon energy.

Quadrupoles can be considered as bending magnets which increase in strength with the distance from the magnet axis. The equations given above also hold for quadrupoles for Gaussian beams if we take as ρ the bending radius of the quadrupole at 1σ offset from the beam axis [19]. The normalised quadrupole power spectrum is

$$s_q(k) = \frac{9\sqrt{3}}{8\pi} k \int_0^\infty (1 - \operatorname{erf}(k/\sqrt{2}s)) K_{5/3}(s) ds, \quad k = \frac{E}{E_{\text{cr},1\sigma}},$$

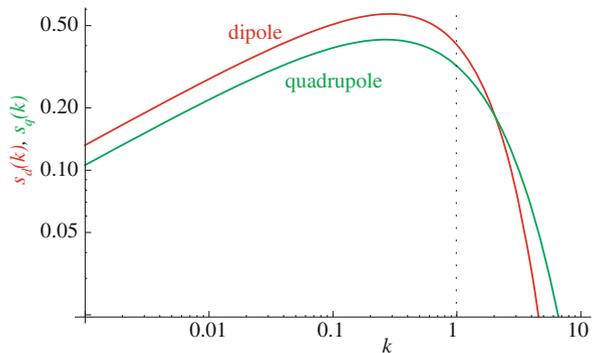
and is shown in Fig. 5.4 together with the spectrum for a dipole.

Averaged over the ring, the synchrotron power radiated in the quadrupoles remains usually very small compared to the power radiated in the main dipole magnets. Due to the vicinity and strength of the quadrupoles around the experiments, it is mandatory to include these in background estimates and also important to consider the possibility of non-Gaussian tails which increase the synchrotron radiation from quadrupoles.

Background estimates for synchrotron radiation depend critically on design details: the magnet lattice, beam pipe geometry, materials and beam parameters. Estimates were often done using dedicated “home-grown” programs with ad-hoc interfaces between separated simulations of the accelerator components, synchrotron radiation generation and simulation of the interactions in the detectors. More recently it has become feasible with the programs BDSIM and MDISim, both based on GEANT4, to perform more flexible integrated simulations which include all relevant components and processes [21, 22].

We will now shortly look at LEP2 which had the strongest synchrotron radiation of all colliders and still tolerable background levels for the detectors. The amount of synchrotron radiation in LEP was huge, particularly at LEP2 energies: about 6×10^{20} photons were emitted per second and a power of 18 MW lost to synchrotron radiation. The experiments had to be very well screened using a sophisticated collimation system with about 100 movable collimators and in addition fixed masks close the experiments [20]. The typical layout of the collimators in a straight section,

Fig. 5.4 Normalised power spectra for synchrotron radiation in a dipole and a quadrupole



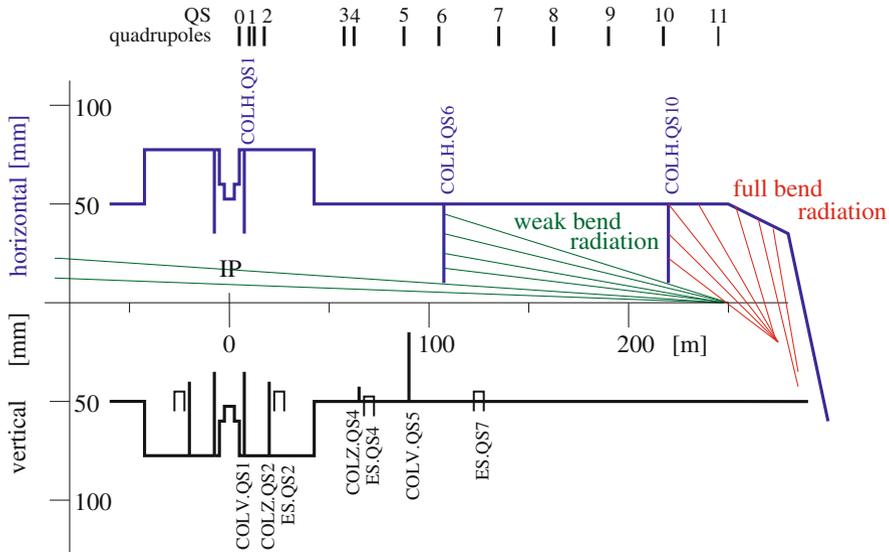


Fig. 5.5 Schematic layout of a straight section at an interaction point (IP) of LEP in the horizontal (top) and vertical (bottom) planes. Shown are the locations of the quadrupoles (QS), electrostatic separators (ES) and collimators (COLH, COLV, COLZ). The solid lines mark the inner vacuum chamber radii

where only scattered synchrotron light could reach the detectors, is shown in Fig. 5.5. The synchrotron radiation spectrum is broad and photons down to about 20 keV can leave the beam pipe. The lower energy X-ray radiation can undergo low angle (multiple) reflection. The strong radiation from the main dipoles of LEP was intercepted close to the arcs, with collimators located between 100 and 220 m away from the interaction point, before the photons could be scattered at low angle towards the experiments. To reduce the radiation shining into the straight sections further, the first dipoles in the arcs had only 10% of the field of the normal arc dipoles.

Local masks were installed about 2.4 m from the interaction points to improve the shielding of the experiments from the increased synchrotron radiation at LEP2. The collimators and masks close to the interaction point were however also a source of scattered background particles. The surface material and inclination of the masks was optimized to minimize the scattering towards the experiment: the masks were made of tungsten and the surface coated with silver and copper layers to reduce the emission of fluorescence photons.

The background photons observed in the detectors originated mainly from synchrotron radiation in the last quadrupoles and was backscattered into the experiment from local collimators. The bunch crossing rate in LEP was about 45 kHz and typically only a few background photons were recorded per bunch crossing in the large wire chambers of the LEP detectors. There was no problem with detector

occupancy, but the currents drawn in the gas-chambers were reported to be not too far from the tolerable limit. The experience gained with LEP has been essential for the design studies for a possible future circular lepton collider at CERN [23, 25].

References

1. G. Battistoni et al., *Overview of the FLUKA code*, *Annals of Nuclear Energy* 82 (2015) 10–18.
2. T.T. Bohlen et al., *The FLUKA Code: Developments and Challenges for High Energy and Medical Applications*, *Nuclear Data Sheets* 120 (2014) 211–214.
3. S. Agostinelli et al., *GEANT4-A simulation toolkit*, *Nuclear Instruments and Methods in Physics Research A* 506 (2003) 250–303.
4. J. Allison et al., *Recent developments in GEANT4*, *Nuclear Instruments and Methods in Physics Research A* 835 (2016) 186–225.
5. N.V. Mokhov and C.C.James, *The MARS Code System User's Guide, Version 15*, (2016), <https://mars.fnal.gov>
6. N.V. Mokhov and S.I. Striganov, *MARS15 Overview*, *Proceedings of the Hadronic Shower Simulation Workshop 2006*, Fermilab 6–8 September 2006, M. Albrow, R. Raja eds., AIP Conference Proceeding 896 (2007).
7. C.J. Werner ed., *MCNP User's Manual, Version 6.2*, Los Alamos National Laboratory report, LA-UR-17-29981 (2017).
8. T. Sato et al., *Features of Particle and Heavy Ion Transport code System (PHITS) version 3.02*, *J. Nucl. Sci. Technol.* 55 (2018) 684–690.
9. Y. Iwamoto et al., *Benchmark study of the recent version of the PHITS code*, *J. Nucl. Sci. Technol.* 54 (2017) 617–635.
10. A. Lechner et al., *Validation of energy deposition simulations for proton and heavy ion losses in the CERN Large Hadron Collider*, submitted to *Physical Review Accelerators and Beams* (2019), <https://doi.org/10.1103/PhysRevAccelBeams.22.071003>.
11. A. Peters, H. Schmickler, K. Witternburg (Editors), “Proceedings of Workshop on DC Current Transformers and Beam-Lifetime Evaluations, Lyon 2004”, [CARE-Note-2004-023-HHH](#)
12. H. Burkhardt, “Beam Lifetime and Beam Tails in LEP”, *Proc. e⁺e⁻ factories*, KEK 1999, [CERN-SL-99-061-AP \(1999\)](#)
13. Y.-S. Tsai, “Pair production and Bremsstrahlung of charged leptons,” *Rev. Mod. Phys.* **46** (1974) 815–851.
14. J. Letaw, R. Silberberg, and C. Tsao, “Proton-nucleus inelastic cross sections: an empirical formula for $E > 10$ MeV”, *The Astrophys. Journ.Suppl.Ser.* 51, (1983), pp 271–276
15. V. I. Telnov, “Scattering of electrons on thermal radiation photons in electron - positron storage rings”, *Nucl. Instrum. Meth.* A260 (1987) 304,
16. H. Burkhardt, “Monte Carlo Simulation of Beam Particles and Thermal Photons”, July, 1993. [CERN SL Note 93-73 \(OP\)](#).
17. I. Agapov, H. Burkhardt, D. Schulte, A. Latina, G. A. Blair, S. Malton, and J. Resta-López, “Tracking studies of the Compact Linear Collider collimation system,” *Phys. Rev. ST Accel. Beams* **12** (Aug, 2009) 081001. [10.1103/PhysRevSTAB.12.081001](https://doi.org/10.1103/PhysRevSTAB.12.081001)
18. H. Burkhardt, G. A. Blair, and L. Deacon, “Muon Backgrounds in CLIC”, [Proc. IPAC 2010](#)
19. E. Keil, “Synchrotron Radiation from a Large electron-Positron Storage Ring”, CERN-ISR-LTD-76-23, June 1976
20. G. von Holtey, A. Ball, et al., “Study of beam-induced particle backgrounds at the LEP detectors”, *Nucl. Instrum. Meth.* A403 (1998) 205–246, [https://doi.org/10.1016/S0168-9002\(97\)01094-2](https://doi.org/10.1016/S0168-9002(97)01094-2).

21. L. Nevay et al., BDSIM: An Accelerator Tracking Code with Particle-Matter Interactions. https://urldefense.proofpoint.com/v2/url?u=https-3A__arxiv.org_abs_1808.10745v1&d=DwIFaQ&c=vh6FgFnduejNhPPD0fl_yRaSfZy8CWbWnIf4XJhSqx8&r=XRE98stjh0DZIXWNYx-jk11s49AfQ9rwMw8TKownaLA&m=nT9ubpnYZV6ASvZRvylWgRRpic9Q8sG0taj0hVsLlIE&s=sVWRORTCWGrRe0NT2jXWKQJUXrvTIg_zydlnZnlwOs&e=
22. H. Burkhardt and M. Boscolo. Tools for Flexible Optimisation of IR Designs with Application to FCC. In *Proc IPAC 2015*, <https://doi.org/10.18429/JACoW-IPAC2015-TUPTY031>.
23. A. Abada et al. FCC-ee: The Lepton Collider. <https://doi.org/10.1140/epjst/e2019-900045-4>. [*Eur. Phys. J. ST*228,no.2,261(2019)].
24. A. Abada et al. FCC-hh: The Hadron Collider. *Eur. Phys. J. ST*, 228(4):755–1107, <https://doi.org/10.1140/epjst/e2019-900087-0>.
25. M. Boscolo, H. Burkhardt, and M. Sullivan. Machine detector interface studies: Layout and synchrotron radiation estimate in the future circular collider interaction region. *Phys. Rev. Accel. Beams*, 20(1):011008, 2017. <https://doi.org/10.1103/PhysRevAccelBeams.20.011008>.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 6

Design and Principles of Synchrotrons and Circular Colliders



B. J. Holzer, B. Goddard, Werner Herr, Bruno Muratori, L. Rivkin, M. E. Biagini, J. M. Jowett, K. Hanke, W. Fischer, F. Caspers, and D. Möhl

6.1 Beam Optics and Lattice Design in High Energy Particle Accelerators

B. J. Holzer

Lattice design in the context we will describe it here is the design and optimization of the principle elements—the lattice cells—of a circular accelerator, and it includes the dedicated variation of the accelerator elements (as for example position and strength of the magnets in the machine) to obtain well defined and predictable parameters of the stored particle beam. It is therefore closely related to the theory of linear beam optics that has been described in Chap. 2 [1].

Coordinated by K. Hanke

B. J. Holzer · B. Goddard · W. Herr · B. Muratori · M. E. Biagini · J. M. Jowett · K. Hanke (✉) · F. Caspers · D. Möhl
CERN (European Organization for Nuclear Research) Meyrin, Genève, Switzerland
e-mail: Bernhard.holzer@cern.ch; Brennan.Goddard@cern.ch; werner.herr@cern.ch; bruno.muratori@stfc.ac.uk; John.Jowett@cern.ch; Klaus.Hanke@cern.ch; Fritz.Caspers@cern.ch

L. Rivkin
Paul Scherrer Institut, Villigen, Switzerland
e-mail: Leonid.Rivkin@psi.ch

W. Fischer
Brookhaven National Laboratory, New York, NY, USA
e-mail: Wolfram.Fischer@bnl.gov

6.1.1 Geometry of the Ring

For the bending force as well as for the focusing of a particle beam, magnetic fields are applied in an accelerator. In principle, electrostatic fields would also be possible but at high momenta (i.e. if the particle velocity is close to the speed of light) the usage of magnetic fields is much more efficient. In its most general form, the force acting on the particles is given by the Lorentz-force

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \quad (6.1)$$

In high energy accelerators, 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 for bending and focusing the particles. Neglecting the \mathbf{E} component therefore in Eq. (6.1), the condition for a circular orbit is defined as the equality of the Lorentz force and the centrifugal force:

$$qvB = \frac{mv^2}{\rho} \quad (6.2)$$

In a constant transverse magnetic field B , the particle will see a constant deflecting force and the trajectory will be a part of a circle, whose bending radius ρ is determined by the particle momentum $p = mv$ and the external B field.

$$\rho = \frac{p}{qB} \quad (6.3)$$

The term $B\rho$ is called beam rigidity. Inside each dipole magnet in a storage ring the bending angle—sketched out in Fig. 6.1—is given by the integrated field strength via

$$\alpha = \frac{\int B ds}{B\rho} \quad (6.4)$$

Requiring a bending angle of 2π for a full circle, we get the condition for the magnetic dipole fields in the ring. In the case of the LHC e.g. for a momentum of $p = 7000 \text{ GeV}/c$ a number of 1232 dipole magnets are needed each having a length of $\sim 15 \text{ m}$ with a B -field of 8.3 T. As a general rule in high energy rings, about 66% (2/3) of the circumference of the machine should be foreseen to install dipole magnets, as they define the maximum particle momentum that can be carried by the machine. This basic dipole structure is completed with focusing elements, beam diagnostic tools etc. and forms the arcs of the ring. They are connected by long

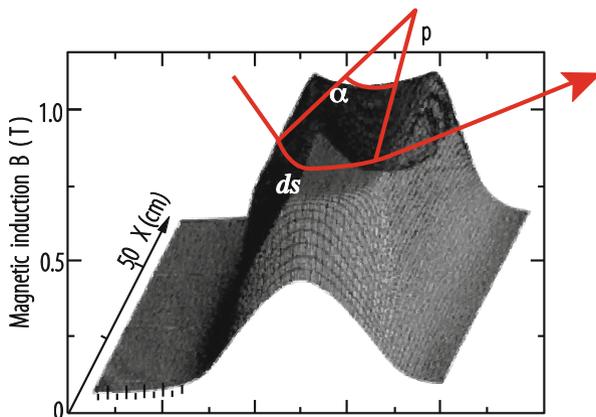


Fig. 6.1 B -field in a storage ring dipole magnet and schematic particle orbit

straight sections, so-called insertions, where the optics are modified to establish conditions needed e.g. for particle injection or extraction and the installation of the radio-frequency resonators for the particle acceleration. In the case of collider rings so-called mini-beta insertions are included, where the beam dimensions are reduced considerably to increase the particle collision rate and where space is needed for the installation of the particle detectors.

The lattice and correspondingly the beam optics therefore are split in different characteristic parts: arc structures that are used to guide the particle beam and define the geometry of the ring; they establish a regular pattern of focusing elements. And the straight sections, that are optimised for the installation of a manifold of technical devices, including the high-energy physics detectors.

6.1.2 Lattice Design

An example of such a high-energy lattice and the corresponding beam optics is shown in Fig. 6.2. In the upper part of the figure the regular pattern of the beta function is plotted in red and green for the two transverse planes. As a consequence of the periodic structure of the lattice, the beta function—and so the beam size—reaches a maximum value in the centre of the focusing, and a minimum in the centre of the defocusing quadrupoles. The lower part of the figure shows the horizontal and vertical dispersion function. The lattice of the complete machine is designed on the basis of small periodic lattice structures—called cells—that repeat many times in the ring. One of the most widespread lattice cells used in high-energy rings is the

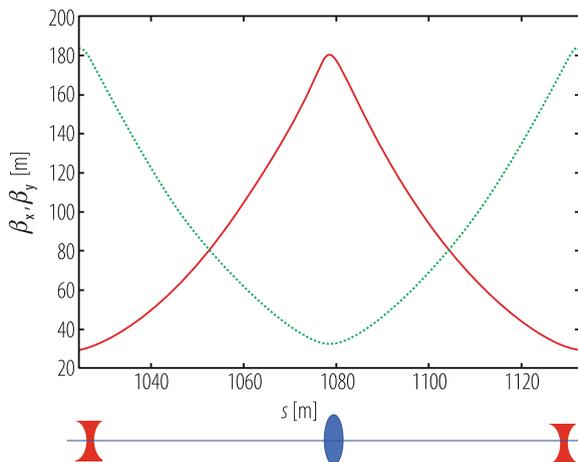


Fig. 6.3 Basic element of a high-energy storage ring: the FODO cell

so-called FODO cell: A magnet structure consisting of focusing and defocusing quadrupole lenses in alternating order. In between the focusing elements the dipole magnets are located and any other machine elements like orbit corrector dipoles, multiple correction coils or diagnostic instruments can be installed.

In Fig. 6.3 the optical solution of such a FODO cell is plotted: The graph shows the β -function in the two transverse planes (red curve for the horizontal, green curve for the vertical plane). In the lower part of the plot the position of the magnet lenses, the lattice, is indicated schematically. In first order the optical properties of such a lattice are determined only by the parameters of the focusing (F) and de-focusing (D) quadrupole lenses. In between these two quadrupole magnets only lattice elements are installed that have zero (“ O ”) or negligible influence on the transverse particle dynamics. Hence the acronym *FODO* for such a structure. Due to the symmetry of the cell the solution for the β function is periodic (in general such a FODO cell is the smallest periodic structure in a storage ring) and it reaches its extreme values at the position of the quadrupole lenses. As a consequence, at these locations in the arc, the beam will reach its maximum dimension $\sigma = \sqrt{\varepsilon\beta}$, and the aperture need will be highest.

Accordingly, the “Twiss” parameter α , which is the derivative of β is generally zero in the middle of the FODO quadrupoles. Based on the thin lens approximation a number of scaling laws and rules can be established to understand the properties of such a FODO structure [2]: How do we arrange the strength and position of the quadrupole lenses in the lattice to obtain a certain beta-function? How does the cell length influence the phase advance of the particle trajectories? How do we guarantee that, turn by turn, a stable particle oscillation is obtained?

In the following we briefly summarise these rules.

- Stability of the motion: the strengths of the focusing (and defocusing) elements in the lattice have to be such that the particle oscillation does not increase. This condition—the stability criterion for a periodic structure in a lattice—is obtained in a FODO if the focal length of the magnets is larger than a quarter of the cell length:

$$f = \frac{1}{kl} = \frac{L_{cell}}{4}. \quad (6.5)$$

- The beta function—and so the beam size—is determined by the phase advance of the cell and its length:

$$\beta_{\max,\min} = \frac{1 \pm \sin(\varphi_{cell}/2)}{\sin \varphi_{cell}} L_{cell}. \quad (6.6)$$

- A similar scaling law is obtained for the dispersion:

$$D_{\max,\min} = \frac{L_{cell}^2}{4\rho} \frac{1 \pm \frac{1}{2} \sin(\varphi_{cell}/2)}{\sin^2(\varphi_{cell}/2)}. \quad (6.7)$$

In general, small values for the β functions as well as for the dispersion are desired. It will be the intention of the lattice designer to minimise the beam size, and so to optimise the aperture need of the beam. In addition the β -function indicates the sensitivity of the beam with respect to external fields and field errors. A change in a quadrupole field e.g. will shift the tune of the beam by

$$\Delta Q = \frac{1}{4\pi} \int \Delta k(s) \beta(s) ds. \quad (6.8)$$

The effect is proportional to the size of the applied change in quadrupole field, Δk but also to the value of the beta function at this position. Therefore, the phase advance of the FODO cell has to be chosen to obtain smallest values for β in both transverse planes, which leads in the case of protons or heavy ions to an optimum phase advance of 90° per cell. It will be no surprise that the focusing structure of typical high energy proton rings like SPS, Tevatron, HERA-p and LHC were optimised for this value.

In addition to the main building blocks, the dipoles and quadrupole magnets, the FODO will be equipped with a number of correction magnets for orbit correction, compensation of higher harmonic field errors of the main magnets, and sextupoles

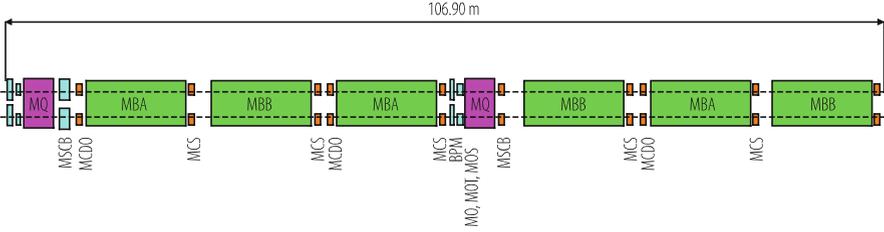


Fig. 6.4 FODO cell of LHC. In addition to the two main quadrupoles and six dipole magnets, diagnostic instruments and multipole compensation coils are included in the arc lattice

for chromaticity compensation of the machine. The FODO cell of LHC, including these corrector magnets is illustrated in Fig. 6.4.

Six dipoles and two main quadrupoles are forming the basic structure of the cell; they are complemented by orbit correction dipoles, trim quadrupoles that are used for fine tuning of the working point and multipole correction coils to compensate higher order field distortions up to 12 pole [3].

Among the higher order correction coils mentioned above the sextupoles play the most critical role in the arc structure, as they are indispensable to compensate the chromatic errors in the lattice. Chromaticity is an optical error that describes the distortion of the focusing properties in a lattice under the presence of momentum spread of the particle beam. In general a sextupole magnet will be installed to support each quadrupole in the arc. At least two sextupole families are required, one for each transverse plane. In some cases several families per plane are installed to improve the region of stability in the transverse plane (the so-called dynamic aperture of the storage ring). They have to be strong enough to correct the chromaticity created in the arc cells as well as in the insertion sections. The mechanism of chromaticity correction is based on the combination of the dispersion function that sorts the particles according to their momentum and the nonlinear field of a sextupole magnet:

$$B_z = \frac{1}{2} \tilde{g} (x^2 - z^2), \tag{6.9}$$

where

$$\tilde{g} = \frac{d^2 B_z}{dx^2} \tag{6.10}$$

describes the sextupole “gradient”.

Normalizing the sextupole field to the beam rigidity we write the contribution of each sextupole to the chromaticity as

$$\Delta Q = \frac{1}{4\pi} \int k_{\text{sext}} D\beta dl \quad (6.11)$$

and it depends indeed on the value of both, beta function and dispersion. Therefore the sextupole magnets that are needed to compensate the natural chromaticity in the ring will be located in the lattice at places where at the same time the dispersion and the beta function are large, i.e. close to the corresponding quadrupole lenses.

6.2 Lattice Insertions

B. J. Holzer

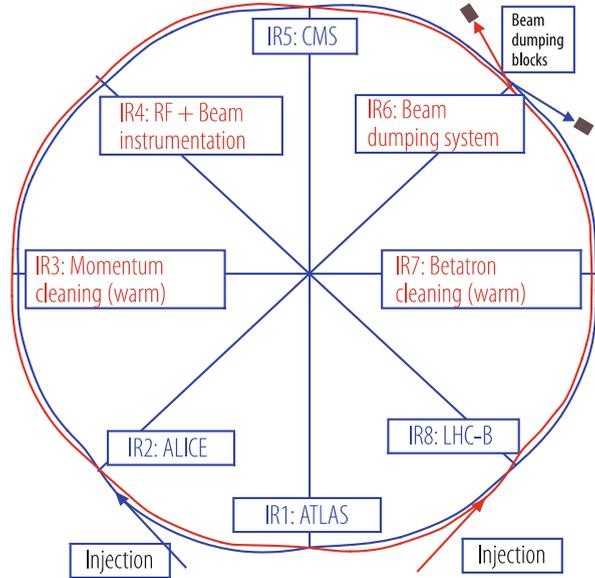
The arc structure of a storage ring is usually built out of regular patterns like FODO cells that are repeated periodically and determine the geometry of the machine. Straight sections are inserted to combine these arcs and provide the space required for beam injection, extraction, or dispersion free lattice parts to install e.g. RF systems. Finally space is needed to establish the conditions that are required for the collisions of the two counter rotating beams. As an example of the general layout of a storage ring we refer again to the LHC lattice. Eight straight sections connect eight arcs: four of them are used for beam injection, extraction and collimation, the remaining four are optimised to house the high-energy detectors (IR1, 5, 2, 8 in Fig. 6.5). Here the storage ring lattice has to provide the free space needed for the installation of a large modern particle detector and the beam optics has to be modified to provide the strong focusing needed at the collision point.

6.2.1 Low Beta Insertions

The most important “insertion” for a particle collider ring is the so-called mini beta structure: The key issue of a collider is its luminosity [4] that defines the rate of produced collision events (particles or particle reactions of interest) in the machine. Its value is defined by the machine lattice and under the assumption of equal beam properties in the two colliding beams it is given by the stored currents in the two beams, I_{p1} , I_{p2} , the revolution frequency f_0 , the number of stored bunches, n_b , and most of all by the transverse size of the two beams, σ_x^* and σ_y^* . In the simplest case we get:

$$L = \frac{1}{4\pi e^2 f_0 n_b} \frac{I_{p1} I_{p2}}{\sigma_x^* \sigma_y^*}. \quad (6.12)$$

Fig. 6.5 Lattice geometry of the LHC



A more general formula that includes geometric and optical reduction factors is presented in Sect. 6.4 [4]. At the interaction point “IP”, the intention of the lattice designer will be to reduce the beta function as much as possible in order to obtain the smallest possible beam. The main limiting factor comes from a basic principle which is valid for any system of particles under the influence of conservative forces (“Liouville’s Theorem”): Under conservative forces, the density of the particle’s phase space volume is constant. Applying this law to a particle beam in an accelerator it means that the beam dimension and divergence are not independent of each other. Namely for the design of symmetric drift space in a storage ring we can deduce a rule for the beta function: Starting from a waist ($\alpha^* = 0$ at the collision point) the beta function develops as

$$\beta(s) = \beta^* + \frac{s^2}{\beta^*}. \quad (6.13)$$

The star refers to the value at the waist (e.g. the interaction point “IP”). This relation is a direct consequence of Liouville’s theorem and therefore of fundamental nature. As a consequence the behaviour of β in a symmetric drift cannot be changed and has a strong impact on the design of a storage ring: Small beta functions at the collision point and a large distance to the first focusing element lead to high values of the beta function and correspondingly to large beam dimensions at the first focusing element in front and after the IP.

The preparation of the beam optics for the installation of modern high-energy detectors therefore needs special treatment in the lattice design to provide the large space needed for the detector hardware. An illustrative example is shown in

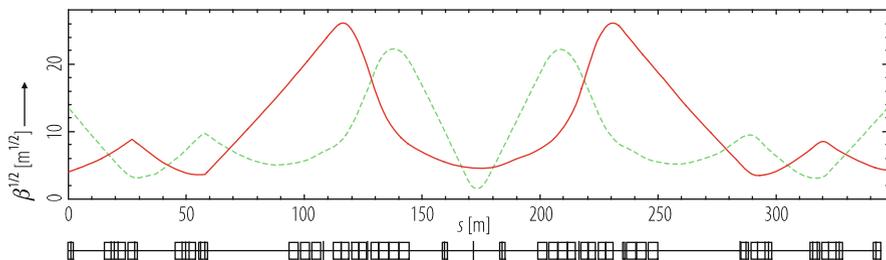


Fig. 6.6 Layout of a mini beta insertion scheme. The example shows a low beta insertion based on a quadrupole doublet. The vertical beta function (green line) starting with smaller values at the IP shows a stronger increase than the beta in the horizontal plane. Accordingly the doublet quadrupoles are powered in QD-QF polarity

Fig. 6.6: a long symmetric drift space that holds the experiment is centred around the interaction point of the colliding beams. Depending on the respective value of beta at the IP the beta functions increase in the horizontal (red) and vertical (green) plane and are focused back using a couple of strong, large aperture and high quality quadrupole lenses. Depending on the particular situation (namely the ratio of the two β^* values in the two planes a quadrupole doublet or triplet arrangement will be the adequate choice for these mini beta quadrupoles. Additional independent quadrupole magnets (i.e. individually powered magnets) will be needed to create a smooth transition of the optics from the IP to the periodic solution of the FODO cells in the arc. In general eight parameters have to be optimised: the β and α values in the two planes, the dispersion and its derivative and the phase advance of the complete mini beta system. As a consequence such a mini beta insertion will have to be equipped with at least eight individually powered quadrupole magnets to fulfil this requirement.

It has been pointed out in the previous chapter that the emittance of a particle beam is not constant during acceleration but depends on the energy of the particle beam. In the case of a proton or ion beam the adiabatic shrinking is the dominant effect and the emittance follows the rule $\varepsilon \propto 1/\beta\gamma$ where β and γ are the relativistic parameters. As a consequence the emittance in a proton storage ring is highest at injection energy and the beam optics has to be optimised to limit the beta function at any place in the machine to values that guarantee sufficient aperture. At high energy (the so-called flat-top) the emittance is small enough that the mini beta concept can be used to full extend and only here the β^* can be reduced to the small values that are required to deliver the design luminosity values. The lattice of the mini beta insertion therefore has to be optimised in a way, that two quite different beam optics can be established by corresponding adjustment of the quadrupole gradients: A low energy optics for injection and the early steps of the acceleration and a true mini beta optics that will be used for the collider run at high energy.

The procedure to pass from the injection optics to the luminosity case is often called “beta squeeze” and is a critical situation as optics, orbits and global beam parameters like tune and chromaticity have to be maintained constant and well

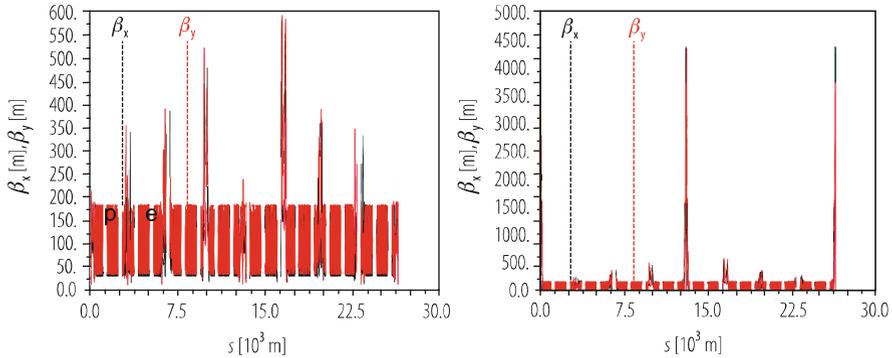


Fig. 6.7 (Left) Beam optics for the LHC: 450 GeV injection optics optimised for small values of beta to gain highest aperture in the machine. (Right) Low beta optics for the LHC luminosity operation: due to the small values at the IP the beta function reaches large values in the low beta quadrupole lenses. (Note the different scale of the vertical axis)

controlled during the changing quadrupole settings. Several intermediate steps might be needed to guarantee a smooth transition between the two operation modes. In the case of the LHC the 450 GeV injection case and the 7 TeV luminosity optics are compared in Fig. 6.7.

6.2.2 Injection and Extraction Insertions

In addition to the mini beta insertions where the beams are optimised for highest collision rates, additional insertions are needed in the storage ring for beam injection and extraction. In these cases the same rules are valid as for the mini beta insertions but in general the consequences are more relaxed. Additional hardware that has to be installed for the injection process (fast kicker magnets and septum dipoles to inject the new beam) is much smaller than the detectors at the collision points. Still, however, some modifications of the lattice will be needed and the optics will have to be re-matched to establish the required space. A special additional feature should be mentioned here: the new beam that is being injected has to match perfectly in energy and in phase space to the optical parameters of the storage ring or synchrotron. At the end of the beam transfer line as well as in the storage ring the focusing fields have to be optimised to obtain the same values of the Twiss functions α and β in both transverse planes. As in the case of the mini beta insertions additional individually powered quadrupole magnets are needed. As an example the beam optics of the SPS-LHC transfer-line is plotted in Fig. 6.8. At the beginning and the end of the lattice structure—indicated by red markers in the figure—the beta function is modified to match the optics from the SPS to the FODO channel of the transfer line and from

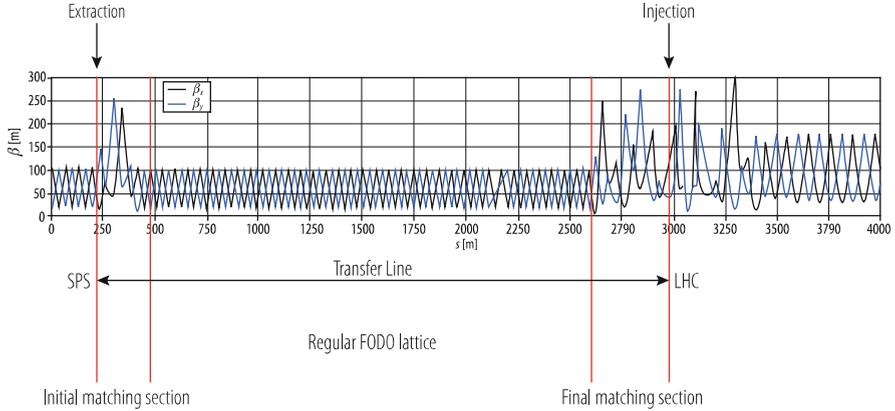


Fig. 6.8 Transfer line between the SPS and the LHC. Two matching sections have to be introduced to adopt the beam optics from the SPS to the transfer line and to the LHC

the FODO to the LHC insertion at IR2 and IR8 where the injection elements are located.

6.2.3 Dispersion Suppressors

The dispersion function $D(s)$ has already been introduced in Sects. 2.4 and 6.1. It describes the trajectory in the case of a momentum deviation of the particle and is the consequence of the corresponding error in the bending strength of the dipole magnets. In the arc structure with its regular pattern of dipole magnets, dispersive effects cannot be avoided (but they should be minimised) and the additional amplitude due to the dispersion has to be considered if we are talking about particle trajectories or beam sizes. In linear approximation and for a small momentum spread $\Delta p/p$ in the beam, the amplitude of a particle oscillation is obtained by

$$x(s) = x_\beta(s) + D(s) \frac{\Delta p}{p_0}, \quad (6.14)$$

where x_β describes the solution of the homogeneous differential equation (the usual betatron oscillations of the particle) and the second term—the dispersion term—corresponds to the additional oscillation amplitude for particles with a relative momentum error $\Delta p/p_0$. At the interaction point where the smallest beam sizes are required to obtain the highest luminosity, we intend to suppress the dispersion and as the collision point is generally located in a straight section of the accelerator, techniques have been developed to obtain dispersion free sections inside the lattice. The insertions that are used to reduce the dispersion function from its periodic value in the arc to zero are called dispersion suppressors [2, 5, 6].

It has to be mentioned in this context that especially in the case of synchrotron light sources a variety of lattice types has been developed with the goal to achieve small or even zero dispersion in the ring or in parts of it. However, these lattices are optimised for the purpose of high brilliant synchrotron radiation and are not ideal for high-energy particle accelerators, where FODO cells are usually the most appropriate choice.

Referring to high energy colliders we will concentrate therefore on the interaction region, i.e. a straight section of a ring where two counter rotating beams collide in a dispersion free part of the storage ring. A non-vanishing dispersion dilutes the luminosity of the machine and leads to additional stop bands in the working diagram of the accelerator (“synchro-betatron resonances”), that are driven by the beam-beam interaction. Therefore sections are inserted in our magnet lattice that are designed to reduce the function $D(s)$ to zero. Three main techniques are widely used: the quadrupole based dispersion suppressor, the missing bend scheme and the half bend scheme. We will not present all of them in detail but instead restrict ourselves to the basic idea behind it.

6.2.3.1 The “Straightforward” Way: Dispersion Suppression Using Quadrupole Magnets

Let us assume here that a periodic lattice is given in the arc (see Fig. 6.2) and that this FODO structure simply is continued through the straight section—but with vanishing dispersion. Given an optical solution in the arc cells, as for example shown in Fig. 6.9, we have to guarantee that starting from the periodic solution

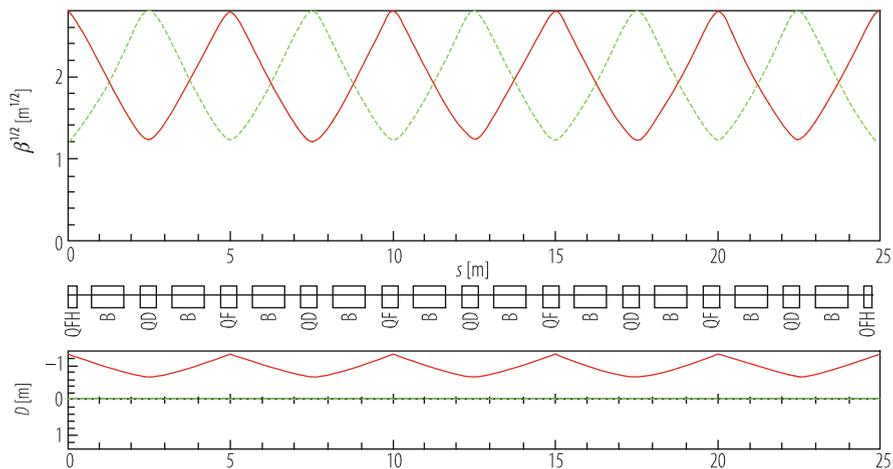


Fig. 6.9 Periodic FODO and horizontal dispersion function in a regular FODO structure

of the optical parameters $\alpha(s)$, $\beta(s)$ and $D(s)$ we obtain a situation at the end of the suppressor where we get $D(s) = D'(s) = 0$ and the values for α and β unchanged.

The boundary conditions after the suppressor section

$$\begin{aligned}
 D(s) &= D'(s) = 0, \\
 \beta_x(s) &= \beta_{x \text{ arc}}, \alpha_x(s) = \alpha_{x \text{ arc}}, \\
 \beta_y(s) &= \beta_{y \text{ arc}}, \alpha_y(s) = \alpha_{y \text{ arc}},
 \end{aligned}
 \tag{6.15}$$

can be fulfilled by introducing six additional quadrupole lenses whose strengths have to be matched individually in an adequate way. This can be done by using one of the beam optics codes that are available today in every accelerator laboratory. An example is shown in Fig. 6.10, starting from a FODO structure with a phase advance of $\varphi \approx 70^\circ$ per cell.

The advantages of this scheme are:

- it works for any phase advance of the arc structure;
- matching works also for different optical parameters α and β before and after the dispersion suppressor as—within a certain range—the quadrupoles can be used to match the Twiss functions to different values;
- the ring geometry is unchanged as the number and location of dipole magnets in the ring is unchanged.

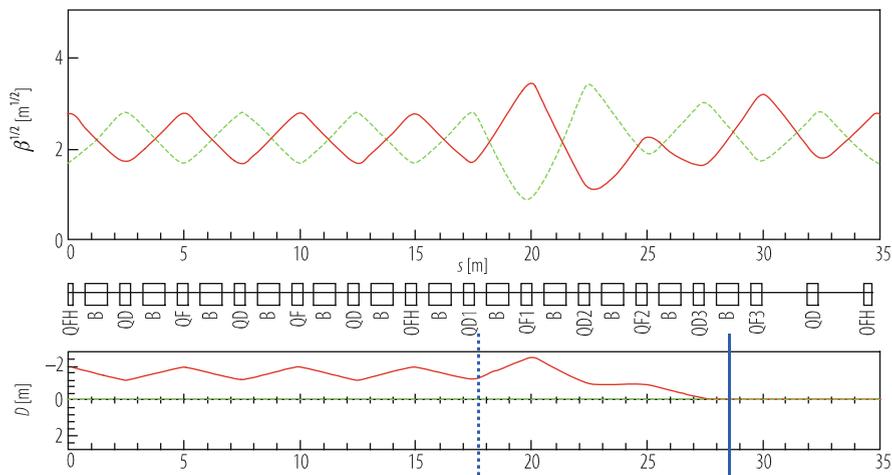


Fig. 6.10 Periodic FODO and horizontal dispersion function in a regular FODO structure dispersion suppressor scheme based on individually powered quadrupole lenses

On the other hand there are a number of disadvantages that have to be mentioned:

- as the strength of the additional quadrupole magnets have to be matched individually the scheme needs additional power supplies and quadrupole magnet types which can be an expensive requirement;
- the required quadrupole fields are in general stronger than in the arc;
- the β function reaches higher values (sometimes *really* high values) which leads to higher beam sensitivity and larger aperture needs.

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

6.2.3.2 The “Clever” Way: Half Bend Schemes

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

Starting from the dispersion free straight section the basic idea of this scheme is to create with a special arrangement of dipole magnets inside the dispersion suppressor—exactly the dispersion that corresponds to the periodic solution of the arc FODO cells. The solution will depend on the phase advance of the cells as well as on the strength of the bending magnets inside the suppressor magnets.

As explained before in the beam optics chapter, the matrix for a periodic part of the lattice (namely one single cell in our case) can be expressed as

$$M_{cell} = \begin{pmatrix} C & S & D \\ C' & S' & D' \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \cos \phi_c & \beta_C \sin \phi_c & D \\ -\frac{1}{\beta_c} \sin \phi_c & \cos \phi_c & D' \\ 0 & 0 & 1 \end{pmatrix}, \quad (6.16)$$

where the index “ c ” reflects the solution of a cell, ϕ_c denotes the phase advance for a single cell and the elements D and D' correspond to its periodic dispersion.

As usual the dispersion elements are obtained by

$$D(l) = S(l) \int_0^l \frac{C(\tilde{s})}{\varrho(\tilde{s})} d\tilde{s} - C(l) \int_0^l \frac{S(\tilde{s})}{\varrho(\tilde{s})} d\tilde{s}. \quad (6.17)$$

The functions $C(s)$ and $S(s)$ are the cosine and sine like matrix elements of the lattice element in the sense that e.g. $C(s) = M[1,1]$, and the integral is executed over one complete cell.

In the dispersion suppressor section, the dispersion $D(s)$ starts with the value D_0 the end of the arc cell and is reduced to zero. Or turning it around and thinking from right to left: the dispersion has to be created inside the suppressor part by proper arrangement of the dipole magnets, starting from $D = D' = 0$ in the straight section to reach the values that correspond to the periodic dispersion of the arc cells. Solving the equation above by integrating over a certain number of cells will determine the bending strength $1/\rho$ and the number n of cells in the suppressor part that is needed to fulfill the boundary condition and get the values of the dispersion in the following periodic arc cell.

For a given phase advance φ_c per cell two conditions for the dispersion matching are obtained that combine the number of suppressor cells, n , and the strength of the suppressor dipoles, δ_{supr} :

$$\left. \begin{aligned} 2\delta_{\text{supr}}\sin^2\left(\frac{n\phi_c}{2}\right) &= \delta_{\text{arc}} \\ \sin(n\phi_c) &= 0 \end{aligned} \right\} \delta_{\text{supr}} = \frac{1}{2}\delta_{\text{arc}}. \quad (6.18)$$

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

$$n\phi_c = k\pi, \quad k = 1, 3, \dots \quad (6.19)$$

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

Figure 6.11 shows such a half bend dispersion suppressor, starting from a FODO structure with 60° phase advance per cell. The focusing strength of the FODO cells before and after the suppressor are identical, with the exception that—clearly—the FODO cells on the right are “empty”, i.e. they have no bending magnets.

It is evident that unlike to the suppressor scheme with quadrupole lenses now the beta function is unchanged in the suppressor region.

Again this scheme has advantages:

- no additional quadrupole lenses are needed and no individual power supplies;
- in first order the β functions are unchanged; aperture needs and beam sensitivity are not increased;

and disadvantages:

- it works only for certain values of the phase advance in the structure and therefore restricts the free choice of the optics in the arc;
- special dipole magnets are needed (having half the strength of the arc types);
- the geometry of the ring is changed.

energy beams) to deflect the beam into or out of the accelerator aperture, and frequently also a closed orbit bump to approach the septum and reduce the required kicker strength. For these single-turn methods, the beam losses can be very low, and the emittance dilution associated with the injection or extraction can be very small, defined by the delivery precision, the optics mismatch, the kicker flat top ripple and septum stability. For both injection and extraction, the circulating beam can be adversely affected by septum stray fields penetrating into the circulating beam region and by the kicker field rise time which can overlap temporally with circulating bunches. Injecting a bunched beam into another accelerator also requires that the momentum spread and phase be matched to the RF bucket, and that the RF system can accept the transient beam loading which arises from the sudden change in beam intensity.

Multiple-turn injection is used to fill the circumference of a receiving accelerator and to accumulate bunch intensity. A wide variety of multiple-turn injection and extraction schemes exist, and these can be very different for lepton and hadron machines. Lepton injection schemes 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” to precisely tailor the transverse and longitudinal distributions, particularly with H^- charge exchange injection or slip stacking; while resonant multiple-turn extraction schemes have been developed to provide quasi-continuous particle fluxes for periods which range from milliseconds to hours. The additional hardware 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.

Overall, injection and extraction techniques share many similarities and hardware requirements [8]: one important difference between them is that extraction is usually at higher beam rigidity, which implies less effect from space charge and also stronger and hence longer deflecting systems, which can have a significant effect on lattice and insertion design [9–11].

6.3.1 *Fast Injection*

Fast injection [12–14] is typically used to fill another machine with bunch-to-bucket transfer, or to fill a collider over several injections with ‘boxcar’ stacking, where bunches or trains of bunches are added sequentially like boxcars (wagons) to a train. The system design depends critically on the aperture needed for the beam, and the kicker rise time, fall time and flat top duration. Very fast kicker rise times are often 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.

6.3.2 Slip-Stacking Injection

In slip-stacking [15], two trains of bunches are merged to increase the bunch intensity, using separate RF systems. 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 as a result 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 can be made to move relative to each other in phase. When the phase difference reaches zero, both sets of bunches are captured together and merged, by a rapid change of the RF frequency. The accelerator needs enough momentum aperture to accept both beams, and sophisticated RF control to make the manipulations. The final longitudinal emittance is the sum of the two individual emittances multiplied by an unavoidable blowup factor, typically around 1.5.

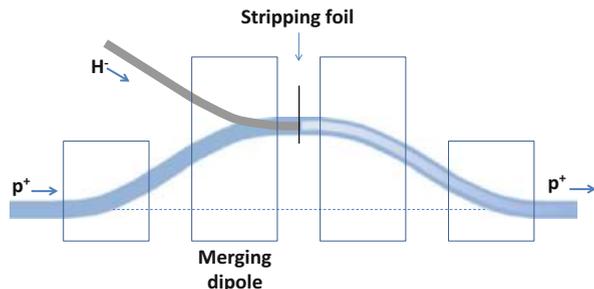
6.3.3 H^- Charge-Exchange Injection

High brightness, low energy proton machines frequently make use of H^- charge exchange injection [16]. In this technique, a linac accelerates H^- ions which are then merged with the circulating proton beam in a dipole magnet, Fig. 6.13, before the two loosely-attached electrons are stripped away in a foil which is almost transparent to the circulating beam.

This technique allows the accumulation of high brightness beams, since unlike other methods it 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 diamond-like carbon, with a thickness typically in the micron range, which is a compromise between obtaining high stripping efficiency and minimizing the beam losses and emittance growth from scattering.

Fast painting bumpers or kickers in both planes are used to displace the circulating beam access with respect to the foil, and the waveform of the bumper

Fig. 6.13 Merging H^- and p^+ beams in H^- charge exchange injection



field can be varied to achieve the desired phase space density distribution. This is the process of phase space painting, where the small emittance LINAC beam is the brush and the large acceptance of the receiving machine is the canvas.

In addition to beam loss from scattering at the foil, another significant source of beam loss can be the field-stripping in the third chicane magnet of excited H^0 . In ISIS [17] the injection is made on the ramp, and the dispersion at the foil provides some of the transverse phase space painting. For SNS [18], where the average beam power is over 1 MW, the uncontrolled beam losses must be kept extremely low and the accumulation is made over 1160 turns.

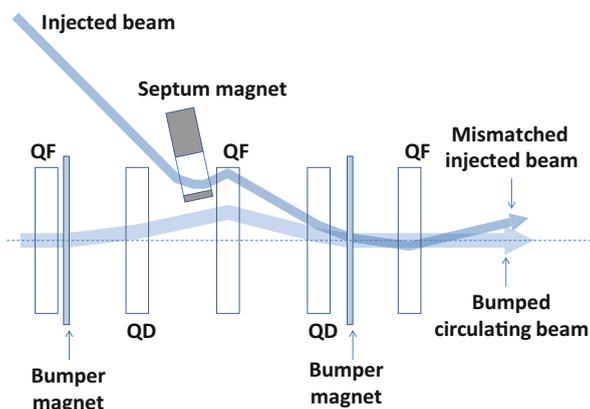
The use of stripping foils is disadvantageous for several reasons, in particular the associated uncontrolled beam losses, but also due to the simple mechanical and radiological difficulties in handling such fragile objects. A foil-free method of H^- stripping using a high-powered laser to resonantly excite neutral H^0 before field stripping in a dipole has been proposed and demonstrated in principle, and is promising for very high energy H^- injection systems [19].

6.3.4 Lepton Accumulation Injection

Injection of leptons can take advantage of the strong damping which is present from synchrotron radiation to accumulate intensity. This is very commonly used at Synchrotron Radiation rings, where top-up operation [20] consists of frequently injecting small amounts of beam to replace beam losses and keep the beam and synchrotron radiation intensities stable in a very small range.

In betatron injection, Fig. 6.14, the injected bunch or train is injected with an orbit offset with respect to the circulating beam, which is moved towards the injection septum with a fast closed-orbit bump. The offset between the injected beam and the circulating beam must be large enough to accommodate the injection septum. The particles of the newly-injected bunches then perform damped betatron oscillations

Fig. 6.14 Betatron injection. The injected beam is mismatched and performs betatron oscillations until damped by emission of synchrotron radiation



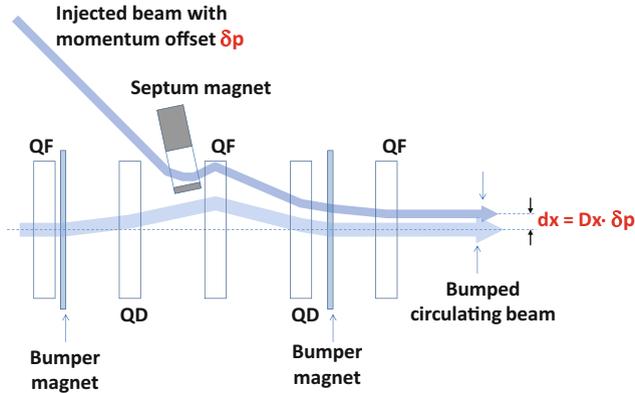


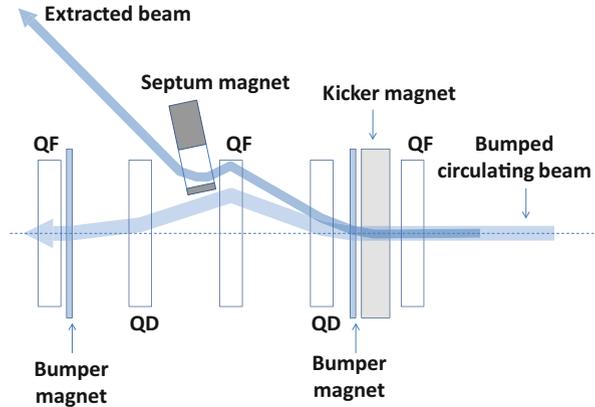
Fig. 6.15 Synchrotron injection. The injected beam has a momentum offset, and the injection trajectory is matched to the local dispersion orbit. The beam then performs oscillations about the closed orbit determined by the dispersion function, as the momentum changes with the synchrotron oscillations

around the closed orbit, until they merge with the already circulating beam. This technique has the disadvantage that the betatron amplitude may be large in regions of the accelerator where the β -function is large. In the alternative synchrotron injection [21], Fig. 6.15, the new particles are injected with a momentum offset δp and a position offset X into a region with dispersion D , such that $X = \delta p \times D$. The particles are injected onto the matched betatron orbit for their momentum, and thus only perform synchrotron oscillations around the stored particles, with the transverse offsets following the dispersion function. For LEP a combination of betatron and synchrotron injection was preferred [22], since the dispersion in the long straight sections was very small and the background to the experiments could be significantly improved.

6.3.5 Fast Extraction

Fast extraction is typically used to provide beam to a higher energy machine with bunch-to-bucket transfer. As for fast injection, the system design depends critically on the aperture needed for the beam, and the kicker rise time, fall time and flat top duration. Achieving fast kicker rise time with sufficient deflection angle at high beam rigidity is a common challenge, as is the design of the extraction insertion where the septum strength must be sufficient to provide enough clearance at the next downstream accelerator element. As beam energies increase, protection from missteered beam of the extraction septum and of other accelerator components becomes important; for the LHC beam extraction system at 7 TeV [23], the synchronization of the kicker system and protection from asynchronous kicker firing is a critical

Fig. 6.16 Schematic of fast extraction system with kicker, septum and orbit bumpers. For higher energy machines, protection devices to intercept and dilute any mis-kicked extracted beam are placed in front of the septum and downstream QF quadrupole



system design feature. Closed orbit bumps can be used to move the beam closer to the septum, to reduce the required kick strength, Fig. 6.16.

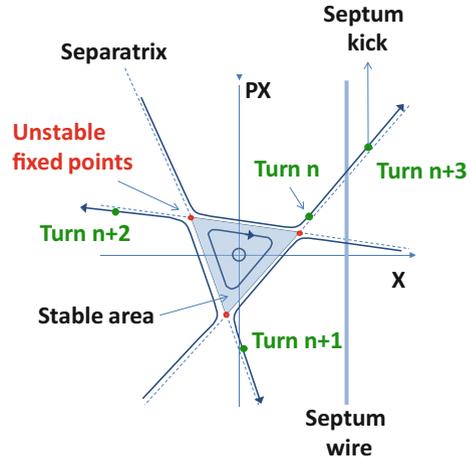
6.3.6 Resonant Extraction

Many rate-limited applications such as physics experiments, test beams or medical treatment beams require a slow flux of particles with as uniform a time structure as possible. Resonant extraction using the third integer is the most common method of providing such uniform spills. In this ‘slow’ extraction [24], 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. Before the start of the extraction process, particles remain stable if their single-particle emittances are smaller than the area of the stable triangle.

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. 6.17. The rate of extraction is controlled either by modulating the excitation process or by controlling the stable area. Several techniques for driving the particles unstable are possible:

- i) the stable area can be reduced by increasing the resonance (sextupole) strength or by moving the tune closer to the third integer. Increasing the resonance strength reduces the stable area, but the smallest amplitude particles cannot be extracted, and changing the resonance affects the machine optics. Crossing the resonant tune offers the advantage that all of the beam can be extracted; however, the optics is still perturbed and in addition the position of the extracted beam in phase space changes as particles are extracted;

Fig. 6.17 Resonant extraction in normalised 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



- ii) the particle amplitudes can be increased by use of a transverse excitation. The stable area is kept fixed and the particle amplitudes increased, as in RF-knockout [25] where a high-frequency damper is used near the betatron resonance frequency to excite the beam. The machine optics is not changed and this method allows very fine control of the spill flux, suitable for medical machines;
- iii) the particles can be accelerated into the resonance where the chromaticity couples the momentum and the tune. A betatron core can be used [26] to accelerate the beam smoothly through the resonance. As the momentum of the beam changes, this is coupled via the chromaticity into a tune change. This method provides stability and insensitivity to power supply ripple. An alternative method (Constant Optics Slow Extraction) is to change the strengths of all machine elements to achieve the same effect, where the beam momentum remains fixed but the accelerator momentum changes [27];
- iv) RF noise can be applied to gradually diffuse particles longitudinally, which through the chromaticity are brought into resonance. This stochastic extraction [28] allows extremely long and uniform spills, and again has the advantage of leaving the machine lattice functions unchanged.

It should be noted that extraction can also be made using the second order resonance, where octupole fields are used to define a stable area in phase space. The amplitude growth with time is much faster, and the beam can be extracted in several hundred turns.

The use of a physical septum means that losses and activation are key performance aspects for slow extraction. Several interesting techniques exist to reduce beam losses at extraction [29], including the use of scatterers to reduce the particle density at the septum, multipoles to manipulate the separatrix density and techniques to reduce the angular spread of the beam and reduce the effective septum width.

6.3.7 *Continuous Transfer Extraction*

A frequent requirement in an accelerator complex is to fill a large circumference machine with the contents of a smaller machine. One way of doing this is boxcar stacking; another technique is continuous transfer [30], where the beam in the first machine is extracted over a number of turns, like peeling the skin from an orange in a continuous strip. The machine tune is brought near to the appropriate integer n , where the beam will be extracted in $n+1$ turns. A fast closed bump is then applied to the circulating beam with kickers to move the beam partly across a septum, such that a fraction of the beam is cut and extracted. The machine tune rotates the beam in phase space such that subsequent slices are extracted—when the n^{th} turn is extracted, the bump amplitude is increased to extract the remaining central part. This process is of use where the injector can service other machines or experiments while the receiving machine is accelerating the beam, since it minimises the time spent filling. The disadvantage of the technique is that large beam losses occur at the septum, with the transfer efficiency typically 85%. The transfer can be made with a bunched beam, leaving space for the kicker rise time, but this means that the receiving machine will need to capture a beam with strong intensity modulation. Another feature of this extraction is that the extracted slices all have different emittances, as the slices in phase space are all different.

6.3.8 *Resonant Continuous Transfer Extraction*

To reduce the beam losses from continuous transfer, a hybrid technique has been developed and deployed called Multi-Turn Extraction [31] where non-linear resonances are excited which define stable areas in phase space. These are populated by the controlled crossing of a resonance, and the islands are then separated by varying the multipole strength to provide a physical separation at the septum, to reduce or avoid transverse losses. The beam needs to be bunched with a gap to avoid losses during the kicker rise time. In addition to the lower losses, another advantage of this technique is that the extracted islands all have the same emittance.

6.3.9 *Other Injection and Extraction Techniques*

More exotic injection and extraction techniques also exist as working systems or concepts. These include radio-frequency stacking [32], pion-decay injection into muon storage rings [33] and combined cooling and stacking [34]. Charge exchange extraction [35] is used in cyclotrons, with a stripping foil, to convert for example H^- to p^+ , or H_2^+ to H_2^{2+} so that the beam is then deflected out of the accelerator. Finally, very high energy particle extraction can be envisaged with a bent crystal replacing the septum [36].

6.4 Concept of Luminosity

Werner Herr · Bernhard Holzer · Bruno Muratori

6.4.1 Introduction

In particle physics experiments the energy available for the production of new effects is the most important parameter. Besides the energy the number of useful interactions (events), is important. The quantity that measures the ability of a particle accelerator to produce the required number of interactions is called the luminosity (see Chap. 2) and is the proportionality factor between the number of events per second dR/dt and the cross section σ_p :

$$\frac{dR}{dt} = \mathcal{L} \cdot \sigma_p \quad (6.22)$$

The unit of the luminosity is therefore $\text{cm}^{-2} \text{s}^{-1}$.

Here we will derive a general expression for the luminosity and give formulae for basic cases. Additional complications such as crossing angle and offset collisions are added to the calculation. Other effects such as the hourglass effect are estimated from the generalized expression.

In the final section we will discuss the measurement and calibration of the luminosity for both e^+e^- as well as hadron colliders.

6.4.2 Computation of Luminosity

In the case of two colliding bunches, both serve as “target” as well as “incoming” beam at the same time. A schematic picture is shown in Fig. 6.18. The overlap integral which is proportional to the luminosity \mathcal{L} can be written as [37]:

$$\mathcal{L} \propto K N_1 N_2 \cdot \int \int \int \int_{-\infty}^{+\infty} \rho_1(x, y, s, -s_0) \rho_2(x, y, s, s_0) dx dy ds ds_0 \quad (6.23)$$

Here $\rho_1(x, y, s, s_0)$ and $\rho_2(x, y, s, s_0)$ are the time dependent beam density distribution functions and N_1 and N_2 the number of particles per bunch. We assume, that the two bunches meet at $s_0 = 0$ and $s_0 = c \cdot t$ is used as the “time” variable. Because the beams are moving against each other, we have to introduce the kinematic factor [38]:

$$K = \sqrt{(\vec{v}_1 - \vec{v}_2)^2 - (\vec{v}_1 \times \vec{v}_2)^2/c^2} \quad (6.24)$$

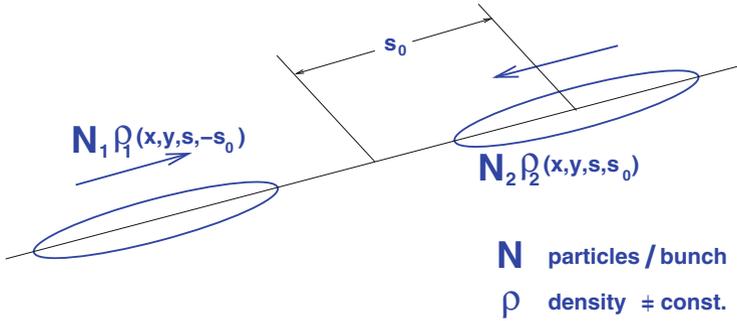


Fig. 6.18 Schematic view of a colliding beam interaction

This factor is needed to make the luminosity and therefore the cross section relativistically invariant.

For the calculation we assume Gaussian profiles in all dimensions of the form:

$$\rho_{iz}(z) = \frac{1}{\sigma_z \sqrt{2\pi}} \exp\left(-\frac{z^2}{2\sigma_z^2}\right) \quad \text{where } i = 1, 2, \quad z = x, y \quad (6.25)$$

in the transverse planes and

$$\rho_s(s \pm s_0) = \frac{1}{\sigma_s \sqrt{2\pi}} \exp\left(-\frac{(s \pm s_0)^2}{2\sigma_s^2}\right) \quad (6.26)$$

in the longitudinal plane.

We further assume that the distributions are independent in the three coordinates and can be factorized. The integral (6.23) can then be evaluated. For the general case of: $\sigma_{1x} \neq \sigma_{2x}$, $\sigma_{1y} \neq \sigma_{2y}$, but assuming approximately equal bunch lengths $\sigma_{1s} \approx \sigma_{2s}$ we get the formula:

$$\mathcal{L} = \frac{N_1 N_2 f_c}{2\pi \sqrt{\sigma_{1x}^2 + \sigma_{2x}^2} \sqrt{\sigma_{2y}^2 + \sigma_{2y}^2}} \quad (6.27)$$

Where N_1 and N_2 are the bunch intensities and f_c the repetition rate. In the case of a circular collider with N_b bunches and a revolution frequency of f_{rev} , we have $f_c = f_{rev} \cdot N_b$.

6.4.3 Luminosity with Correction Factors

The Eq. (6.26) requires correction factors when the beam do not fully overlap (crossing angle and offset), the beam size varies in the longitudinal plane (hour glass effect) or in the case of non-Gaussian beams.

6.4.3.1 Effect of Crossing Angle and Transverse Offset

Here we give the correction to the luminosity calculation in the case where two bunches do not collide exactly head-on, but with a crossing angle and/or transverse offset. In that case the luminosity is reduced and we must apply a correction factor to compute the correct value. For simplicity we assume crossing angle and offset in the horizontal (x) plane, but this is not a restriction. The integration (6.23) can be carried out by rotating the coordinate systems of the two beams each by half the crossing angle [37] and can be simplified introducing the factors:

$$A = \frac{\sin^2 \frac{\phi}{2}}{\sigma_x^2} + \frac{\cos^2 \frac{\phi}{2}}{\sigma_s^2}, \quad B = \frac{(d_2 - d_1) \sin(\phi/2)}{2\sigma_x^2}, \quad W = e^{-\frac{1}{4\sigma_x^2}(d_2 - d_1)^2} \tag{6.28}$$

$$S = \frac{1}{\sqrt{1 + \left(\frac{\sigma_s}{\sigma_x} \tan \frac{\phi}{2}\right)^2}} \approx \frac{1}{\sqrt{1 + \left(\frac{\sigma_s}{\sigma_x} \frac{\phi}{2}\right)^2}} \tag{6.29}$$

where $\Phi/2$ is half the crossing angle and d_1 and d_2 are the transverse offsets of the two beams (Fig. 6.19).

We can re-write the luminosity with three correction factors:

$$\mathcal{L} = \frac{N_1 N_2 f N_b}{4\pi \sigma_x \sigma_y} \frac{N_1 N_2 f N_b}{4\pi \sigma_x \sigma_y} \cdot W \cdot e^{\frac{B^2}{A}} \cdot S \tag{6.30}$$

This factorization enlightens the different contributions and allows straightforward calculations. The last factor S is the luminosity reduction factor for a crossing angle. One factor W reduces the luminosity in the presence of beam offsets and the factor $e^{\frac{B^2}{A}}$ is only present when we have a crossing angle and offsets simultaneously in the same plane. The formulae for the luminosity under very general conditions can be found in [39]. A popular interpretation of this result is to consider it a correction to the beam size and to introduce an “effective beam size” like:

$$\sigma_{eff} = \sigma / \sqrt{1 + \left(\frac{\sigma_s}{\sigma} \frac{\phi}{2}\right)^2} \tag{6.31}$$

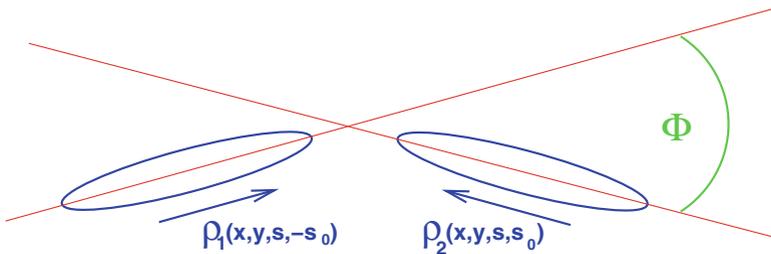


Fig. 6.19 Schematic view of a colliding beam interaction at a crossing angle

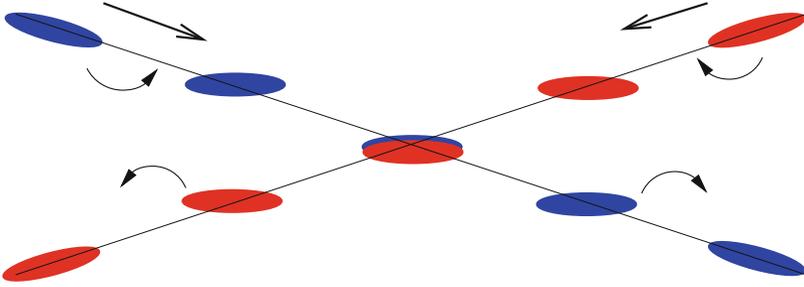


Fig. 6.20 Scheme of crab crossing with transversely deflecting cavities

This equation is valid when $\sigma_z \gg \sigma$. The effective beam size can then be used in the standard formula for the beam size in the crossing plane. This concept of an effective beam size is interesting because it also applies to the calculation of beam-beam effects of bunched beams with a crossing angle [40, 41].

In the case of flat beams, (i.e. $\sigma_z \ll \sigma$) a more general expression has to be used, (see e.g. [39]).

To avoid the loss of luminosity, the use of crab cavities is an option, where the bunches are deflected transversely before and after the collision Fig. 6.20.

6.4.3.2 Hour Glass Effect

In a low- β region the β -function varies with the distance s to the minimum like:

$$\beta(s) = \beta^* \left(1 + \left(\frac{s}{\beta^*} \right)^2 \right) \quad (6.32)$$

For very small β^* comparable to the bunch length, the β -function is not a constant along the longitudinal dimension of the bunch. It cannot be considered a constant in Eq. (6.23). It follows a parabola and rises very fast and can become very large for small β^* .

In our formulae we have to replace σ by $\sigma(s)$ and get a more general expression for the luminosity (assuming equal parameters in both beams, the most general expression can be found in [39]):

$$\frac{\mathcal{L}(\sigma_s)}{\mathcal{L}(0)} = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi}} \frac{e^{-u^2}}{\sqrt{\left[1 + \left(\frac{u}{u_x} \right)^2 \right] \cdot \left[1 + \left(\frac{u}{u_y} \right)^2 \right]}} du \quad (6.33)$$

Using the expressions: $u_x = \beta_x^*/\sigma_s$ and $u_y = \beta_y^*/\sigma_s$
 For the case of round beams it can be simplified and the integral becomes:

$$\frac{\mathcal{L}(\sigma_s)}{\mathcal{L}(0)} = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi}} \frac{e^{-u^2}}{\left[1 + \left(\frac{u}{u_x}\right)^2\right]} du = \sqrt{\pi} \cdot u_x \cdot e^{u_x^2} \cdot \operatorname{erfc}(u_x) \quad (6.34)$$

Here $\operatorname{erfc}(u)$ is the complex error function. The hourglass effect depends strongly on the relative value of β^* and the bunch length σ_s . For small β^* the effect becomes relevant since the beam size varies rapidly along the longitudinal bunch direction, i.e. when s becomes comparable to the bunch length in Eq. (6.32). A loss of luminosity according to Eq. (6.34) is the consequence.

6.4.3.3 Crabbed Waist Scheme

In the case of a large crossing angle, the collision point of particles is displaced. Schematically this is shown in Figs. 6.21 and 6.22.

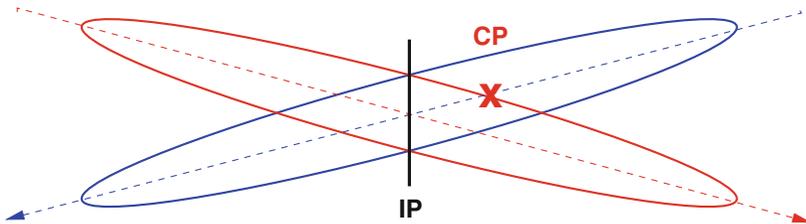


Fig. 6.21 Collision with large crossing angle and longitudinally displaced collision point

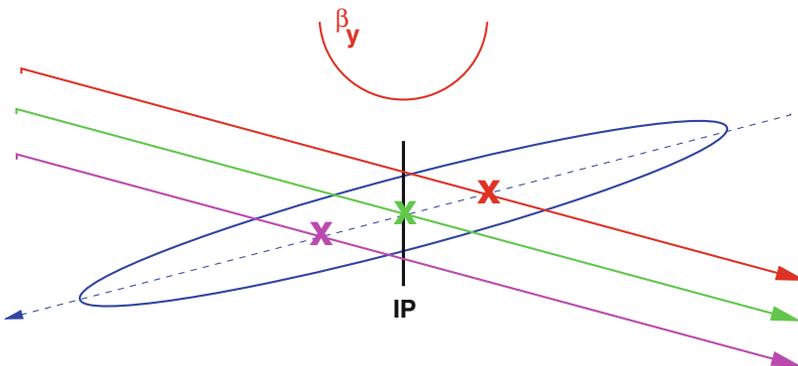


Fig. 6.22 Collision with large crossing angle and longitudinally displaced collision point. Shown for three particles with different amplitudes

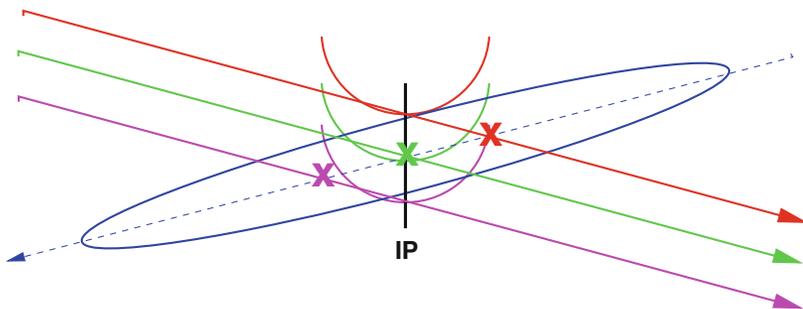


Fig. 6.23 Collisions with different vertical β -functions

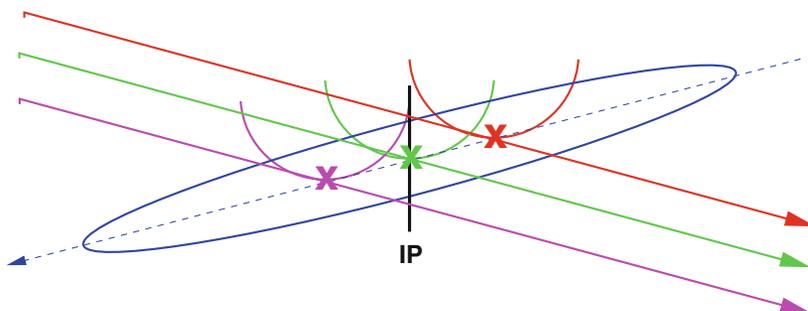


Fig. 6.24 All collisions at a minimum of the vertical β -functions using a crabbed waist scheme

One possible consequence can be the coupling between the transverse and longitudinal plane. Such a coupling is particularly bad for flat beams since the vertical beam size will increase significantly.

In Fig. 6.23 the vertical β -function is indicated and the result of this effect is that the particles collide at positions with different vertical β -functions.

This can be mitigated [42] by making the vertical waist (β_y^{min}) amplitude dependent in the horizontal plane Fig. 6.24. All particles collide now at the minimum of the vertical β -function.

It should be emphasized that the main purpose of such a scheme is not to reduce a geometrical loss but to reduce the coupling. Therefore it is of interest only for flat beams.

This scheme is established using two sextupoles.

6.4.4 Integrated Luminosity and Event Pile Up

The maximum luminosity, and therefore the instantaneous number of interactions per second, is very important, but the final figure of merit is the so-called integrated

luminosity:

$$\mathcal{L}_{\text{int}} = \int_0^T \mathcal{L}(t') dt' \quad (6.35)$$

because it directly relates to the number of observed events:

$$\mathcal{L}_{\text{int}} \cdot \sigma_p = \text{number of events of interest} \quad (6.36)$$

The integral is taken over the sensitive time, i.e. excluding possible dead time. The unit of the integrated luminosity is cm^{-2} and often expressed in inverse barn ($1 \text{ barn}^{-1} = 10^{24} \text{ cm}^{-2}$).

Another important parameter for a beam with high luminosity and bunched beams are the number of collisions per bunch crossing, the so-called pile up. In particular for collisions with a large cross section this can become a problem. In the case of the LHC, bunch crossings occur every 25 ns and the expected pile up is more than 20 for proton-proton collisions. The challenge is to maximise the useful luminosity while keeping the pile up to a level that can be handled by the particle detectors.

6.4.5 Measurement and Calibration of Luminosity

To obtain the exact integrated luminosity, it has to be recorded continuously. It is rather straightforward to obtain a counting rate directly proportional to the total interaction rate dR/dt . This relative signal has to be calibrated to deliver the absolute luminosity. We have already seen some effects that affect the absolute luminosity and therefore to a large extent the luminosity measurement. In particular the crossing angle and the luminous region are of importance since they have immediate implications for the geometrical acceptance of the instruments.

In principle one can determine the absolute luminosity when all relevant beam parameters are known, i.e. the bunch intensities, beam sizes (r.m.s. in case of unknown beam profiles) and the exact geometry. However the precise measurement of beam sizes is a challenge, in particular for hadron colliders when a non-destructive measurement is required. When the energy spread in the beams is large (e.g. some e^+e^- colliders), a residual dispersion at the interaction point increases significantly the beam size and must be included.

There exist other methods which relate the counting rate to well known processes which can be used for calibration. We shall discuss several methods for both, lepton and hadron colliders.

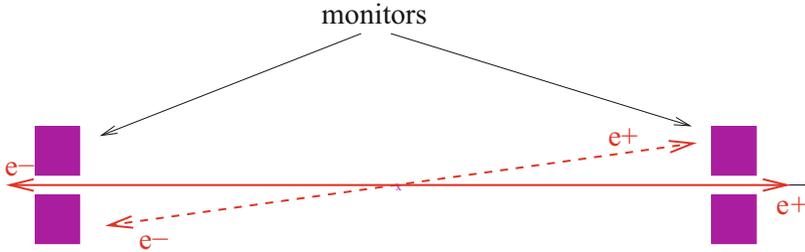


Fig. 6.25 Principle of luminosity measurement using Bhabha scattering for e^+e^- colliders

6.4.6 Absolute Luminosity: Lepton Colliders

Once the relative luminosity is known, a very precise method is to compare the counting rate to well known and calculable processes. In case of e^+e^- colliders these are electromagnetic processes such as elastic scattering (Bhabha scattering). The principle is shown in Fig. 6.25. Particle detectors are used to measure the trajectories at very small angles and with a coincidence of particles on both sides of the interaction point. For a precise measurement one has to go to very small angles since the elastic cross section σ_{el} has a strong dependence on the scattering angle ($\sigma_{el} \propto \Theta^{-3}$).

Furthermore, the cross section diminishes rapidly with increasing energy ($\sigma_{el} \propto \frac{1}{E^2}$) and the result may be small counting rates. At LEP energies with $\mathcal{L} = 10^{30} \text{ cm}^{-2} \text{ s}^{-1}$ one can expect only about 25 Hz for the counting rate. Background from other processes can become problematic when the signal is small.

6.4.7 Absolute Luminosity: Hadron Colliders

For hadron colliders two types of calibration have become part of regular operation, the measurement of the beam size by scanning the beam and the calibration with the cross section for small angle scattering. The determination of the bunch intensities is usually easier, although non-trivial in the case of a collider with several thousand bunches.

6.4.7.1 Measurement by Profile Monitors and Beam Displacement

Typical profile measurement devices are wire scanners where a thin wire is moved through the beam and the interaction of the beam with the wire gives the signal. For high intensity hadron beams this has however limitations. Non-destructive devices

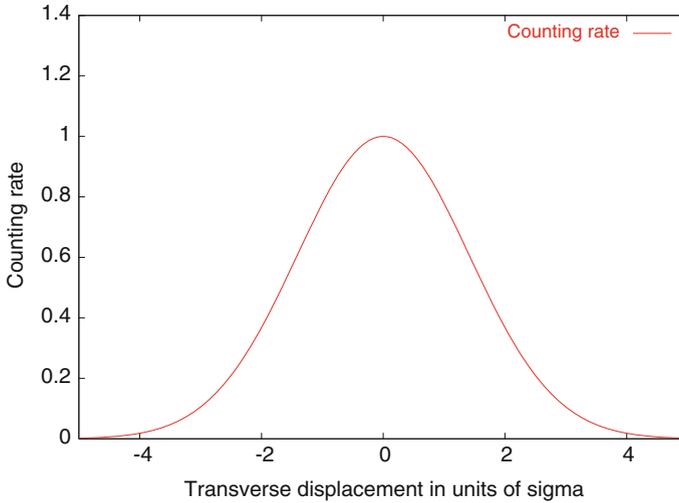


Fig. 6.26 Principle of luminosity measurement using transverse beam displacement

such as synchrotron light monitors are available but the emitted light from hadrons is often not sufficient for a precise measurement.

An alternative is to measure the beam size by displacing the two beams against each other. The relative luminosity reduction due to this offset can be measured and is described by the formula (6.28) developed earlier:

$$\mathcal{L}(d)/\mathcal{L}_0 = W = e^{-\frac{d^2}{4\sigma^2}} \quad (6.37)$$

where d is the separation between the beams and the measurement of the luminosity ratio is a direct measurement of W . This method was already used in the CERN Intersection Storage Rings (ISR) and known as “van der Meer scan”.

The expected counting rate of such a scan is shown in Fig. 6.26. A fit to the above formula gives the beam size. A drawback of this method is the distortion of the beam optics in case of very strong beam-beam interactions [40]. This effect has to be evaluated carefully.

6.4.7.2 Absolute Measurement with Optical Theorem

This method is similar to the measurement of Bhabha scattering for e^+e^- colliders but requires dedicated experiments and often special machine conditions.

The total elastic and inelastic counting rate is related to the luminosity and the total cross section (elastic and inelastic) by the expression:

$$\sigma_{tot} \cdot \mathcal{L} = N_{inel} + N_{el} \quad (\text{Total counting rate}) \quad (6.38)$$

The key to this method is that the total cross section is related to the elastic cross section for small values of the momentum transfer t by the so-called optical theorem [43]:

$$\lim_{t \rightarrow 0} \frac{d\sigma_{el}}{dt} = (1 + \rho^2) \frac{\sigma_{tot}^2}{16\pi} = \frac{1}{\mathcal{L}} \frac{dN_{el}}{dt} \Big|_{t=0} \quad (6.39)$$

Therefore the luminosity can in principle be calculated directly from experimental rates through:

$$\mathcal{L} = \frac{(1 + \rho^2) (N_{inel} + N_{el})^2}{16\pi (dN_{el}/dt)_{t=0}} \quad (6.40)$$

All counting rates, the total number of events $N_{inel} + N_{el}$ and the differential elastic counting rate dN_{el}/dt at small t have to be measured with high precision. This requires a very good detector coverage of the whole space (4π) for the inelastic rate and the possibility to measure to very small values of t .

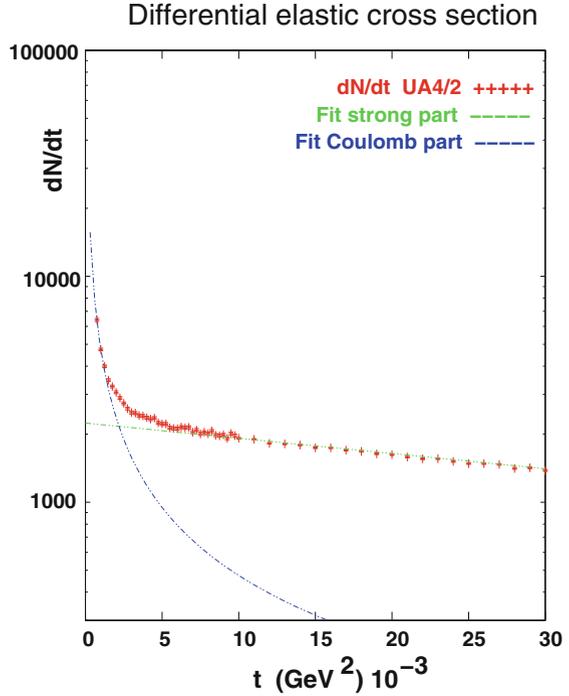
A slightly modified version of the above uses the Coulomb scattering amplitude which can be precisely calculated. The elastic scattering amplitude is a superposition of the strong (f_s) and Coulomb (f_c) amplitudes, the latter dominates at small t . We can re-write the differential elastic cross section $\frac{d\sigma_{el}}{dt}$:

$$\lim_{t \rightarrow 0} \frac{d\sigma_{el}}{dt} = \frac{1}{\mathcal{L}} \frac{dN_{el}}{dt} \Big|_{t=0} = \pi |f_c + f_s|^2 \simeq \pi \left| \frac{2\alpha_{em}}{-t} + \frac{\sigma_{tot}}{4\pi} (\rho + i) e^{B\frac{t}{2}} \right|^2 \simeq \frac{4\pi\alpha_{em}^2}{t^2} \Big|_{|t| \rightarrow 0} \quad (6.41)$$

If the differential cross section is measured over a large enough range, the unknown parameters σ_{tot} , ρ , B and \mathcal{L} can be determined by a fit. A measurement [44–46] together with some crude fits is shown in Fig. 6.27 to demonstrate the principle. The advantage of this method is that it can be performed measuring only elastic scattering without the need of a full coverage to measure N_{inel} . It is therefore a good way to measure the luminosity (and total cross section σ_{tot} and interference parameter ρ !) although the previous method is of more practical importance for regular use.

The measurement of the Coulomb amplitude usually requires dedicated experiments with detectors very close to the beam (e.g. with so-called Roman Pots) and therefore special parameters such as reduced intensity and zero crossing angle. Furthermore, in order to measure very small angle scattering, one has to reduce the divergence in the beam itself ($\sigma' = \sqrt{\epsilon/\beta}$). For that purpose special running conditions with a **high** β^* at the collision point are often needed ($\beta^* > 1000$ m) [45]. The precision of such a measurement is however as good as a few percent.

Fig. 6.27 Principle of luminosity measurement using optical theorem in proton proton (antiproton) collisions



6.4.8 Luminosity in Linear Colliders

In linear colliders the beams collide only once and to get a high luminosity a very small beam size and therefore small β at the collision point are required.

This implies additional effects such a beam disruption and an enhanced luminosity due to the so-called pinch effect.

Due to very strong field of the quadrupoles of the final focusing, significant synchrotron radiation is produced.

6.4.8.1 Disruption and Luminosity Enhancement Factor

The basic formula for the luminosity of a linear collider is shown in Eq. (6.42).

$$\mathcal{L} = \frac{N^2 f_{rep} n_b}{4\pi \overline{\sigma_x} \overline{\sigma_y}} \rightarrow \mathcal{L} = \frac{H_D \cdot N^2 f_{rep} n_b}{4\pi \sigma_x \sigma_y} \tag{6.42}$$

The revolution frequency has to be replaced by the repetition rate f_{rep} of the colliding bunches.

The luminosity is increased by the enhancement factor H_D which takes into account the reduction of the nominal beam size by the disruptive field (pinch effects).

This enhancement factor is related to the beam disruption parameter

$$D_{x,y} = \frac{2r_e N \sigma_z}{\gamma \sigma_{x,y} (\sigma_x + \sigma_y)} \quad (6.43)$$

For weak disruption ($D \ll 1$) and round beams the enhancement factor can be written as:

$$H_D = 1 + \frac{2}{3\sqrt{\pi}} D + \mathcal{O}(D^2) \quad (6.44)$$

When the disruption is strong or for flat beams, computer simulations are necessary.

6.4.8.2 Beamstrahlung

The strong synchrotron radiation (beamstrahlung) has two main effects:

- Spread of the centre of mass energy.
- Pair creation and background in the detectors.

It is parametrized by the parameter Y which can be written as the mean field strength in the rest frame, normalized to the critical field B_c :

$$Y = \frac{\langle E + B \rangle}{B_c} \approx \frac{5}{6} \frac{r_e^2 \gamma N}{\alpha \sigma_z (\sigma_x + \sigma_y)} \quad (6.45)$$

$$B_c = \frac{m^2 c^3}{e \hbar} \approx 4.4 \times 10^{13} G \quad (6.46)$$

6.5 Synchrotron Radiation and Damping

L. Rivkin

6.5.1 Basic Properties of Synchrotron Radiation

Charged particles radiate when they are deflected in the magnetic field [47] (transverse acceleration) [see also Sect. 11.1 for a more detailed treatment]. In the

ultra-relativistic case, when the particle speed is very close to the speed of light, $\beta \approx c$, most of the radiation is emitted in the forward direction [48] into a cone centred on the tangent to the trajectory and with an opening angle of $1/\gamma$, where γ is the Lorentz factor (since for a few GeV electron or a few TeV proton, $\gamma \approx 1000$, the photon emission angles are within a milliradian of the tangent to the trajectory).

The power emitted by a particle is proportional to the square of its energy E and to the square of the deflecting magnetic field B :

$$P_{SR} \propto E^2 B^2, \quad (6.47)$$

and in terms of Lorentz factor γ and the local bending radius ρ can be written as follows:

$$P_{SR} = \frac{2}{3} \alpha \hbar c^2 \frac{\gamma^4}{\rho^2}, \quad (6.48)$$

where α is the fine-structure constant and the Plank's constant is given in a convenient conversion constant:

$$\alpha = \frac{1}{137} \quad \text{and} \quad \hbar c = 197 \text{ MeV fm}. \quad (6.49)$$

The emitted power is a very steep function of both the particle energy and particle mass, being proportional to the fourth power of γ .

Integrating the above expression around the machine we obtain the amount of energy lost per turn:

$$U_0 = \frac{4\pi}{3} \alpha \hbar c \frac{\gamma^4}{\rho}. \quad (6.50)$$

The emitted radiation spectrum consists of harmonics of the revolution frequency and peaks near the so-called critical frequency or critical photon energy. It is defined such that exactly half of the radiated power is emitted below it:

$$\varepsilon_c = \frac{2}{3} \hbar c \frac{\gamma^3}{\rho}. \quad (6.51)$$

On the average a particle then emits $n_c \approx 2\pi\alpha\gamma$ photons per turn.

6.5.2 Radiation Damping

In a storage ring the steady loss of energy to synchrotron radiation is compensated in the RF cavities, where the particle receives each turn the average amount of energy

lost. The energy lost per turn is normally a small fraction of the total particle energy, typically of the order of one part per thousand.

Transverse Oscillations

Since the radiation is emitted along the tangent to the trajectory, only the amplitude of the momentum changes. As the RF cavities increase the longitudinal component of the momentum only, the transverse component is damped exponentially with the damping rate of the order of U_0 per revolution time. A typical transverse damping time corresponds simply to the number of turns it would take to lose the amount of energy equal to the particle energy. The damping times are very fast, in case of a few GeV electron ring being on the order of a few milliseconds.

$$A_{\perp} = A_0 e^{-\frac{t}{\tau}}, \text{ where } \frac{1}{\tau} = \frac{U_0}{2ET_0}. \quad (6.52)$$

In a given storage ring the damping time is inversely proportional to the cube of the particle energy.

Longitudinal or Synchrotron Oscillations

Synchrotron oscillations are damped because the energy loss per turn is a quadratic function of the particle's energy. The damping rate is typically twice the rate for transverse oscillations.

Damping Partition Numbers and Robinson Theorem

For particles that emit synchrotron radiation the dynamics is characterized by the damping of particle oscillations in all three degrees of freedom. In fact, the total amount of damping (Robinson theorem [49]), i.e. the sum of the damping decrements depends only on the particle energy and the emitted synchrotron radiation power:

$$\frac{1}{\tau_x} + \frac{1}{\tau_y} + \frac{1}{\tau_{\varepsilon}} = \frac{2U_0}{ET_0} = \frac{U_0}{2ET_0} (J_x + J_y + J_{\varepsilon}) \quad (6.53)$$

where we have introduced the usual notation of *damping partition numbers* that show how the total amount of damping in the system is distributed among the three degrees of freedom. A typical set of the damping partition numbers is (1,1,2) and their sum is, according to the Robinson theorem, a constant.

$$J_x + J_y + J_{\varepsilon} = 4. \quad (6.54)$$

Adjustment of Damping Rates

The partition numbers can differ from the above values, while their sum remains a constant. In fact, under certain circumstances, the motion can become “anti-damped”, i.e. the damping time can become negative, leading to an exponential growth of the oscillations amplitudes. From a more detailed analysis of damping

rates [50] the damping time can be written as

$$\frac{1}{\tau_\varepsilon} = \frac{U_0}{2ET_0} (2 + \mathcal{D}), \quad \text{and} \quad \frac{1}{\tau_x} = \frac{U_0}{2ET_0} (1 - \mathcal{D}), \quad \text{where} \quad \mathcal{D} \equiv \frac{\oint \frac{D}{\rho} \left(2k + \frac{1}{\rho^2}\right) ds}{\oint \frac{ds}{\rho^2}}. \quad (6.55)$$

The constant introduced above is an integral of the dispersion function \mathcal{D} and the magnetic guide field functions, i.e. bending radius and gradient around the ring and is independent of the particle energy. It deviates substantially from zero only when a particle encounters combined function elements, i.e. where the product of the field gradient and the curvature is non-zero. The damping partition numbers then are:

$$J_x = 1 - \mathcal{D}, \quad J_\varepsilon = 2 + \mathcal{D}, \quad J_x + J_\varepsilon = 3. \quad (6.56)$$

The vertical damping partition number is usually unchanged as the vertical dispersion is zero in storage rings that are built in one (horizontal) plane.

The amount of damping can be repartitioned between the horizontal and energy-time oscillations by altering the value of the \mathcal{D} constant [50]. This can be achieved by either using combined function magnetic elements in the lattice, or by introducing a special combined function wiggler magnet (so-called Robinson wiggler). Values of horizontal partition number as high as 2.5 have been obtained that way. Values of $\mathcal{D} > 1$ lead to anti-damping of horizontal betatron oscillations, while for $\mathcal{D} < -2$ the synchrotron oscillations become unstable.

6.6 Computer Codes for Beam Dynamics

Werner Herr

6.6.1 Introduction

The design and operation of an accelerator today is unthinkable without the help of computer codes, the reason being large, complex structures (like in the case of big accelerators and colliders, e.g. LHC) or complications in the beam dynamics of small or special purpose machines (e.g. FFAG). Their complexity does not allow the computation with pencil and paper. Here we address only the codes for beam dynamics, i.e. special codes for the design of accelerators components such as magnets or RF equipment will not be treated but can be found in the literature. The main fields where beam dynamics codes are essential are:

- Determination of parameters and the design of beam lines and accelerators
- Evaluation of performance
- Control, machine protection and operation

Different classes of codes are used in these fields which also resemble the life cycle of an accelerator.

Given the scope of this handbook and the rapid evolution of computer codes and software techniques, we do not attempt to provide a list of existing codes, but rather will describe the main features, techniques and applications of the different types of codes. Details and access to existing codes can be found in computer code libraries on the internet. A supported library is provided by the Los Alamos Accelerator Code Group (LAACG) [51], another one supported by Astec (UK) [52]. It contains links to popular and frequently used codes from many laboratories and institutions.

6.6.2 *Classes of Beam Dynamics Codes*

The different classes of codes can be divided according to their application:

- General purpose optics codes
- Beam dynamics of single particles
- Beam dynamics of multi particles

Optics codes are used mainly in the initial design phase of an accelerator, rings as well as beam lines and linear accelerators. The evaluation of the performance (stability etc.) is done using codes to simulate the beam dynamics of single particles as well as ensembles of particles and their interaction with the environment or other particles in the beam(s).

6.6.3 *Optics Codes*

A large group of computer codes for beam dynamics are used to design the lattice of an accelerator or beam line and to compute and optimize the optical parameters. The range of available codes extends from small codes for pedagogical purpose to large general purpose programs. Such codes can have easily 100,000 lines of codes or more. The accelerator physics is described in the existing literature [53] and in this handbook. The main applications of general purpose optics codes are:

- Determination of main parameters and the computation of linear and non-linear optics. This implies to find periodic solutions for the optical parameters and the closed orbit.
- Parameter matching (optical/geometrical) and lattice optimization, i.e. the properties of elements are varied until the optical functions assume their desired values.
- Simulation of imperfections and algorithms for their corrections.
- Simulation of synchrotron radiation and evaluation of radiation integrals to derive estimates for parameters (e.g. equilibrium emittances) in lepton machines.

The result should be a consistent set of parameters fulfilling the design requirements. They are the basis for the design of machine elements.

Depending on the complexity of the problem, different techniques are in use for optics codes. The majority of these codes rely on the description of machine elements using maps, which can be of higher order for non-linear elements. In the simplest case for the description of linear machines the maps become matrices and are therefore often referred to as “matrix codes” [54]). The concatenation of the matrices provide a matrix for the entire ring and its analysis gives the optical parameters, closed orbit etc.

Another technique is to follow the particles through the accelerator, i.e. integrating the equation of motion in the electromagnetic fields of the machine elements. The analysis of the results of these “tracking programs” provides the required parameters and information about the stability of the machine (for some details see [54]).

Dealing with complex machines, other considerations may become important such as e.g.:

- Definition of an input language which can be used by other programs. This input language defines the sequence of elements, i.e. the ring or a beam line, as well as the properties of the elements such as e.g. their types (dipole, quadrupole,...), lengths and strengths.
- For large machines with a large number of elements the interface to a data base may be required. Large machines such as the LHC or future colliders have several thousand elements.
- An interface to the control system for on-line modelling is desirable

6.6.4 Single Particle Tracking Codes

To evaluate the performance of accelerators, in particular multi pass, i.e. circular machines, one has to deal with complex iterative processes. The standard perturbation theories can fail to correctly describe the behaviour beyond leading orders. Single particle tracking codes are successfully used when analytical methods fail to describe the effect of non-linear forces on the stability of the particles. Many tracking codes have been developed together with the necessary tools to analyse the results and from the simulation point of view the treatment of non-linear effects is well established. Conceptually, in a tracking code the equation of motion of a particle in an accelerator element is solved and the phase space coordinates of the particle are followed through all elements of the accelerator or beam line. To obtain the desired information, it may be necessary to repeat this process for up to 10^7 turns which require appropriate algorithms and techniques to avoid numerical problems. Similar problems exist and some of these techniques have been developed for celestial mechanics. In order to draw conclusions from the tracking data it is necessary to provide tools to allow a qualitative and quantitative understanding of

the results [55]. The outcome of the analysis allows to answer the most important questions for the design of a machine such as:

- Stability of particle motion
- Dynamic aperture
- Specifications for the properties of machine elements
- Optimization of the particle stability

In general the results of these studies are used in an iterative procedure to improve and optimize the design of the machine.

6.6.4.1 Techniques

A requirement for all techniques employed for particle tracking is that the associated maps must be symplectic. To solve the equation of motion, most programs use explicit canonical integration techniques, e.g.:

- Thin lens tracking (most common since they are automatically symplectic and fast)
- Ray tracing (accuracy by slicing into large number of steps, but time consuming)
- Symplectic integration (see [54] and references therein)

6.6.4.2 Analysis of Tracking Data

Some of the analysis techniques are discussed in the chapter on non-linear dynamics in this handbook in more detail and some are mentioned here for completeness:

- Taylor maps using Truncated Power Series Algebra (TPSA, [54])
- Lie algebraic maps [54, 56]
- Normal form analysis

The results of the analysis include non-linear resonances and distortion, non-linear tunes shift with amplitude and an evaluation of the long term stability. In all cases the interpretation of the results requires a careful analysis of the range where the data is meaningful to avoid wrong conclusions. Typical problems are numerical effects which can lead to unphysical features.

6.6.5 *Multi Particle Tracking Codes*

Multi particle tracking codes are used when we are concerned with the behaviour of an ensemble of particle. The calculations largely rely on techniques developed for single particle dynamics. Typical applications are the simulation of:

- Space charge effects, mutual interaction of particles within the same beam.

- Collective instabilities and interaction with environment (impedance)
- Beam-beam effects in case of particle colliders, i.e. the interactions with the fields produced by the counter-rotating beam.
- Electron cloud effects, i.e. secondary electron production by synchrotron radiation

A key issue for multi particle simulation codes is the evaluation of the electromagnetic fields produced by the beams or the environment. New techniques and the availability of parallel computing facilities have allowed vast progress in this field in the last 20 years.

6.6.6 *Machine Protection*

For large energy and high intensity machines the protection of the machine elements becomes an important part of the design. Simulation codes have to include the interaction of particles with matter.

6.7 **Electron-Positron Circular Colliders**

M. E. Biagini · J. M. Jowett

Electron-positron (e^+e^-) collider rings have been a mainstay of both discovery and precision physics for half a century: discovery, since the simple initial state can create any particle coupled to the electromagnetic field; precision, from the combination of high luminosity and large cross-sections at a rich spectrum of resonances up to $\sqrt{s} \simeq 200\text{GeV}$. While the fundamentals of these machines have remained in essence the same, the technology has matured to the point where luminosities of the latest “factories” exceed what was thought possible in the 1970s and early 1980s by 2–3 orders of magnitude.

These colliders are based on the principle of the synchrotron (Sect. 1.2.6) although the name is barely appropriate for those which enjoy the advantage of full-energy injection. Beams are necessarily bunched by an RF system, which must provide sufficient voltage to compensate the energy lost by synchrotron radiation.

6.7.1 *Physics of Electron-Positron Rings*

Consider an ideal storage ring constructed with bending and focussing magnets such that a particle of charge e and *constant* momentum p_0 could circulate on a stable closed orbit, O_{xy} , in transverse phase space (x, p_x, y, p_y) , with local radius of

curvature $\rho(s)$. The orbits and optical functions ($\beta_{x,y}(s)$, dispersion $D_{x,y}(s)$, etc., Chap. 2) of such *hypothetical, non-radiating* particles are a construct useful in the description of e^\pm dynamics. Real, radiating, e^+ of energy $E = \sqrt{p^2 c^2 + m^2 c^4} \simeq pc \simeq p_0 c$ can circulate in a phase-space neighbourhood of O_{xy} provided RF cavities of a proper frequency and sufficient voltage are added to compensate the average radiative energy loss and provide longitudinal phase stability (e^- can circulate in the opposite direction). In a semi-classical picture [49, 53, 57, 58], e^\pm emit photons at random times according to the classical synchrotron radiation spectrum [47, 48] and make stochastic transitions between betatron trajectories corresponding to their instantaneous momenta. This picture can be understood [59] by recognising that a storage ring differs from an atom in that changes, $\Delta n = nu/E$, in orbital quantum number, n , corresponding to typical photon emissions of energy u , satisfy $n \gg \Delta n \gg 1$.

There is no deterministic closed orbit but the full 6D *central orbit*, O_{xyz} of a bunch of many electrons normally coincides with the attractive stable orbit calculated by averaging over photon emissions to include only the classical deterministic part of the synchrotron radiation (this includes the stable phase with respect to the RF system). If the domain of attraction of this orbit is large enough, the beam can have a good lifetime (Eq. 6.60 below). Because of the energy variation round the ring (localised RF cavities giving “energy-sawtooth”), the transverse projection of O_{xyz} does not coincide with O_{xy} . Figure 6.28 shows an example.

Neglecting intensity-dependent phenomena, the equilibrium dimensions of the beam are macroscopic quantum effects determined by the balance between radiation damping (the dependence of the classical radiation lost in magnetic fields on the energy, [49, 57, 58] and Sect. 6.5), and the quantum fluctuations (discrete photon nature) of the synchrotron radiation [57, 58]. Generally, the effects are linear enough that the core of the distribution is gaussian in each normal mode coordinate.

The mean-square fractional energy spread in the beam is

$$\frac{\sigma_E^2}{E^2} = \frac{55}{32\sqrt{3}} \frac{\hbar}{mc} \left(\frac{E_0}{mc^2} \right)^2 \frac{\oint |G^3| ds}{J_z \oint G^2 ds} \simeq \frac{1}{2} \gamma^2 \frac{\lambda_e}{\rho_0}, \quad (6.57)$$

where $G = eB/p_0 c = \rho^{-1}$ is the inverse of the local bending radius of O_{xy} , $\oint \dots ds$ denotes an integral around O_{xy} , J_z is the longitudinal damping partition number (Sect. 6.5), $\lambda_e = \hbar/mc$ is the reduced Compton wavelength of the electron and the last equality holds to the extent that $G(s)$ is zero or has a constant value $1/\rho_0$ (isomagnetic ring).

Economic arguments, balancing construction cost against power consumption, are sometimes invoked to derive a scaling of radius with energy squared but this only applies for the highest energy rings with a few bunches (see [60] for the scaling of design parameters). More generally, the chromaticity correction and dynamic aperture constraints (Sect. 3.4) in collider rings require $6\sigma_E/E \lesssim 1\%$, so imposing a minimum radius $\rho/m \approx 0.26(E_0/\text{GeV})^2$. The spread in centre-of-mass energies of collisions $\sigma_{\sqrt{s}} = \sqrt{2}\sigma_E$ (if $D_x = 0$ at the collision point) should also be kept small.

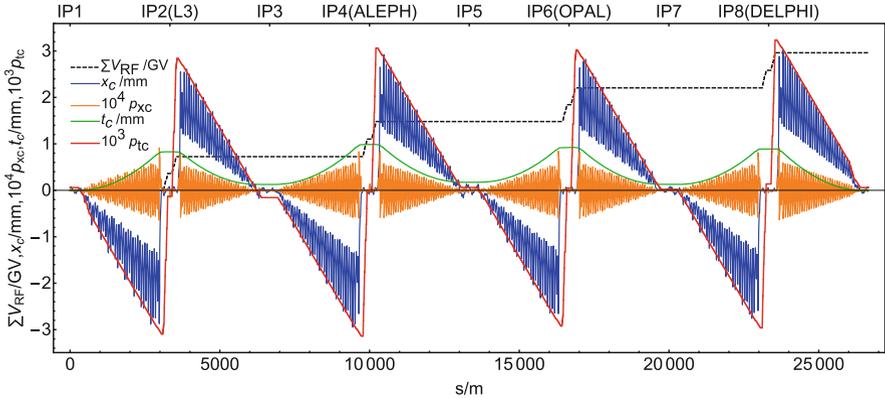


Fig. 6.28 The cumulative RF voltage (black dashed line) around the ring and four components of the ideal six-dimensional closed orbit of the e^+ beam in CERN's LEP collider, at a central beam energy of 94 GeV in an optics used in 1998. The fractional deviation of the beam energy on the closed orbit p_{tc} (red) exhibits the “energy sawtooth” due to the energy lost by synchrotron radiation in the eight arcs and its replenishment by RF systems located around the experimental interaction points. The conjugate time-lag coordinate t_c (green) reflects the corresponding path-length changes. The horizontal closed orbit $x_c = D_x p_{tc} + x_{cB}$ (blue) is a combination of the local dispersion orbit and a forced betatron oscillation and its conjugate p_{xc} . Without radiation, these four orbit components would be zero. LEP was a single ring collider with e^+ and e^- beams of similar intensity circulating in the same beam pipe. In normal operation, the average x_c for the two beams, which were approximately equal and opposite, $x_c^+ \simeq -x_c^-$, was measured and corrected to the central trajectory. At higher energies, these orbits could be separated by a few cm near the RF systems

The equilibrium horizontal emittance for flat rings without betatron coupling is

$$\varepsilon_x = \frac{55}{32\sqrt{3}} \frac{\hbar}{mc} \left(\frac{E_0}{mc^2} \right)^2 \frac{\oint |\mathcal{H}G^3| ds}{J_x \oint G^2 ds}, \quad (6.58)$$

where $\mathcal{H} = \beta_x D_x^2 + 2\alpha_x D_x D'_x + \gamma_x D_x'^2$ is a quadratic form constructed from the dispersion and betatron matrix (Sect. 2.1). Together, Eqs. (6.57) and (6.58) give the mean-square equilibrium beam size at any point in the ring

$$\sigma_x^2 = \left\langle (x_\beta + D_x (E - E_0) / E_0)^2 \right\rangle = \beta_x \varepsilon_x + D_x^2 (\sigma_E / E)^2. \quad (6.59)$$

The vertical emittance is usually smaller and due to some coupling of horizontal betatron motion into the vertical and vertical dispersion from orbit errors or other vertical bends. More general formalisms [53, 61] describe the radiation-generated emittances for the eigenmodes of linear oscillations about general six-dimensional central orbits.

A true equilibrium (strictly, stationary) state does not exist because the quantum fluctuations lead to loss from the tails of the beam with lifetimes for the three modes

given by

$$\tau_{q,u} = \frac{1}{2} \tau_u \frac{e^{\xi_u}}{\xi_u}, \quad \text{where} \quad \xi_u = \frac{A_u^2}{2\sigma_u^2}, \quad \text{for} \quad u = x, y, z, \quad \xi_u \gtrsim 20, \quad (6.60)$$

where the τ_u are the radiation damping times and A_u are appropriate acceptances [53, 57, 58]. For the synchrotron mode, A_z is the RF bucket half-height

$$\left(\frac{A_z}{E_0}\right)^2 = \frac{2U_0}{\pi |\eta| h_{\text{RF}} E_0} \left[\sqrt{(eV_{\text{RF}}/U_0)^2 - 1} - \arccos(U_0/eV_{\text{RF}}) \right]. \quad (6.61)$$

For adequate lifetime at small intensity, the mechanical and dynamic apertures and RF voltage must be large enough.

The bunch length is given by $\sigma_z = c|\eta|\sigma_E/(\omega_s E_0)$ where ω_s is the angular synchrotron frequency and $\eta \simeq \alpha_c$ the frequency slip factor ([53], Sects. 2.5.2 and 2.5.3).

6.7.2 Design of Colliders

Colliders are designed from the interaction point outwards. The classical design is based on head-on collisions of flat beams. However a number of other configurations have been explored and the most promising among them is described in the following section. In the classical scheme, luminosity (Sect. 6.4) is maximised by achieving very flat beams, $\kappa = \varepsilon_y/\varepsilon_x \ll 1$; we consider only beams of equal energy, size and single bunch population, N_b , colliding head-on, with $\sigma_y^* \ll \beta_y^*$; for generalisations see [53]. The beam-beam effect (Sect. 4.6.1) generally imposes maximum attainable values on the horizontal and vertical beam-beam parameters

$$\xi_{x,y} = \frac{r_e N_b \beta_{x,y}^*}{2\pi (E_0/mc^2) \sigma_{x,y}^* (\sigma_x^* + \sigma_y^*)}, \quad (6.62)$$

where r_e is the electron classical radius, $\beta_{x,y}^*$ and $\sigma_{x,y,z}^*$ are the optical functions and beam sizes at the collision point. Typically, one finds $\max \xi_y = 0.03 - 0.1$ with the highest values attained when the machine is very well corrected (favourable tunes, central orbits close to design, minimised vertical dispersion) and when radiation damping is strong. Then the luminosity (Sect. 6.4) can be expressed as

$$L = \frac{f_c N_b}{2r_e} \left(\frac{E_0}{mc^2} \right) \frac{(1 + \kappa) \xi_y}{\beta_y^*}, \quad (6.63)$$

where f_c is the frequency at which identical bunches collide; in the simplest case $f_c = k_b f_0$ where k_b is the number of bunches per beam.

The number of bunches in a single-ring collider is limited by the possibilities for separating the opposing beams at unwanted encounters, e.g., by local or long-range (“pretzel scheme”) electrostatic orbit bumps [53]. Collective effects limiting the single-bunch intensity (bunch-lengthening, transverse mode-coupling, see Chap. 4) are a major concern. In recent double-ring colliders, many more bunches can be stored. A crossing angle at the collision point separates the beams at encounters in the adjacent common section of the beam pipe. In recent years, the highest luminosity collider designs have adopted a new scheme described in the following section.

Multi-bunch collective effects and other limits related to total beam current (e.g., component heating by wakefields or synchrotron radiation, beam-loading, electron-cloud, ion-trapping [53]) tend to dominate. The impedance and surface properties of the vacuum chamber are critical.

Integrated luminosity can be further maximised in moderate energy rings for which a full-energy injector is available by topping up the intensity of the stored beam rather than dumping and refilling. The static magnetic configuration (no ramp and squeeze cycle) simplifies operation dramatically.

The arcs of collider rings are usually composed of FODO cells whose length and phase advance determine the emittance through Eq. (6.58). To minimise radiation power, the bending magnets are made as long as possible. In the highest energy rings, the quadrupoles must also be lengthened.

Low- β insertions (Sect. 6.2.1) provide small values of β_y^* at the interaction point(s) of the experiment(s) in long straight sections. These can also accommodate the accelerating cavities of the RF system, beam instrumentation and wiggler magnets and are connected to the arcs via dispersion suppressors.

Wiggler magnets modify the radiation damping, bunch length and/or emittance by contributing additional terms [53] with large $|G|$ to the integrals in Eqs. (6.57) and (6.58), so providing additional flexibility to maximise performance (e.g., at lower energy).

Sextupoles incorporated in the arcs must correct the large chromatic aberrations generated in the low- β quadrupoles while preserving adequate dynamic aperture (Sect. 3.4.4).

Many variations on this classical e^+e^- collider design are possible with new interaction region concepts showing promise (Sect. 6.4) in overcoming the need for ever-increasing beam current and ever-shorter bunches.

At higher intensities, phenomena such as the Touschek effect and intra-beam scattering [53], sometimes in combination with non-linear single particle dynamics or beam-beam effects, can reduce the lifetime below the values implied by Eq. (6.60); see Chap. 3 and Sect. 4.6.

6.7.3 Large Piwinski Angle and Crab Waist Collision Scheme

The need for precision measurements of rare decay modes with small cross sections at e^+e^- factories has driven requirements on peak luminosity to unprecedented levels. Conventional collision schemes, see Eq. (6.63), are based on pushing up the beam currents, lowering the β_y^* , and increasing the beam emittance so as not to exceed the beam-beam tune-shift limits. Passing from single to double ring colliders allowed the number of bunches to be increased considerably. However in order to avoid luminosity reduction due to parasitic (or long-range) bunch encounters near the collision point, beams had to be collided with a small horizontal crossing angle rather than head-on. However, this approach has come to a dead end since high currents result in high power losses, beam instabilities and increased power consumption.

Because of the parabolic variation of $\beta_y(s) = \beta_y^* + s^2/\beta_y^*$ in the vicinity of the interaction point (IP), the longitudinal region in which individual particle collisions occur will include places where the effective $\beta_y(s) \gg \beta_y^*$ at the IP, and will therefore contribute less to the luminosity. This so-called *hour-glass* effect imposes a condition on the bunch-length: $\sigma_z \lesssim \beta_y(s)$. Unfortunately, shortening the bunch length is costly since it requires high voltage in the RF cavities, can excite collective instabilities, induce higher-order mode (HOM) heating in the beam pipe, and lead to coherent synchrotron radiation emission, which in turn deteriorates the bunch shape. On the other hand, increasing the bunch current leads to coupled bunch instabilities, HOM heating of the beam pipe, and higher wall-plug power.

A solution to these problems came with the idea of a new collision scheme, called “*Large Piwinski Angle and Crab Waist Sextupoles*” (LPA&CW), by P. Raimondi in 2006 [62]. This scheme has two main ingredients:

1. A large horizontal crossing angle at the IP, combined with very small horizontal beam size, resulting in a large Piwinski angle;
2. a pair of sextupoles, each placed on one side of the IP at a specific betatron phase from it.

The Piwinski angle is defined as:

$$\Phi = \frac{\sigma_z \tan(\theta)}{\sigma_x} \approx \theta \frac{\sigma_z}{\sigma_x}. \quad (6.64)$$

Consider two bunches with RMS beam size σ_x and bunch length σ_x , colliding at a horizontal crossing angle 2θ . For flat beams colliding at a small crossing angle $\theta \ll 1$ and large Piwinski angle $\Phi \gg 1$, the luminosity L and the tune-shifts scale as [63]:

$$L \propto \frac{N\xi_y}{\beta_y^*} \quad (6.65)$$

$$\xi_y \propto \frac{N}{2\theta\sigma_z} \sqrt{\beta_y^*/\epsilon_y} \quad (6.66)$$

$$\xi_x \propto \frac{N}{(2\theta\sigma_z)^2} \quad (6.67)$$

In the LPA scheme the Piwinski angle is increased by decreasing σ_x and increasing θ . The most relevant consequence is that the overlap area of the two colliding beams is now reduced, since it is proportional to σ_x/θ . As a plus, as can be seen from Eq. (6.67), the horizontal tune shift in this case drops like $(2\theta\sigma_z)^2$, so the beam-beam interaction can be considered as one-dimensional and only the vertical plane is relevant.

Now, the vertical β_y^* function at the IP can be decreased, as much as the focussing magnet technology allows, to be comparable to the overlap area size that, in this case, is smaller than the bunch length. In this case that is much smaller the bunch length, so relaxing the problems of HOM heating, coherent synchrotron radiation and excessive power consumption:

$$\beta_y^* \approx \frac{\sigma_x}{2\theta} \ll \sigma_z \quad (6.68)$$

This scheme has several advantages:

- a smaller spot size at the IP, leading to higher luminosity,
- a reduction of the vertical tune-shift parameter,
- the mitigation of synchro-betatron resonances.

Long range beam-beam interactions no longer limit the maximum achievable luminosity when the distance between bunches is short. These parasitic crossings become negligible because of the larger crossing angle and the smaller horizontal beam size. The separation at each encounter is larger in terms of σ_x .

However the large Piwinski angle itself may introduce new beam-beam resonances which can limit the maximum achievable tune shifts. The second ingredient of the LPA&CW scheme, the pair of *Crab Waist sextupoles*, is designed to solve this problem. The CW transformation causes the horizontal oscillations to modulate the vertical motion modulation and thereby suppresses the betatron and synchro-betatron resonances. The CW scheme is realised by installing a couple of sextupole magnets on the two sides of the IP, preferably in a high β and zero dispersion region. To provide the exact compensation the sextupoles be at π horizontal and a $\pi/2$ vertical betatron phase advance from the IP.

The CW transformation can be described by the Hamiltonian:

$$H = H_0 + \frac{1}{2\theta} x p_y^2 \quad (6.69)$$

where H_0 is the Hamiltonian of the particle's motion without the CW, x is the horizontal particle coordinate and p_y the vertical momentum. The effect of the CW

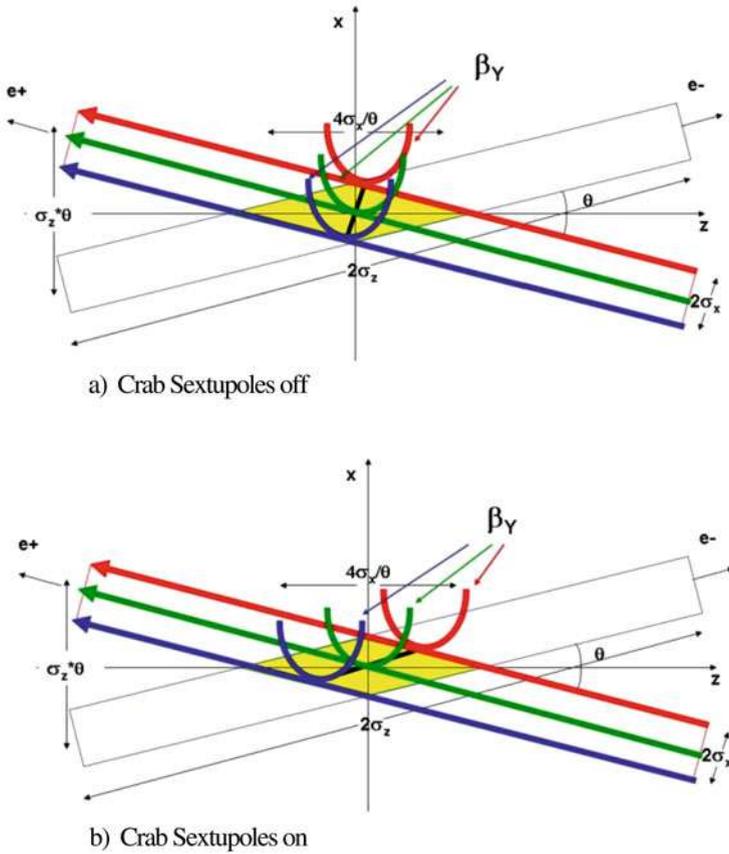


Fig. 6.29 Crab Waist collision scheme

transformation is a vertical betatron function *twist* according to:

$$\beta_y = \beta_y^* + \frac{(s - x/2\theta)^2}{\beta_y^*} \tag{6.70}$$

In this case, the β_y waist of one beam is twisted to be oriented along the central trajectory of the other beam. As a consequence, all particles, independently of their x position, collide at the minimum β_y spot of both beams, with an increase of few percent in the geometric luminosity due to the β_y redistribution along the overlapping beams area. A sketch of this shown in Fig. 6.29.

However the main CW effect is to suppress the betatron and synchro-betatron resonances which would arise due to the vertical motion modulation induced by the horizontal oscillations. This increases the space for the working betatron tunes

of the collider. Moreover beam-beam simulations showed that beam tails are very much reduced and the beam-beam blow-up is also suppressed.

The CW sextupoles strength should satisfy the following condition:

$$K = \frac{1}{2\theta} \frac{1}{\beta_y^* \beta_y} \sqrt{\frac{\beta_x^*}{\beta_y}} \quad (6.71)$$

where starred β values are those at the IP and the others are at the sextupole location. The CW sextupoles can reduce the dynamic aperture if there are other non-linearities between them. For this reason they should ideally be installed before the chromaticity correction sextupoles.

The (LPA&CW) collision scheme was first tested at the DAΦNE Φ-Factory in Frascati (Italy) in 2008 [64], by modifying the interaction region to increase the crossing angle, decrease both the β^* and allocate space for the sextupoles. The result was a boost in luminosity of about a factor of 4 and measurement of the beam profile showed that the bunches kept their Gaussian shape. This scheme has since then been adopted by all new collider designs worldwide (SuperKEKB, FCC-ee, various τ -charm Factory proposals). The collider SuperKEKB in Japan, however, has adopted the LPA scheme (which they called “*nano-beams*”) without the CW sextupoles because of the lack of space in the IR.

6.8 Hadron Colliders and Electron-Proton Colliders

K. Hanke · B. J. Holzer

6.8.1 Principles of Hadron Colliders

Hadron colliders are discovery machines which provide high centre-of-mass energy and cover a wide energy range. Contrary to electron-positron colliders, where the energy and quantum state of the initial particles is precisely known, the input conditions are less well defined in the case of proton-proton or proton-antiproton collisions. In fact, such collisions are, unlike e^+e^- annihilation, collisions of quarks and antiquarks the momenta of which are distributed according to the structure function of the hadron and are hence not precisely defined. As far as the analysis of the events is concerned, hadronic collisions result in a much larger number of tracks in the detector than in the case of e^+e^- annihilation, providing an additional challenge. From a machine physics point of view hadron machines have the enormous advantage that the particle beam energy is not limited by synchrotron radiation. This is because the proton mass is 2000 times higher than the one of the electron, and the energy loss due to synchrotron radiation scales with m^{-4} . What

is limiting the achievable beam energy in a hadron collider is the magnetic field to be provided by the dipole magnets in order to bend the particle beam on a circular trajectory. The radius of curvature of a particle in a dipole field is given by

$$\frac{1}{\rho} [m^{-1}] = \frac{eB}{p} = 0.2998 \frac{B [T]}{p [GeV/c]}$$

where ρ is the radius of curvature, B is the magnetic field, e is the elementary charge and p is the particle momentum. The quantity $B\rho$ is called the rigidity [65, 66]. The beam rigidity determines the B-field required to bend the beam on a circular trajectory with given bending radius. The maximum achievable B-field being limited to about 2 Tesla for normal conducting magnets, today's high-energy hadron colliders use superconducting bending magnets.

The quasi absence of synchrotron radiation leads to another feature of hadron colliders, which is the fact that there is no synchrotron radiation damping and the transverse emittance is hence determined and preserved throughout the injector chain. Emittance blow up, e.g. via injection mismatch, is therefore critical.

6.8.2 Proton-Antiproton Colliders

A machine with one single vacuum chamber, e.g. the Super Proton Synchrotron SPS ("SppbarS in this operation mode) in the 1980s or the Tevatron can accomplish the acceleration of protons and antiprotons.

During the years 1981–1987, the CERN SPS was operated as a proton-antiproton collider, providing high energy collisions for two major experiments located in adjacent sextants of the accelerator. This operation was first with three dense bunches of protons in collision with three rather weak bunches of antiprotons, with no separation of the beams at the unused crossing points. After increasing the antiproton production rate, six bunches per beam were used. The SPS has normal conducting bending magnets and a circumference of 6.9 km. The beam energy provided by the SPS as proton-antiproton collider was 315 GeV [67, 68]. The SppS was the first hadron collider operating with bunched beams. Before the commissioning of the machine it was debated if it was possible to collide proton and antiproton bunches, or if the beams would become unstable due to the presence of the beam-beam interaction without damping as in e^+e^- colliders. Its success demonstrated the feasibility of high energy hadron colliders [69].

Higher energy was achieved by the Tevatron, using superconducting magnets with a maximum B field of 4.5 T. The circumference of the machine is 6.28 km; comparable to that of the SPS. In the final stage of operations ("run II"), beams are injected at 150 GeV and accelerated to 980 GeV. The bunches (36+36) circulate in the same aperture, the protons clockwise and the antiprotons anticlockwise. The machine has a lattice with four dipoles followed by a quadrupole, with a total of 772+2 dipoles and 90+90 quadrupoles, plus a number of corrector magnets.

6.8.3 Proton-Proton Colliders

Proton-proton colliders require a dedicated magnet design with two separate vacuum chambers for the two equally charged beams. The first proton-proton collider was the ISR (Intersecting Storage Rings) at CERN. It consisted of two rings of 943 m length which were intersecting at eight points. Out of these eight intersection points six were used for experiments. The ISR was operated between 1970 and 1984. The top energy achieved for protons was 31.4 GeV/c. The ISR allowed not only proton-proton collisions, but stored and collided later also deuterons, alpha particles and antiprotons. The ISR pioneered a number of techniques which were beneficial which paved the path for future high energy colliders like the SPS.

The proton-proton collider with the highest energy ever built is the Large Hadron Collider (LHC) at CERN [70]. It uses superconducting magnets with two separate vacuum chambers for the two equally charged beams. The design field is 8.36 T, and the machine circumference 26.659 km which yields a design beam energy of 7 TeV. Higher field levels are being studied in the frame of possible further energy upgrades.

A machine with an even higher beam energy of up to 50 TeV is presently being studied by an international collaboration. The Future Circular Collider (FCC) has a hadron-hadron option (FCC-hh) with a beam energy of 50 TeV [71]. The latest design features a machine circumference of 97.75 km with a maximum dipole field of 15.7 T. The size of the machine is a compromise of civil engineering constraints and dipole feasibility.

6.8.4 Electron-Proton Colliders

Collisions between electrons and protons are used to study the inner structure of the proton e.g. the quark gluon distribution underneath the valence quarks. The electrons are used as a point like probe to determine the inner structures in the target. This deep inelastic scattering studies were performed in the beginning using an accelerated electron beam colliding on a fixed target. Due to kinematic considerations however a much higher resolution is obtained if two accelerated beams are brought into collision.

Due to the different nature and beam dynamics of the two particles an electron-proton collider cannot be built as a single ring machine: It consists of two storage rings of equal circumference, one being optimised for the acceleration and storage of electrons, the other for a high energy proton beam. The design of these two rings looks quite different and completely different effects determine the performance limitations of the rings. Figure 6.30 shows the two storage rings of the HERA collider:

HERA was built as a 6.3 km long double ring collider with beam energy of 27.5 GeV for the electron beam, and 920 GeV for the proton beam [72]. The fundamental layout was based on four arcs and four straight sections where



Fig. 6.30 View of the two independent storage rings for electron and proton acceleration in HERA. The super conducting proton lattice is placed on top of the conventional electron ring

the high-energy detectors were located. The proton machine was designed as a superconducting magnet lattice in the arcs to achieve the highest possible beam rigidity (or particle energy). The electron storage ring was built in conventional magnet technology: here the limiting factor was the synchrotron radiation emitted by the electrons which was too strong to justify super conducting magnet technology. Basic limits for the achievable beam energy therefore were in the case of the protons the magnetic field of the bending magnets ($B = 5.1$ T) and for the electron ring the available RF power that was needed to compensate the synchrotron radiation losses. Both rings had been built on top of each other to guarantee an equal revolution time of the circulating particle bunches.

The interaction region of such a two ring collider deserves special attention: While the two beams are brought into collision in a common vacuum system and magnet lattice, they have to be separated after the IP and guided into their respective magnet lattices. Especially in the case of the electron beams the separation has to be performed fast enough, as the strong focusing fields of proton mini beta magnets can only be applied after a full separation of the beams.

Two mini beta insertions therefore have to be installed and combined with an effective beam separation scheme. In the case of HERA the separation has been achieved by using the different momenta of the beams: The mini beta quadrupoles of the electron beam have been placed offset with respect to their magnetic axis and acted as combined function magnets. Consequently the electron beam was bent due to its smaller beam rigidity to the inner side of the ring and at a distance $s^* = 20$ m the first proton magnet could be installed. A schematic view of this nested interaction region is shown in Fig. 6.31.

The advantage of this scheme is its compactness as beam separation and focusing are obtained at the same time. Special care however is needed as the electron quadrupoles of the mini beat section will have an effect on the proton beam that depends on the corresponding energy of the electron beam. This dynamic influence

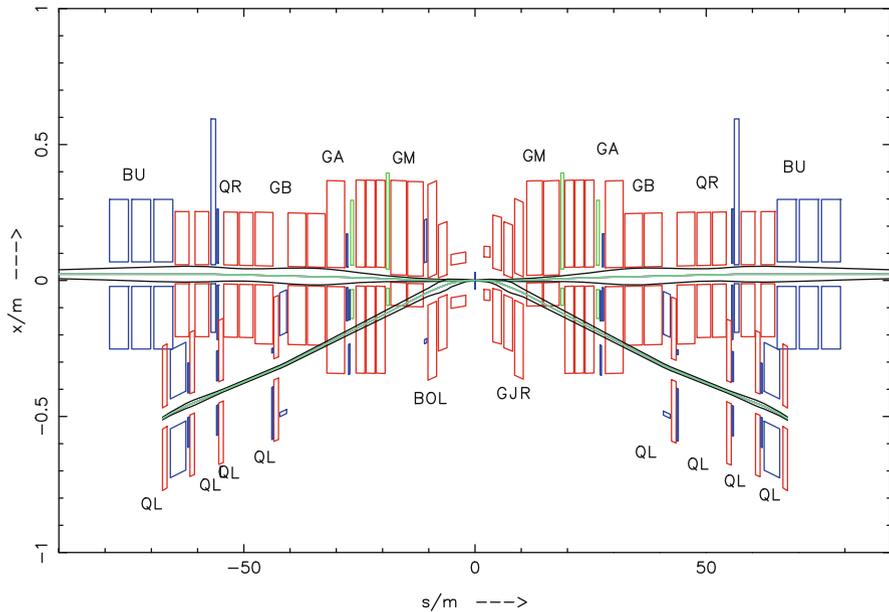


Fig. 6.31 Layout of the HERA interaction region: The inner triplet of the electron lattice is combined with the beam separation scheme and embedded inside the doublet quadrupoles of the proton mini beta insertion

on the optics and orbit of the protons therefore has to be compensated during the acceleration of the electrons as well as during the beta squeeze.

The luminosity formula for such a double ring collider is given by

$$L = \frac{1}{2\pi e^2 f_0} * \frac{\sum_i (I_{pi} * I_{ei})}{\sqrt{\sigma_{xp}^2 + \sigma_{xe}^2} * \sqrt{\sigma_{yp}^2 + \sigma_{ye}^2}}$$

It depends on the product of the single bunch currents I_{pi} and I_{ei} and the sum of this contribution over the overall number i of colliding bunches in the rings. As the beams are guided in different magnet lattices the beam sizes σ are independent of each other. Nevertheless the beams have to be matched, i.e. the beam sizes of the two beams at the IP have to be equal in both planes: $\sigma_{xp} = \sigma_{xe}$ and $\sigma_{yp} = \sigma_{ye}$. This condition deserves special attention as the beam emittances of protons and electrons are quite different and independent beam optics have to be established to achieve matched beam sizes.

Another special feature of an electron proton collider is the synchrotron radiation that is emitted by the electron beam. Usually this effect is present in the arc structure where the dipole fields bend the beam on the design orbit. Due to the separation fields needed in the interaction region the synchrotron radiation is also emitted close to the IP and special care is needed to shield the high energy detector from the

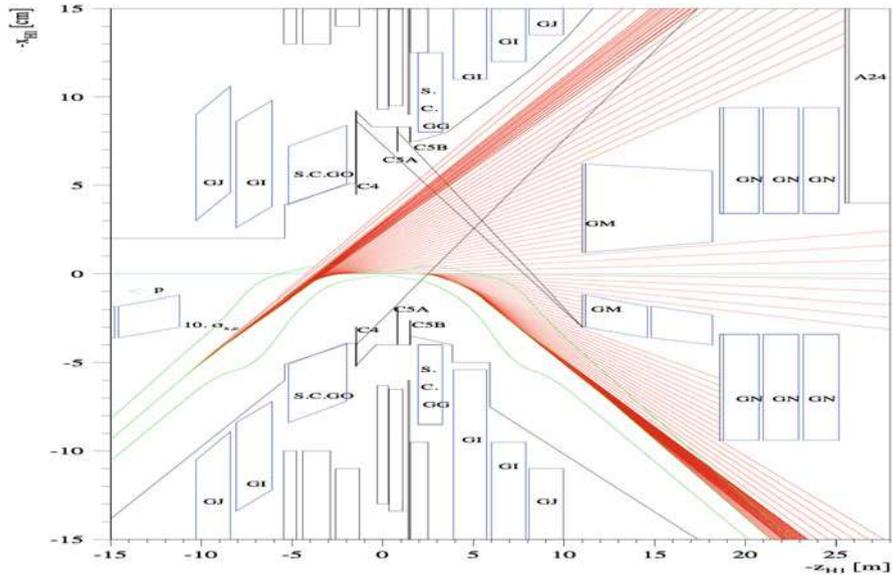


Fig. 6.32 Synchrotron radiation emitted during beam separation in the HERA interaction region. The plot shows schematically the interaction region lattice, the beam dimension and the direction and density of the synchrotron light

emitted photons. For the HERA collider this problem has been studied in detail (see Fig. 6.32) and a combination of absorbers and movable collimation masks have been used to avoid hits from direct or back scattered photons into the detector parts.

The performance limitations of such an e-p-collider are given by the single bunch intensity of the protons (limited by the particle source), the overall current of the electrons (limited by RF power or beam instabilities), the number of bunches that can be stored in the machine (limited by technical reasons of the injection elements) and the usual limits of the mini beta insertions that have been mentioned above.

The main parameters of HERA are summarised in Table 6.1:

Table 6.1 Summary of the main parameters of HERA

	Electrons	Protons
Energy	27.5 GeV	920 GeV
Beam current	58 mA	140 mA
Particles per bunch	4×10^{10}	1×10^{11}
Number of bunches	189	180
Beta function at IP x/y	0.63 m/0.26 m	2.45 m/0.18 m
Hor. emittance	20 nm	5.1 nm
Emittance ratio ϵ_y/ϵ_x	0.18	1.0
Beam size (IP) σ_x/σ_y	112/30 μm	112/30 μm
Luminosity	$7 \times 10^{31} \text{ cm}^{-2} \text{ s}^{-1}$	

6.9 Ion Colliders¹

W. Fischer · J. M. Jowett

Ion colliders are research tools for high-energy nuclear physics. The collisions of fully stripped high-energy ions, that is, atomic nuclei, create matter of a temperature and density that existed in the first microseconds after the Big Bang. The matter created in these high-energy ion collisions is known as the Quark Gluon Plasma (QGP), and interactions between the quarks and gluons is the subject of the theory of quantum chromodynamics (QCD). The basic interactions are studied in simpler collisions such as e^+e^- or pp but heavy-ion collisions allow the study of more complex collective phenomena in QCD. The collisions in ion colliders can create hadronic matter at much higher densities and temperatures than fixed target experiments although at a much lower luminosity.

The collisions of heavy ions in RHIC and the LHC have yielded a number of new results and revealed phenomena that were unexpected on the basis of previous theoretical understanding. The QGP generated in the heavy ion collisions in RHIC was expected to be weakly interacting, but found to be strongly interacting like an almost perfect liquid [73, 74]. Hadronic jets created in the collisions have a rather short mean free path in the QGP leading to a phenomenon termed “jet quenching” [74], and the largest ever measured vorticity was seen in heavy ion collisions [75]. The collisions also created the heaviest artificially made antimatter nuclei, anti-helium-4 [76, 77]. The higher energies in the LHC create many more hard probes and heavy bound states such as charmonium (J/ψ) or bottomonium (Υ) and, in the highest-energy p-Pb collisions, toponium. Z and W bosons, particles that do not interact with the QGP via the strong interaction, were never before seen in heavy ion collisions. The ALICE experiment also reported the highest temperatures directly measured in the laboratory [78].

The colliding nuclei also have high electric charges ($Z \sim 80$). Together with the powerful Lorentz-compression at high energies, these generate enormous electromagnetic fields outside the nuclear radius. As first shown by Fermi, Weizsäcker and Williams, these fields can be represented as a beam of high energy quasi-real photons, leading to so-called ultraperipheral photonuclear and photon-photon collisions. Besides their intrinsic interest, the high cross-sections for these processes have consequences for the operation of the collider. The ATLAS experiment at the LHC has published the first evidence for light-on-light elastic scattering, a long-predicted fundamental process of nonlinear quantum electrodynamics, transcending Maxwell’s equations.

¹This section has been authored by Brookhaven Science Associates, LLC under Contract No. DE-SC0012704 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

The first ion collider was the CERN Intersecting Storage Rings (ISR), which briefly collided light ions [79, 80] in the late 1970s. The BNL Relativistic Heavy Ion Collider (RHIC) has been in operation since 2000 and collided a number of species at numerous energies. The CERN Large Hadron Collider (LHC) started its Run 1 heavy ion program in 2010 and has provided mainly p-p, p-Pb and Pb-Pb at increasing luminosity with a substantial increase in energy in Run 2 (2015–2018). Both RHIC and the LHC have an expected operating time exceeding 20 years. Further upgrades to the LHC, its injector complex and its experiments, foreseen in the shutdown after Run 2, should allow the integrated luminosity in Runs 3 and 4 (up to 2029) to exceed Runs 1 and 2 by an order of magnitude. Table 6.2 shows all species combinations and energy ranges demonstrated to date for the ISR, RHIC and LHC. All three machines also collide protons. In RHIC the protons are spin-polarized, making the machine the only collider of spin-polarized protons ever built. The LHC is the highest energy proton-proton and heavy-ion collider ever built. Critically, proton-proton collisions at the same energy per nucleon provide reference data for heavy ion collisions. In the following, we will limit our comments to the ion operation in RHIC and the LHC.

Ion colliders differ from proton or antiproton colliders in a number of ways: the preparation of the ions in the source and the pre-injector chain is limited by other effects than for protons; frequent changes in the collision energy and particle species, including asymmetric species, are typical; and the interaction of

Table 6.2 Ion species and energies achieved in ISR, RHIC and LHC as of 2017

Machine	Species	Energies [GeV/nucleon]
ISR	α - α	13.3–15.7
	p- α	26.6–31.4 (p), 13.3–15.7 (α)
	d-d	13.3–15.7
	p-d	26.6–31.4 (p), 13.3–15.7 (d)
	p-p	13.5–31.2
RHIC	U-U	96.4
	Au-Au	3.85–100
	Cu-Au	100
	Cu-Cu	11.2–100
	h-Au	103.5 (h)–100 (Au)
	d-Au	9.9–100
	p \uparrow -Au	103.9 (p)–98.6 (Au)
	p \uparrow -Al	103.9 (p)–98.7 (Al)
p \uparrow -p \uparrow	31.2–255	
LHC	Pb-Pb	1380–2511 (2563 briefly)
	Xe-Xe	2721
	p-Pb	4000–6500 (p), 1577–2563 (Pb)
	p-p	3500–6500

p, d, h and α denote the nuclei of the hydrogen, deuterium, helium-3 and helium-4 atoms respectively. All three machines also collide proton beams, which are spin-polarized in RHIC. The quoted energy is the sum of rest and kinetic energy per nucleon

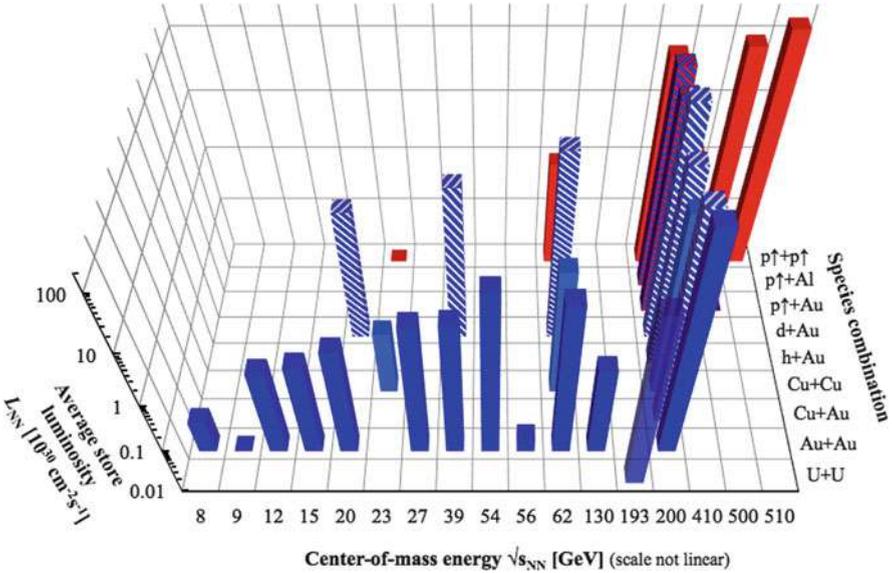


Fig. 6.33 Achieved nucleon-pair luminosity L_{NN} , averaged over a store, for all species combination and energies in RHIC

ions with each other and accelerator components is different from protons. This has implications for collision products, collimation, the beam dump, and intercepting instrumentation devices such as profile monitors. Thus, the performance limitations of heavy-ion colliders are also different from proton-proton colliders.

Figure 6.33 shows the achieved nucleon-pair luminosities L_{NN} , averaged over a store, for all species combinations and energies in RHIC. The plot demonstrates the flexibility of RHIC in colliding different species combinations (all of them at or near the center of mass energy $\sqrt{s_{NN}} = 100 \text{ GeV}$), energy scans for a number of species combinations (Au+Au, Cu+Cu, d+Au p \uparrow +p \uparrow), and a luminosity that is strongly decreasing with the collision energy.

In the preparation for the collider use, the charge state Z of the ions is successively increased. A high charge state Z increases the bending and acceleration efficiency, but also increases the effects of space charge and intrabeam scattering (IBS). The direct space charge tune shift ΔQ , typically limited to values of less than 0.5, is given by [81]

$$\Delta Q = -\frac{\lambda R}{2\varepsilon_n \beta \gamma^2} \frac{r_0 Z^2}{A},$$

where λ is the particle line density, R the machine circumference, ε_n the normalized emittance, β and γ the relativistic factors, r_0 the classical proton radius, and A the

Table 6.3 Preparation of the heavy ions for RHIC and LHC

RHIC (Au)			LHC (Pb)		
	Charge state Z	Ion energy [eV/nucleon]		Charge state Z	Ion energy [eV/nucleon]
LION ^a	1+	150	ECR	27+	2.5 k
EBIS ^b	32+	0.9 M	LINAC3	54+	4.2 M
Booster	77+	101 M	LEIR	54+	72.2 M
AGS	79+	8.8 G	PS	82+	5.9 G
RHIC	79+	99 G	SPS	82+	177 G
			LHC	82+	2.51 T

For each accelerator, the kinetic energy of the ions is given at extraction, and the charge state in the following transfer line

^aLaser Ion Source; ^bElectron Beam Ion Source

mass number. IBS growth rates $1/T_{x,y,s}$ scale like [81]

$$\frac{1}{T_{x,y,s}} \propto \frac{Z^4}{A^2} \frac{N_b}{\gamma \varepsilon_x \varepsilon_y \varepsilon_s}$$

where N_b is the bunch intensity, and $\varepsilon_{x,y,s}$ are the normalized emittances. High charge states also reduce the electron stripping probability, and electron stripping at higher energies is generally more efficient.

Table 6.3 shows the charge states and energies in the RHIC and LHC injector chains for the Au and Pb respectively, the heavy ion species most often used in these machines. For RHIC singly charge ions are generated in a hollow cathode or laser ion source (LION) [82], and transferred into an Electron Beam Ion Source (EBIS) [83]. With EBIS, beams of almost any element can be prepared for RHIC including uranium and spin-polarized ^3He . After increasing the charge state to $Z = +32$ the ions are accelerated through an RFQ and short linac, and injected into the Booster. After acceleration in the Booster, all but two electrons are stripped before injection into the AGS, and the ions are further accelerated. To increase the intensity of the ion bunches, bunches are merged in both the Booster and AGS. The last two electrons are stripped in the transfer line from the AGS to RHIC. In RHIC all ions except protons have to cross the transition energy, when bunches become short and peak currents high. In addition, the longitudinal motion is frozen for a short period, and the short bunches can trigger the creation of an electron cloud [84]. This situation makes the beams vulnerable to instabilities [85], which limited the bunch intensity for a number of years [84].

At CERN an ECR ion source is used, followed by an RFQ and the heavy ion LINAC3 [86, 87]. After passing a carbon foil that strips electrons, the ions are then accumulated in Low Energy Ion Ring (LEIR) [88]. During the 71-turn injection and before acceleration the ions are cooled with an electron beam, with a transverse cooling time of 0.2 s. To minimize dynamic vacuum effects from charge-change

processes the LEIR vacuum system is designed for a dynamic pressure of less than 10^{-12} mbar. From LEIR the ions are injected into the PS where the bunches are split to obtain the bunch spacing needed for the LHC. After acceleration in the PS the last remaining electrons are stripped before injection into the SPS. In the SPS at injection space charge and intrabeam scattering were a concern, and an emittance growth of about 20% is observed at injection. Acceleration in the SPS requires a special fixed frequency acceleration scheme since the main 200 MHz RF system does not have the frequency range required to accelerate heavy ions with a constant harmonic number. The SPS acceleration scheme takes advantage of the fact that the ion bunch train only fills a fraction of the circumference allowing for an adjustment of the RF phase during the time without beam [89].

The luminosity is given by

$$L = (\beta\gamma) \frac{f_{rev}}{4\pi} k_c \frac{N_{b1}N_{b2}}{\varepsilon_n\beta^*} H$$

where f_{rev} is the revolution frequency, k_c the number of bunch-bunch collisions per turn, N_{b1} and N_{b2} the bunch intensities in the two beams respectively, and β^* the lattice envelope function at the interaction point. The factor H accounts for the hourglass effect and crossing angles, and is smaller than and of order 1. The luminosity is limited by different effects in RHIC and the LHC.

In RHIC bunches of fully stripped heavy ions like Au⁷⁹⁺ with the same number of charges as proton bunches have IBS growth rates an order of magnitude larger. In RHIC at injection IBS leads to bunch lengthening, and at store to particle loss out of the RF buckets and an increase in the transverse emittance. Longitudinal and transverse bunched beam stochastic cooling at store has been implemented [90] to counteract IBS. This and an increase in the bunch intensity have significantly increased the average store luminosity (Fig. 6.34). Table 6.4 shows the latest RHIC parameters for Au–Au operation.

Fig. 6.34 RHIC instantaneous Au+Au luminosity in 2007 with longitudinal stochastic cooling in the Yellow ring only, and in 2014 with 3D cooling in both rings. The increase in the initial luminosity is due to an increase in the bunch intensity

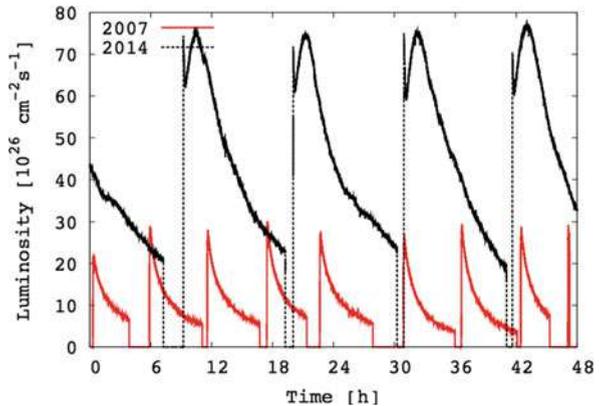
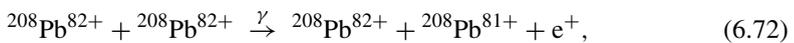


Table 6.4 Main operating parameters achieved for the most commonly heavy ions in RHIC and LHC as of 2017

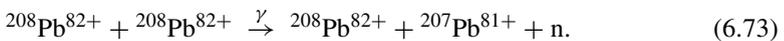
Parameter	Unit	RHIC	LHC
Circumference C	km	3.8	26.7
Ion species		$^{197}\text{Au}^{79+}$	$^{208}\text{Pb}^{82+}$
Maximum energy	GeV/nucleon	100	2511
Bunch intensity	10^9	2.0	0.20
Number of colliding bunches		111	492
Peak luminosity	$10^{26} \text{ cm}^{-2}\text{s}^{-1}$	155	36
Average store luminosity	$10^{26} \text{ cm}^{-2}\text{s}^{-1}$	87	17

Other effects that have limited the heavy ion performance in the past include: the availability of high intensity bunches from the injector chain, instabilities at transition [91] driven by the machine impedance and electron clouds (RHIC is the only superconducting accelerator that crosses the transition energy), dynamic pressure increases including pressure instabilities caused by electron clouds [84], beam loading in the storage RF system (bunches are accelerated with $h = 360$ and transferred into a $h = 7 \times 360$ system at store), and chromatic lattice aberrations at $\beta^* < 70 \text{ cm}$.

The LHC heavy ion operation started in 2010 and the luminosity is principally limited by two effects [92, 93]. Firstly, secondary beams generated in collision and having a Z/A ratio different from the primary beam will be lost in the dispersion suppressor, a location with superconducting magnets with a limited ability to absorb heat [94]. Secondly, the collimation efficiency for ions is lower than for protons leading again to losses in uncontrolled regions [95]. Expected LHC ion parameters are shown in Table 6.4. The two most important processes for the generation of secondary beams in collisions are Bound-Free Pair Production (BFPP),



with a cross section of 281 barn; and Electromagnetic Dissociation (EMD) with a total cross section of 226 barn, about half of which is from the 1-neutron reaction



Beam losses due to BFPP were observed in RHIC with $^{63}\text{Cu}^{29+}$ ions [96] and effective mitigation measures have now been implemented at the LHC [97, 98]. These have allowed Pb-Pb luminosities far beyond the design value from 2015 onwards. Collimation of heavy ions is fundamentally different from protons. Protons are scattered at a primary collimator and collected at a secondary collimator. Heavy ions undergo nuclear fragmentation and electromagnetic dissociation in the primary collimator. The fragments created have a wide range of Z/A ratios that are not collected by the secondary collimators. Measurements of collimation efficiency

were done in the CERN SPS and compared with detailed simulations to obtain reliable estimates of the heavy ion collimation efficiency in the LHC [95].

The LHC also collided Xe nuclei in 2017 [98] and may collide other species in future, generally with a view to increasing the nucleon-nucleon luminosity.

In the collision of asymmetric species the 2-in-1 magnet design of the LHC requires that the magnetic fields in two rings are the same (the two RHIC rings are independent and can have different fields). For p–Pb operation it is then necessary to have different revolution frequencies at injection and during the energy ramp [99]. Lead beams in the LHC at design energy have noticeable synchrotron radiation damping times (6 h and 13 h longitudinally and transversally) that are of the same order as the IBS emittance growth times (8 h and 13 h longitudinally and transversally) [92].

6.10 Beam Cooling

F. Caspers · D. Möhl

6.10.1 Introduction

Beam cooling aims at reducing the size and the energy spread of a particle beam circulating in a storage ring or in an ion trap. This reduction of size should not be accompanied by beam loss; the goal is to increase the particle density [100]. Since the beam size varies with the focusing properties of the storage ring, it is useful to introduce normalized measures of size and density. Such quantities are the (horizontal, vertical and longitudinal) emittances and the phase-space density. For our present purpose they may be regarded as the (squares of the) horizontal and vertical beam diameters, the energy spread, and the density, normalized by the focusing strength and the size of the ring to make them independent of the storage ring properties. Phase-space density is then a general figure of merit of a particle beam, and cooling improves this figure of merit. The terms beam temperature and beam cooling have been taken over from the kinetic theory of gases. For visualization one may imagine a beam of particles going around in a storage ring. Particles will oscillate around the beam centre in much the same way that particles of a hot gas bounce back and forth between the walls of a container. The larger the mean square of the velocity of these oscillations in a beam, the larger the beam size. The mean square velocity spread is used to define the beam temperature in analogy to the temperature of the gas which is determined by the kinetic energy of the molecules.

There are several basic motivations for the application and development of different beam cooling techniques:

- Collection and accumulation of rare particles, e.g. antiprotons or short lived particles such as muons.
- Improvement of interaction rate and resolution, e.g. collision experiments with antiprotons or with ions; increase in luminosity. For fixed target experiments: sharply collimated and/or highly mono-energetic beams for precision experiments.
- Preservation of beam quality, mitigation and suppression of beam blow-up.
- Preparation of crystalline beams.

Several cooling techniques are operational or have been discussed:

- Radiation cooling (often referred to as radiation damping); linked to energy loss of particles via synchrotron radiation (used in virtually all modern electron synchrotrons).
- Stochastic cooling (works well for “hot” beams to get them “tempered”).
- Electron cooling (most suitable for “tempered” beams to get them “cold”).
- Laser cooling (essentially for ions where two level transitions of electrons can be excited).
- Ionization- and friction-cooling (mainly discussed in the context of muon cooling).
- Resistive cooling; used to cool charged particles in a trap where the kinetic energy of the particle is dissipated in the resistive losses of a resonant circuit.
- Coherent electron cooling, a kind of blend from stochastic cooling at very high frequencies and electron cooling (under development at BNL these days)

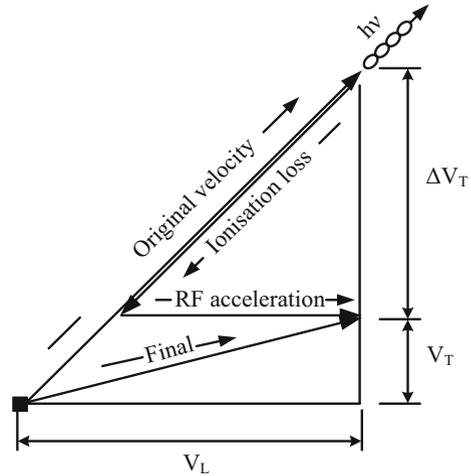
The use of the terms cooling and damping is not always well distinguished and unambiguous in the literature. Even in the context of stochastic cooling the authors were using the term damping in the early days. A similar observation can be made for radiation damping and cooling. One may consider defining any action on individual particles as “cooling” and any action on groups of particles as damping. Examples are the feedback systems in circular machines which are commonly referred to as dampers and which prevent emittance blow up, while a very similar feedback system just having a much higher electronic gain can work as (stochastic) cooler and reduce the emittance. However typically such damper systems have a much smaller bandwidth and lower operation frequency as compared to stochastic cooling hardware.

6.10.2 Beam Cooling Techniques

6.10.2.1 Radiation Cooling

Back in 1956, A.A. Kolomenski and A.N. Lebedev [101] pointed out that the ‘synchrotron light’ emitted by an electron moving on a curved orbit can have a

Fig. 6.35 The principle of transverse cooling by synchrotron radiation (transverse velocities exaggerated)



damping effect on the motion of the particle. This is because the radiation is sharply peaked in the forward direction. The continuous emission of synchrotron radiation leads to a friction force opposite to the direction of the motion. For a particle moving on the design orbit, the energy loss is restored and the friction force is on average compensated by the RF-system. For a real particle the residual friction force tends to damp the deviation from the design orbit (Fig. 6.35). This cooling force is counteracted by the ‘radiation excitation’: synchrotron light is really emitted in discrete quanta and these many small kicks tend to heat the particle. The final emittances result from the equilibrium of radiation damping and excitation. We will see that a similar interplay between a specific cooling and heating mechanism is characteristic also for the other cooling methods.

The theory of cooling by synchrotron radiation is in a mature state. Following up on Sands’ classical treatment on “the physics of electron storage rings”, radiation cooling has found its place in text books. The immense success of modern electron–positron machines, both ‘synchrotron light facilities’ (e.g. ESRF, ALS, APS, BESSY, SPRING8) and colliders (e.g. LEP, PEP II, KEKB) would not have been possible without the full understanding of radiation effects. Virtually all these machines depend critically on radiation cooling to attain the minute emittances necessary in their application. Linear e^+e^- -collider schemes (like CLIC, TESLA, NLC, JLC) too, have to rely on ‘damping rings’ in their injector chain to produce the ultra-high phase-space density required. For historical reasons the reduction of beam emittance due to the emission of synchrotron radiation (typically from leptons) is usually referred to as radiation damping, although the term “cooling” might be more consistent.

The cooling rates as well as the final beam size and momentum spread depend on the lattice functions in regions where the orbit is curved. The art is then to ‘arrange’ these functions such that the desired beam property results. The strategy for ‘low emittance lattices’ is well developed and ‘third-generation machines’ providing

beams of extremely high brightness have come into operation. To enhance the cooling, wiggler magnets are used, producing a succession of left and right bends. This increases the radiation and thereby the damping rates. The heating can be kept small by placing the wiggler at locations where the focusing functions of the ring are appropriate to make the particle motion insensitive to kicks. More details for cooling by synchrotron radiation are given in Sect. 6.5.

Radiation cooling and lattice properties of the storage ring are thus intimately linked and by smart design, orders of magnitude in the equilibrium emittances have been gained. This may serve as example for other cooling techniques for which the art of ‘low emittance lattices’ is only now emerging.

6.10.2.2 Microwave Stochastic Cooling

For (anti-)protons and heavier ions, radiation damping is almost negligible at the energies currently accessible in accelerators except for the LHC. One of the ‘artificial’ cooling methods devised for these heavy particles is stochastic cooling by a broadband feedback system (Fig. 6.36). The name “stochastic damping” was coined by Simon van der Meer who invented this method in 1968 (first published in 1972) [102] to underline the statistical basis of the method. First successful tests and observations were done at the CERN ISR (Intersecting Storage Rings) [102] followed by a dedicated “Initial Cooling Experiment” ICE [103]. In 1984 Simon van der Meer shared the Nobel Prize [104] in physics with Carlo Rubbia for his contribution to the observation of the intermediate vector boson. Microwave stochastic cooling was considered a key ingredient for reaching sufficient phase space density of the precious and rare antiprotons to produce a small number of W- and Z-Bosons in the CERN Super Proton-Antiproton Synchrotron (SppS) experiment in 1982. At the core of stochastic cooling is the observation, that the phase-space density can be increased by a system that acts to reduce the deviation of small sections, called samples, of the beam. By measuring and correcting the statistical fluctuations (‘Schottky noise’) of the sample averages, the *spreads* in the corresponding beam properties are gradually reduced. Stochastic cooling may thus be viewed as a ‘sampling procedure’ where samples are continuously taken from the beam and the average of each sample is corrected. The basic principle of (transverse) stochastic cooling is sketched in Fig. 6.36.

A somewhat different picture is based on the behavior of a test particle. At each passage it receives its own ‘coherent’ kick plus the ‘incoherent’ random kicks due to all other sample members. The sample length T_s (response-time) is given by the bandwidth W of the system through $T_s \approx 1/2W$ and the number N_s of particles per sample is proportional to T_s . Hence a large bandwidth is important to work with small samples. Present day cooling rings have a revolution time between a fraction of a μs (e.g. CERN AD) up to about 20 μs (Relativistic Heavy Ion Collider (RHIC), Fermilab bunched beam cooling systems). The sample length T_s amounts usually to less than 1 ns which corresponds to a cooling system bandwidth of 500 MHz in this case assuming the generalized Nyquist criterion for band-limited signals under

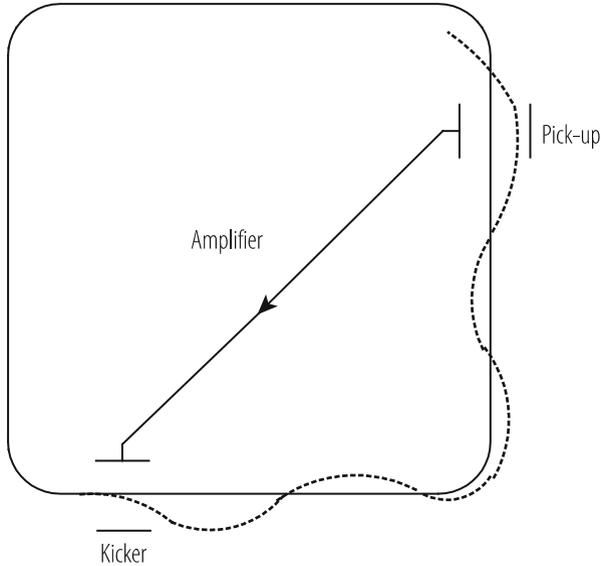


Fig. 6.36 The basic set up for (horizontal) stochastic cooling

ideal assumptions. Thus each sample contains only a small fraction of the total beam population circulating in the machine. Another important ingredient is ‘mixing’, i.e. the renewal of the sample population due to the spread of the particle revolution frequencies.

Based on the ‘sampling’ and/or the ‘test particle picture’ one derives in a few steps [105] a simplified relation for the cooling rate $1/\tau$ of the transverse emittance ε with ($1/\tau = (1/\varepsilon)d\varepsilon/dt$) or for longitudinal phase space the momentum deviation ($1/\tau = (1/\Delta p)d\Delta p/dt$):

$$\frac{1}{\tau} = \frac{W}{N} \left[\underset{\text{<coherent effect>}}{2g \left(1 - \tilde{M}^{-2} \right)} - \underset{\text{<incoherent effect>}}{g^2 \left(M + U/Z^2 \right)} \right]. \tag{6.74}$$

The parameters appearing in Eq. (6.74) have the following significance:

N	number of particles in the coasting beam	
W	cooling system bandwidth	
g	gain parameter (fraction of sample error corrected per turn)	$(g < 1)$
M	desired mixing factor (mixing on the way kicker–pick-up = good mixing)	$(M > 1)$
\tilde{M}	undesired mixing factor (slippage on the way pick-up–kicker = bad mixing)	$(\tilde{M} > 1)$
U	noise to signal power ratio (for single charged particles)	$(U > 0)$
Z	charge number of beam particles (\leq atomic number of the ion!)	$(Z \geq 1)$

There is an optimum value of g for which Eq. (6.74) has a maximum. As to the other parameters, N and Z are properties of the beam, W is a property of the cooling system and M , \tilde{M} and U depend on the interplay of cooling system-, beam- and storage ring characteristics. The term in the bracket can at best be 1 but is more like 1/10 to 1/100 in real systems, depending on how well the mixing and noise problems are solved. The ideal cooling rate W/N can be interpreted as the maximum rate at which information on single particles can be acquired. Note that the gain parameter g (fractional sample error correction) should not be confounded with the electronic gain of cooling system which is typically 120 db or 12 orders of magnitude in power.

Lattice parameters are especially important for the achievement of ‘good’ values of M , \tilde{M} and U , maximising the bracket in Eq. (6.74). In addition to the struggle for large bandwidth, the advance in stochastic cooling is intimately linked to progress in dealing with the noise and mixing factors. In summary it can be said that present-day systems are working with a bandwidth of around 1 GHz for an individual cooling system with the possibility of extensions up to nearly 10 GHz by using several cooling bands in the same ring. Limitations on W are discussed in [106].

Turning to the mixing dilemma discussed at length in [107], we note that stochastic cooling only works if after each correction the samples (at least partly) re-randomise (desired mixing), and at the same time a particle on its way from pick-up to kicker does not slip too much with respect to its own signal (undesired mixing). The mixing rates $1/M$ and $1/\tilde{M}$ are related to the fraction of the sample length by which a particle with the typical momentum deviation slips with respect to the nominal particle. Here M refers to the way from kicker to pick-up (‘K to P’), and \tilde{M} to the way pick-up to kicker (‘P to K’). Both depend on the flight-time dispersion which in turn is given by the local ‘off-momentum factors’,

$$\eta_{kp} = \left(\frac{dT}{T} / \frac{dp}{p} \right)_{kp}, \quad (6.75)$$

and the similar quantity η_{pk} respectively. For a regular lattice the beam paths ‘K to P’ and ‘P to K’ consist of a number of identical cells and one has

$$\eta_{kp} \approx \eta_{pk} \approx \eta = \left| \gamma_{tr}^{-2} - \gamma^{-2} \right|, \quad (6.76)$$

i.e. the local η -factors are close to the off-momentum factor for the whole ring. In this situation the ratio \tilde{M}/M is simply given by the corresponding path lengths (T_{pk} and T_{kp}). Then, e.g. in the case of the CERN AD (antiproton decelerator) where the cooling loop cuts diagonally across the ring, one has $\tilde{M} \approx M$ instead of the desired $\tilde{M} \gg 1$, $M = 1$. The usual compromise is to accept imperfect mixing, letting both \tilde{M} and M be in the range of 3–5, say. The price to pay is a slower cooling rate, for

example $1/\tau \leq 0.28W/N$ in the case of $\tilde{M} = M$ instead of $1/\tau \leq W/N$ for perfect mixing.

‘Optimum mixing lattices’ (also referred to as ‘split ring designs’) have been proposed for the 10 GeV ‘SuperLEAR’ ring [108] (which was, however, never built). The idea is to make the path P to K isochronous ($\eta_{pk} = 0$) and the path K to P strongly flight-time dispersive ($\eta_{kp} \gg 0$). These lattice properties have to be reconciled with the many other requirements of the storage ring. The next generation of stochastic cooling rings will use such split rings lattices. They were discussed for RIKEN in Japan [109] and are under construction for GSI and FAIR in Germany [110, 111]. It should be mentioned that the condition $\eta_{pk} = 0, \eta_{kp} \gg 0$ can increase the cooling rate for transverse and for longitudinal ‘Palmer-Hereward’ cooling where the transverse displacement concurrent with the betatron amplitude and the momentum error of the particles is used. For momentum cooling by the filter (‘Thorndahl’) method, the split ring design brings less improvement since here the time of flight over a full revolution is used as a measure of momentum. A storage ring with $\eta_{pk} = 0$ and $\eta \approx 1\text{--}2\%$ is under construction for GSI and FAIR in Germany, meeting best conditions for both transverse cooling and filter momentum cooling of antiprotons [112].

Regarding the situation at GSI it should be mentioned that a first successful experiment was performed at the ESR to measure the nuclear radius of the radioactive nucleus ^{56}Ni . To this purpose stochastic precooling and subsequent electron cooling were used in order to accumulate enough intensity for a sufficient S/N in a scattering experiment with an internal hydrogen target [113].

This is not the end of the mixing dilemma: during momentum cooling, as $\Delta p/p$ decreases, the M-factors increase (c.f. Eq. 6.75) and the mixing situation tends to degrade. One can in principle stay close to the optimum by changing η (‘dynamic transition tuning’) as cooling proceeds. Similar considerations hold for machines with variable working energy where, through a change of η , good mixing can be maintained. Again these improvements might be incorporated in the next generation of cooling rings (e.g. at FAIR [112]).

As for the noise, from Eq. (6.74) it is clear that a balanced design aims at $U/Z^2 \ll M$. The noise to signal (power-)ratio depends on the technology of the pre-amplifier and other ‘low level components’ on the one hand and on the sensitivity of the pick-up device on the other hand. There has been great progress in the design of the pick-up and kicker structures and the other components of the cooling loop. These components developed in different labs (e.g. BNL [114], CERN [107], Fermilab [115], Forschungszentrum Jülich (FZJ) [116], GSI [112, 117]) are in fact formidable ‘high-fidelity (HiFi) systems’ with an unprecedented combination of high sensitivity, low noise, great bandwidth, large amplification, very linear phase response, and excellent compatibility with the ultra-high vacuum of the storage ring.

A more detailed discussion of stochastic cooling hardware progress over the last 30 years can be found in [118]. Regarding pick-up and kicker structures we have seen printed versions arriving in the late 1980s of the classical $\lambda/4$ strip-line couplers which are referred to as printed loop or printed slotline couplers which are normally

used as “phased arrays” [119, 120]. As for travelling wave structures starting from the TEM type slotted line version of Faltin [121] McGinnis developed a related device [122] not based on a TEM line, but essentially a waveguide directional coupler with slots masks for the coupling. Those waveguide type slot array couplers have the advantage (in contrast to the Faltin version) that they can operate efficiently also for highly relativistic beams and they exhibit a very high longitudinal and transverse sensitivity over a bandwidth of several 100 MHz in the GHz region. As a particular development the kicker structure for the BNL RHIC bunched beam stochastic cooling system [123–125] is worth mentioning. It consists of an array of cavities which are cut in length and can be opened by a mechanical plunging mechanism in order to let the injected beam pass without aperture limitations.

Another travelling wave structure is the perforated structure which was originally proposed in 2011 [126] and later developed for HIRFL-CSRe stochastic cooling. A large number of small slots in the electrode provides distributed inductive loading, slowing down the phase velocity of the travelling wave structure for the low beta beams. This device is very broadband and operates from low frequencies onwards as a forward coupler. Even for 2.76 m long electrodes used in HIFRL-CSRe, it can be used from a few MHz to 1.2 GHz [127].

Another very promising recent development for pick-ups and kickers are “slot ring” structures [128]. These structures were originally developed for the High Energy Storage Ring (HESR) of the FAIR project at GSI, Germany and successfully tested at the Nuclotron (JINR, Russia) for longitudinal cooling and at COSY (FZJ) for longitudinal and transverse cooling. Slot ring couplers have a fixed aperture and can be used for all three cooling planes simultaneously [128].

In CERN’s anti-proton decelerator AD stochastic cooling is employed at 3.57 GeV/c and 2 GeV/c in both transverse planes and for the longitudinal plane (filter cooling) [129]. The current system uses a set of two kickers and pick-ups, each combining one transverse plane and the longitudinal cooling, with a total of 4.8 kW installed power. It is undergoing a consolidation and upgrade [130] which is including a notch filter with optical delay lines. Cooling times of 15–20 s reduce transverse emittances to 3–4 π mm rad and Dp/p to $\pm 0.3 \times 10^{-3}$ at 3.57 GeV/c and to $\pm 0.08 \times 10^{-3}$ at 2 GeV/c at intensities of 5×10^7 antiprotons. The system uses a bandwidth of one octave between 850 MHz and 1.7 GHz. This is the actual status in early 2019.

In parallel, CLASS A solid state amplifiers [131] (kicker driver) gradually took over from TWT (travelling wave tube) units, although TWTs are still in operation for stochastic cooling e.g. at Fermilab where they work reliably. Notch filters, required for Thorndahl type longitudinal cooling (filter cooling) are implemented since about 1990 with good success in optical fibre technology [125, 132].

Optical signal transmission across the ring (Fermilab de-buncher) has been realized with a laser beam in an evacuated metal pipe (no signal fluctuation from temperature effects of air and humidity on the laser beam). The driving force to select this method of signal transmission was the very tight requirement in terms of transmission delay and delay stability. Anything slower than speed of light would not have permitted timely arrival of the correction signal at the kicker. In 2017 very

fast (around 99% speed of light) hollow optical fibres were applied successfully for analog and wideband signal transmission across the ring at COSY (FZ-Jülich, Germany) [133].

Front end amplifiers showed slow but steady progress and these days we can easily get an uncooled 1–2 GHz or 2–4 GHz device with a noise temperature of 30 K.

Examples of remarkable recent progress in the field of microwave stochastic cooling are

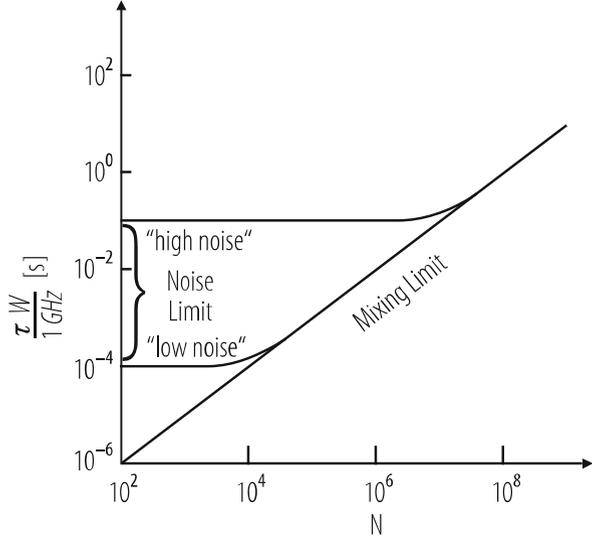
- bunched beam stochastic cooling at RHIC and Fermilab [134–137],
- the impressive improvements of the performance and the interplay of all stochastic cooling and stacking systems at Fermilab together with elaborate beam handling methods such as “slip stacking” [135, 136].

It should be noted that there have been unsuccessful attempts to get bunched beam stochastic cooling operational in large machines despite the fact that one of the first evidence on stochastic cooling at all, in ICE [137] already worked with a bunched beam. However the bunch length was very large. Attempts which failed were in the frame of the SPS p-pbar program at CERN [124] and later (around 1990) also in the Tevatron [138, 139]. Bunched beam cooling is of course hampered by the higher particle density in the bunch. In fact in Eq. (6.74) the number N for the coasting beam has to be replaced by $N_b/B_f = N_b \cdot 1.4 \cdot 2\pi R/l_b$ (with R radius and l_b length of bunch) for a rough estimate [137, 140]. In addition to those expected effects the direct (coherent) bunch signal (proportional to N at low frequencies) tends to mask the very weak Schottky signals required for cooling [137, 140]. This is one of the reasons to place the cooling bands towards high frequencies. In addition a subtle but important difficulty is related to the presence of unexpected and rather strong coherent signals in the bunched beam spectrum which lead to saturation of the front end amplifiers via intermodulation [139, 140]. A theoretical treatment of these persisting “turbulence islands” in the bunch was given by Blaskiewicz [141]. Just in the recent years this problem became mastered at BNL (gold ions) [135] and also in the Fermilab recycler [140]. However, bunched beam stochastic cooling has always been working reasonably well in small machine like the CERN AC [107], LEAR [142], Fermilab de-buncher and accumulator [143] and others since the relative intensity of those coherent signals was less violent compared to large machines. However for the small machines there was little interest in bunched beam stochastic cooling.

New applications of stochastic cooling may include:

- fast cooling and stacking of low intensity radioactive ion beams with cooling times of 100 ms or less as discussed for RIKEN [109] and under construction at GSI [110],
- fast optical stochastic cooling [144–146] (e.g. of intense muon beams but also for bunched beam cooling in large rings) for which a bandwidth of 10^{12} – 10^{13} Hz and a new pick-up, kicker and amplifier technology, and new lattice designs have been contemplated.

Fig. 6.37 Sketch of cooling time vs. intensity (for the ‘mixing limit’ $\tau = 10 N/W$ is taken in the figure)



Let us have a quick look at these developments. The challenge of fast low-intensity cooling can be discussed with reference to Fig. 6.37 [108], which illustrates the optimum cooling time vs. intensity N . For large N the cooling time increases linearly with N with the slope $1/W \left[M / (1 - \tilde{M}^{-2})^2 \right]$. This is the mixing and bandwidth limit. For small N , cooling time levels off to a constant ‘noise limited’ value reached for $U/Z^2 \gg M$ (note that $U \propto 1/N$).

The art is to shift the levelling off to small N by improving the signal to noise ratio. Theoretically, short cooling times are then possible (e.g. 10 ms for $N = 10^5$ Sn⁵⁰⁺ ions and a few 100 MHz bandwidth as discussed for RIKEN). However, other difficulties like the broadband power needed for such a rapid emittance decrease, and the residual RF-structure after debunching may pose new problems for fast cooling and stacking.

Optical stochastic cooling (OSC) proposed by Mikhailchenko, Zholents and Zolotarev [144, 145] in 1993 is an extension of certain basic concepts of microwave stochastic cooling into the optical frequency range using different pickup and kicker mechanism and structures. It is a potentially very promising technique but has not been tested in practice so far. Challenges may be amongst other items the stability and linearity of the optical signal transmission chain as well as of the circulating hadron beam. Maybe we shall soon see important steps towards this technology at BNL in a forthcoming “coherent electron cooling experiment”.

At BNL stochastic cooling has been implemented at top energy in the RHIC [147]. The bunch cores have full length 5 ns and are spaced by 100 ns. The root mean square Schottky voltage is typically 10% of the coherent voltage generated by the average bunch shape and multi-kilovolt kicker voltages are required for optimal longitudinal cooling. Several novel technologies were required to meet the

challenges. The kicker voltage was obtained by periodically extending the pickup signal, passing it through narrow band filters spaced by 200 MHz (1/5 ns) and driving individual cavities. Taming coherent lines while meeting timing requirements was a serious challenge [148]. When all was said and done the cooling system increased the integrated luminosity of uranium-uranium collisions by a factor of five and typically doubled the gold-gold luminosity.

At Fermilab, OSC is being pursued at the IOTA facility [149]. In OSC a particle emits electromagnetic radiation in the first (pickup) wiggler. Then, the radiation amplified in an optical amplifier (OA) makes a longitudinal kick to the same particle in the second (kicker). A magnetic chicane is used to make space for the OA and to delay a particle so that to compensate for a delay of its radiation in the OA resulting in simultaneous arrival of the particle and its amplified radiation to the kicker wiggler. The chosen optical wavelength is 800 nm, resulting in bandwidths approaching 10^{14} Hz. In the proposed test, the use of 100-MeV ($\gamma = 200$) electrons instead of protons greatly reduces the cost of the experiment but does not limit its generality and applicability to hadron colliders. Conceptual design of the system is complete, with engineering design of the wiggler and optical hardware underway.

Already in late 1970s the need for “stochastic stacking” has been realized [150]. In the “old” CERN AA (antiproton accumulator) [151] early stacking methods were tested and applied in routine operation. In the CERN AAC (antiproton accumulator complex) [152] the antiprotons (pbar or p) coming from the AC (collector ring) were transferred to the inner ring (AA = accumulator). There dedicated stack tail and stack core systems took over the antiprotons after they have passed a pre-cooling system in the AA and were transferred to another orbit by means of RF manipulations. At Fermilab [138] stacking is done in the accumulator ring and later also in the recycler. For the future stacking with stochastic cooling is planned in the frame of the FAIR project [110]. Stochastic stacking of rare radioactive ions has been considered during the planning phase of RIKEN [110] upgrades between 1900 and about 2000 and for FAIR [110].

At Fermilab huge progress has been made since the year 2000 [153], this includes stacking with stochastic cooling was done in three separate machines. The debuncher [154], which accepted $\sim 1.5 \times 10^8$ antiprotons every 2.1 s, used the McGinnis waveguide directional couplers in eight bands over the frequency range 4–8 GHz for a factor of 10 reduction in longitudinal and transverse size. A key piece was the implementation of ramping the amplifier gain down during the cycle, to counter act noise to signal for the momentum bands in the notch filters. A 6 dB decrease in gain resulted in a 12% decrease in the 95% momentum width after 2 s of cooling. The Accumulator [154] accepted the same $\sim 1.5 \times 10^8$ antiprotons and used the Palmer method to build a ‘stack’. Peak performance reached 2.6×10^{11} antiprotons in an hour, with regular transfers to the Recycler to mitigate the known decrease in performance with larger stacks. The Recycler, using a combination of stochastic and electron cooling [155], reached intensities of greater than 4×10^{12} regularly, with peak intensity of 6.1×10^{12} and delivering over 4×10^{13} per week to the collider program.

6.10.2.3 Electron Cooling

The concept of cooling a “hot” beam of ions by mixing it over a short distance in a circular machine with a cold electron beam had been developed by Budker [156] in 1966. It was first tested in 1974 with 68 MeV protons at the NAP-M storage ring at in Novosibirsk. The notions of ‘beam temperature’ and ‘beam cooling’ were introduced and become lucid in the context of electron cooling, which is readily viewed as temperature relaxation in the mixture of a hot ion beam with a co-moving cold electron ‘fluid’. The equilibrium emittances, obtainable when other ‘heating mechanisms’ are negligible, can easily be estimated from this analogy, assuming equalisation of the temperatures ($(M\Delta v^2)_{\text{ion}} \rightarrow (m\Delta v^2)_{\text{electron}}$). For a simple estimate of the cooling time, another resemblance, namely the analogy with slowing down of swift particles in matter, can be helpful. A nice presentation of this subject is given in Jackson’s book [157]: the energy loss in matter is due to the interaction with the shell electrons and in first approximation these electrons are regarded as free rather than bound. Results for this case can be directly applied to the ‘stopping of the heavy particles in the co-moving electron plasma’. The calculations are performed assuming ‘binary collisions’ involving only one ion and one electron at a time.

Using this approximation the cooling time can be written as

$$\frac{1}{\tau} \approx \frac{1}{k} \frac{q^2}{A} \eta_c L_C r_e r_p \frac{j}{e} \frac{1}{\beta^4 \gamma^5 \theta^3}, \quad (6.77)$$

where

$k = 0.6$: for a Gaussian distribution (not realistic),

$k = 0.16$: for a flattened distribution,

q : ion charge number,

A : ion mass number,

η_c : length of cooling section/circumference,

$L_C \approx 10$: Coulomb logarithm (log of max/min impact parameter),

$r_e \approx 2.8 \times 10^{-13}$ cm: classical electron radius,

$r_p \approx 1.5 \times 10^{-16}$ cm: classical proton radius,

j (A/cm²): electron beam current density,

$e \approx 1.6 \times 10^{-19}$ C: elementary charge

$\theta = (\theta_e^2 + \theta_i^2)^{1/2} = \left(\frac{T_e}{m_e c^2} + \frac{T_i}{m_i c^2} \right)$: r.m.s. angle between electron and ion beams,

β, γ : relativistic factors.

The cooling rate ($1/\tau$) thus obtained exhibits the dependence on the main beam and storage ring parameters [158]. Notable is the dependence on both the electron and the ion (both longitudinal and transverse) velocity spreads: $\tau \propto \theta^3 \propto \left(|\Delta \mathbf{v}_{e_{\text{rms}}}|^3 + |\Delta \mathbf{v}_{i_{\text{rms}}}|^3 \right)$. This indicates an ‘ion spread dominated regime’, where cooling gets faster as the ions cool down until it saturates for $|\Delta \mathbf{v}_{i_{\text{rms}}}| < |\Delta \mathbf{v}_{e_{\text{rms}}}|$ (‘electron dominated regime’). Remarkable also is the strong energy dependence

predicted in this model: $\tau \propto \beta^4 \gamma^5$, with all other parameters (including the electron current density j) kept constant [159].

Neglected in the simple theory are the ‘flattened distribution’, the ‘magnetisation’ and the ‘electron space-charge’ effects, all three (also) discovered and explained at Novosibirsk [159, 160]. In essence the flattened distribution effect takes into account that (due to the acceleration) the electron velocity spread is not isotropic but contracted (by $[E_{\text{cathode}}/E_{\text{final}}]^{1/2}$) in the longitudinal direction. The magnetisation effect is due to the spiraling (Larmor-) motion of the electrons in the magnetic field of the solenoid that is used to guide the electron beam. Then for electron-ion encounters with long ‘collision times’ (impact parameter \gg Larmor radius), the transverse electron velocity spread averages to zero. Finally the electron space-charge induces a potential that leads to a parabolic velocity profile $v(r)$ over the beam whereas the ions exhibit a linear dependence $v(x)$ and $v(y)$ given by the storage ring lattice. Hence the difficulty arises to match the ion and electron velocities. Flattening and magnetisation can have a beneficial outcome, whereas space-charge has a hampering influence on the cooling process. All three effects complicate the theory, spoil the hope for simple analytical formulae and obscure the comparison between measurements at different machines, and even different situations at the same cooler. As an example the cooling assembly used in the low energy antiproton ring (LEAR) is sketched in Fig. 6.38.

The electrons are produced in a gun and directed into the cooling region where they overlap the ion beam over a length 1 m. At the end of the cooling section the electrons are steered away from the ions into a collector where their energy is recuperated. On their whole way from the cathode of the gun to the collector the electrons are usually immersed in a longitudinal magnetic guiding field. This field is constant over the full length or stronger in the gun region. In the latter case the transverse electron temperature in the overlap region decreases (due to “magnetic expansion”) at the expense of the longitudinal temperature. This can reduce the cooling time in situations where the electron temperature dominates.

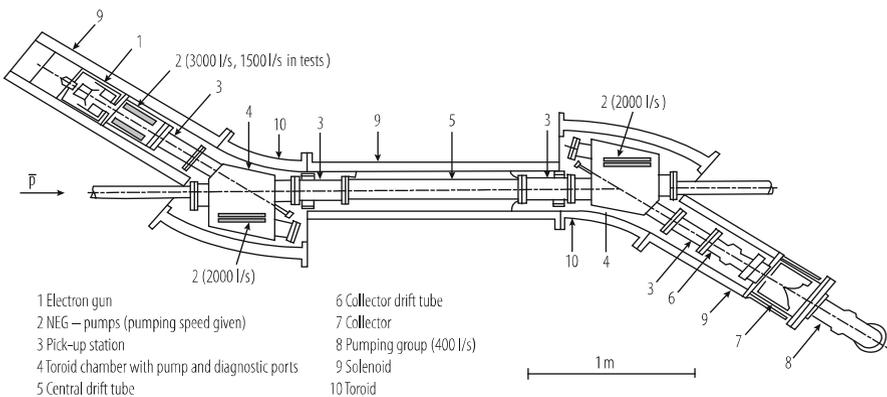


Fig. 6.38 An electron cooling assembly (LEAR electron cooler) from [158]

An important technical problem is electron beam power consumption. To reduce direct losses of the beam power the recuperation (recovering) method is used. It assumes biasing of the collector to negative potential slightly above the cathode potential. Then the power consumption is defined mainly by product of the beam current by the difference of the collector and gun potentials.

In the two toroidal sections, adjacent to the overlap region, the solenoid to create the longitudinal field is curved to guide the electron beam parallel to the ions at the entrance and away from them at the exit. Also in the toroidal regions the solenoid has a larger diameter to permit the penetration of the ion beam.

Many papers deal with the ‘exact and general theory’ [161] and computer programs like BETACOOOL [162] try to include all the subtle effects. Numerous also are the experimental results from 11 (or so) present and past cooling rings. It is not easy to compare the data from different experiments because the cooling in each plane depends in a complicated way on the emittances in all three directions both of the ion and the electron beam. Moreover different quantities are used to measure/define ‘cooling strength’ (examples: cooling of a large injected beam, response of a cold beam to a ‘kick’ or to a transverse or an energy displacement, equilibrium with heating by noise).

In the context of the accumulation of lead ions for the future Large Hadron Collider (LHC) [163], a program of experiments [164] was performed at the LEAR ring to determine optimum lattice functions [165]. Results indicate rather small optimum betatron functions (3–5 m instead of the expected 10 m) and large dispersion ($D = 2\text{--}3$ m instead of the expected 0–1 m). The dependence on dispersion is not fully reproduced by simple analytical formulae. There are other old questions: e.g. the (dis)advantage of magnetic expansion, the dependence of the cooling time on the charge of the ion, the (dis)advantage of neutralising the electron beam, the enigma of the stability of the cooled beam [166], the puzzle of the anomalously fast recombination of certain ions with cooling electrons [167], the (dis)advantage of a hollow electron beam [168].

Considerations so far concern electron cooling at ‘low energies’ ($T_e = 2\text{--}300$ keV) where cooling rings have flourished since the 1980s. More recently medium energy cooling ($T_e = 1\text{--}10$ MeV) has re-gained a lot of interest [167–169]. Clearly the higher energy requires new technology and extrapolation to a new range of parameters. At Fermilab high energy e-cooling (with up to 5 MeV electrons using electrostatic acceleration for cooling of 8 GeV bunched antiprotons in the recycler) has been successfully developed and implemented. The generation and recirculation of the 4.3 MeV and 0.5 Ampere electron beam and its adaptation to the antiproton beam over a cooling length of 20 m are remarkable achievements. Finally the idea of ‘very high energy electron cooling’ ($T_e \geq 50$ MeV) has been revived as this might improve the luminosity of RHIC [168, 170]. At this energy the electron beam could circulate in a small ‘low-emittance storage ring’ with strong radiation damping. An attractive alternative is a scheme [170], in which the low-emittance beam after acceleration is re-decelerated after the passage through the cooling section to recuperate its energy.

In summary: 45 years after its invention, the field of electron cooling continues to expand with exciting old and new questions to be answered. Bunched beam cooling is no longer a magic barrier and even a merger between electron cooling and stochastic cooling i.e. the “coherent electron cooling” [171] appears at the horizon. In the concept of coherent electron information of the particle distribution of the hadron beam to be cooled is sampled by the electron beam, amplified and further downstream fed back onto the hadron beam.

6.10.2.4 Laser Cooling

Due to the pioneering work of the Heidelberg (TSR) [172] and Aarhus (ASTRID) [173] groups in the 1990s, laser cooling in storage rings has evolved into a very powerful technique. Longitudinal cooling times as short as a few milliseconds and momentum spreads as small as 10^{-6} are reported. These bright perspectives are somewhat mitigated by two specific attributes [174]: laser cooling takes place (mainly) in the longitudinal plane and it works (only) for special ions that have a closed transition between a stable (or meta-stable) lower state and a short-lived higher state. The transition is excited by laser light, and the return to the lower state occurs through spontaneous re-emission (Fig. 6.39). ‘Unclosed’ transitions, where the de-excitation to more than one level is possible, are not suited because ions decaying to the ‘wrong’ states are lost for further cooling cycles. This limits the number of ion candidates (although extended schemes with additional lasers to ‘pump back’ from the unwanted states could enlarge the number of ion species susceptible to cooling). Up to now, a few singly charged ions (like Li^{1+} , Be^{1+} or Mg^{1+}) have been used with ‘normal’ transitions accessible to laser frequencies. Transitions between fine structure, or even hyperfine levels of highly-charged heavy ions have also been considered, but in that case the cooling force is less pronounced and not so much superior to the electron cooling force which increases with charge (like $Q^{1.5}$ or even Q^2).

The laser irradiates the circulating ions co-linearly over the length of a straight section of the storage ring [174]. The absorption is very sharply resonant at the transition frequency. Then the Doppler shift ($\omega = (1 \pm v/c)\gamma\omega_{\text{laser}}$) seen by

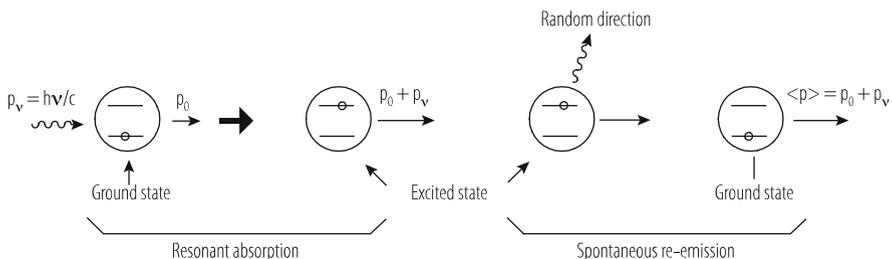


Fig. 6.39 Sketch of Laser-ion interaction

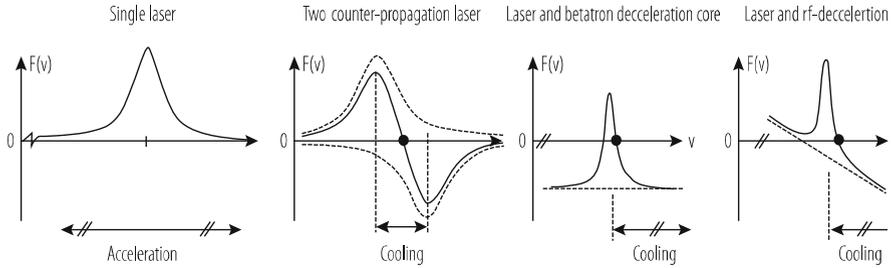


Fig. 6.40 Force $F(v)$ due to a single laser and different schemes for cooling to a fixed velocity

the ion makes the interaction strongly dependent on its velocity. This leads to a sharp resonance of the absorption as a function of the velocity (Fig. 6.40). The corresponding recoil (friction) force accelerates/decelerates the ions with a maximum rate at the resonant momentum. To obtain cooling to a fixed momentum, a second force $f(v)$ is necessary. It can be provided by a second (counter-propagating) laser or by a betatron core or by an RF-cavity, which decelerate the ions ‘towards the resonance of the first laser’ (Fig. 6.40). The interaction with the laser photons (and hence the cooling) takes place in the direction of the laser beam (longitudinal plane of the ions). De-excitation proceeds by re-emission of photons in all directions and this leads to heating of the ions in all three planes.

Through transverse-to-longitudinal coupling, part of the cooling can be transferred to the horizontal and vertical planes. Intra-beam scattering [175], dispersion [176] and special coupling cavities [177] have been considered for this purpose. Transfer by scattering and by dispersion has been demonstrated at the cooling rings, although the transverse cooling thus obtained was weak, a fact explainable by the weakness of the coupling.

The main motivation for laser cooling has been the goal of achieving ultra-cold *crystalline beams* [178] where the ions are held in place because the Coulomb repulsion overrides the energy of their thermal motion. A second application, cooling of low-charge states of heavy ions, was proposed [179] in order to prepare high-density drive beams for inertial confinement fusion. Several years ago a study [180] on the use of laser cooling of ions for the LHC was published. All these applications for the moment meet with difficulties: crystallisation, in full three-dimensional beauty, is hampered by the lattice properties of (present) storage rings and by the relative weakness of transverse cooling. Cooling for fusion is not fast enough [181] to ‘compress’ the high-intensity large-momentum-spread beam during the few milliseconds lifetime given by intra-beam charge exchange between the ions. And, finally laser cooling of highly charged ions for colliders meets with the competition of electron cooling and also with the restrictions on the choice of suitable ion species and states [180]. The investigations on laser cooling to obtain crystalline beams continue [182] and a special storage ring (S-LSR) with lattice properties apt to reach this goal [183, 184] has been built at Kyoto university.

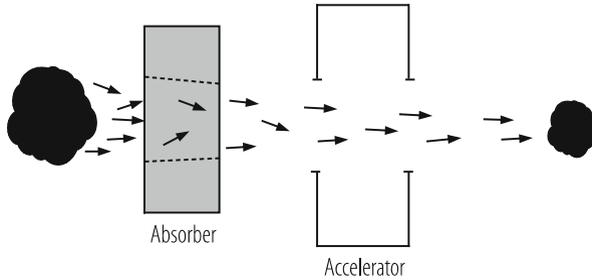


Fig. 6.41 Basic setup for (transverse) ionisation cooling (adapted from [186])

In conclusion: laser cooling in storage rings has led to very interesting and important results concerning the physics of cooling and cooling rings and, also atomic and laser physics. However, ‘accelerator applications’ like for electron or stochastic cooling are not realistic for the near future. The goal to obtain crystalline beams in special storage rings is under intense investigation.

6.10.2.5 Ionisation Cooling

Excellent reviews of ionisation cooling are given in papers by Skrinsky [185] and Neuffer [186]. The basic setup (Fig. 6.41) consists of a block of material (absorber) in which the particles lose energy, followed by an accelerating gap (RF-cavity) where the energy loss is restored. Losses in the absorber reduce both the longitudinal and the transverse momentum of the particle. The RF-cavity (ideally) only restores the longitudinal component and the net result is transverse cooling (Fig. 6.41). There is an obvious resemblance to radiation damping (Fig. 6.35), in which energy loss by synchrotron radiation followed by RF-acceleration results in cooling. Longitudinal ionisation cooling is also possible, especially in the range where the loss increases with energy (i.e. above the energy where the minimum of dE/ds occurs). At the expense of horizontal cooling, the longitudinal effect can be enhanced by using a wedge-shaped absorber in a region where the orbits exhibit dispersion with energy.

The statistical fluctuations (‘straggling’) of the loss and the angular (multiple) scattering introduce heating of the longitudinal and transverse emittances. The ratio of ionisation loss due to angular scattering favours light absorber material. Equilibrium emittances depend strongly on the lattice functions at the position of the absorber and the cavity. As in the case of radiation damping, the sum of the cooling rates (also in the case of a wedge absorber) is invariant with a value $J_x + J_y + J_E \approx 2 + J_E \approx 2$ for the ‘damping partition numbers’, instead of $J_x + J_y + J_E = 4$ for radiation damping. The quantity J_E depends on the slope of the dE/ds vs. E curve and is about constant and roughly equal to 0.12 for light materials above the minimum of dE/ds , but is strongly negative below. In terms of the partition numbers, the three emittance damping rates can be expressed by the energy loss ΔE_μ of the

muons in the absorber and the length Δs of the basic cell (Fig. 6.41) as:

$$\frac{1}{\varepsilon_i} \frac{d\varepsilon_i}{ds} = J_i \frac{1}{E_\mu} \frac{\Delta E_\mu}{\Delta s}. \quad (6.78)$$

A large number of cells or traversals through a cell is necessary to obtain appreciable emittance reduction.

Almost by a miracle, the muon mass falls into a narrow ‘window’ where ionisation cooling within the short life of the particle looks possible (although not easy). For electrons as well as for protons and heavier particles, the method is not practical, because the effect of bremsstrahlung (for e’s) and non-elastic processes in the absorber (for p’s), leads to unacceptable loss.

With the revival of interest for muon colliders and, related to that, neutrino factories [187], large collaborations (including more than 15 institutes, [188]) is undertaking a demonstration experiment. The ISIS accelerator at the Rutherford lab. is chosen for this task. Neutrino factory and muon collider proposals have to rely critically on muon cooling: typically 50 m to several 100 m long channels with solenoidal focussing (superconducting solenoids) are foreseen to reduce the phase-space of the muons emerging from pion decay. Liquid hydrogen absorbers, each 0.5–1 m in length, alternate with high-field accelerating cavities.

The variant selected by MICE is a ‘single particle experiment’ where one muon at a time is traced. Fast spectrometers, capable of resolving 1 muon per 25 ns, record/compare the three position coordinates and the three velocity components of the muon at the entrance and the exit of a short cooling section. Typically such a test-section should lead to 10% emittance reduction. The emittance pattern is ‘painted’ by a scatterer or a steering magnets changing the entrance conditions of the particle at random (scatterer) or in a programmed manner. A large number of muons are necessary to establish the six-dimensional phase-space reduction with sufficient statistics.

Apart from the spectrometers, other challenges can be identified: long term mechanical stability, muon decay and birth, contamination with other particles and non-linearities in focussing which deform the emittance pattern. In the coming years we will see a large effort on muon cooling scenarios and tests.

6.10.2.6 Cooling of Particles in Traps

In many experiments utilizing ion traps, the ions must first be cooled in order to perform high precision measurements. Cooling refers here to the reduction of kinetic energy of confined particles. A detailed review of cooling traps is given in [189] and the implementation of several cooling methods into a big project is described in [190].

With adequate modifications, most of methods discussed above for storage rings stochastic- [191], electron- [192], or laser cooling [193] can also be applied to traps.

Some (like stochastic cooling) are more difficult others (e.g. laser cooling) are much more powerful in the traps.

There are however several techniques which are especially adapted to or even working only in the environment of particles confined in a trap which is frequently cryogenically cooled. A gross classification is to divide them into lossy and lossless methods in terms of conservation of number of particles. A lossy technique (which in the strict definition of [100] would not be classified as “phase density cooling”) is evaporative cooling. In this case, just as in the evaporation of water, the more energetic molecules leave the trap and the temperature of the condensate is thereby strongly reduced. Experiments at the forefront of physics making Bose-Einstein condensates at a temperature of a tiny fraction of a degree have thus become possible [194].

An example of a widely used lossless process is resistive cooling: the trap electrodes are connected to an external circuit to dissipate energy from the ions through induced currents [189, 195] (Fig. 6.42).

In other words the particle’s kinetic energy is dampened by I^2R losses in a resistive circuit [195, 196]. Idealistically speaking, the resistor or the losses in a resonant circuit and absorb the particle’s energy to create a thermal equilibrium when there is no other heating source involved. Since the resistor has a specific physical temperature, it generates Johnson noise that in turn stochastically drives the trapped particles. Resistive cooling was first applied by H. Dehmelt and collaborators in 1975 [195].

To estimate the cooling time, a simple single particle model is used, where-by it is harmonically bound between two capacitor plates [195]. Due to this model, the energy is dampened with a time constant τ calculated by:

$$\tau = \frac{4mz_0}{q^2R}. \quad (6.79)$$

Here $2z_0$ is the separation of the capacitor plates (the electrodes of the trap) and R stands for the real part of the impedance from the attached external circuit, q is the charge and m the mass of the trapped particles. From Eq. (6.79) [189] one can easily conclude that light, highly charged particles are efficiently cooled. The cooling rate can be further improved by developing a high resistance in the external circuit.

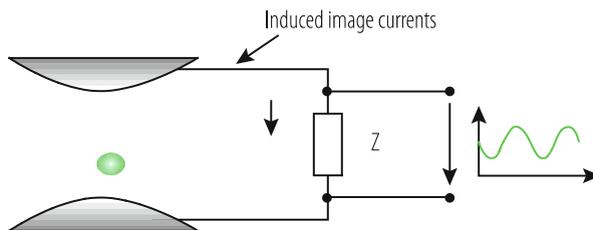


Fig. 6.42 Principle of resistive cooling of a trapped ion

In general, the external circuit which is often in vacuum and at cryo-temperature includes a low-noise amplifier to couple the induced current signal to room temperature, and thus enable plasma diagnostics. The input noise temperature of the amplifier is, depending on its coupling to the resonant circuit, closely related to the minimum achievable temperature of the particles in the trap. Most frequently, the impedance Z (with the real part R) shown in Fig. 6.42 is implemented as inductance L so that the circuit becomes resonant at the oscillation frequency of the ions. This is to tune out (compensate at resonance) the parasitic capacitance of the electrodes. This inductance may be implemented as discrete solenoid coil made of copper or superconducting wire. The quality factor $Q = R/\omega L$ of the tuned circuit has to be large to guarantee efficient resistive cooling. A high Q in turn means to incorporate a low loss network. As a caveat it should be mentioned that extremely high Q values (above say 10^5) may be problematic if the bandwidth of the resonance becomes smaller than the width of the particle spectrum. For further reading we refer to [197], where Shockley presents the basic equations for trapped and charged particles in a Penning trap.

There exist a large number of other cooling techniques used in traps, such as collisional cooling, RF- and optical sideband cooling (resolved and unresolved sideband methods) and sympathetic cooling. The list of examples given here is certainly not exhaustive and a detailed description can be found in review articles [189]. As for the term RF cooling (which may be confounded with RF-sideband cooling) [189] it should be pointed out that this refers to a reduction of temperature or vibration amplitude of a microscopic cantilevered bar in vacuum using an RF resonant circuit which is fed externally with a few milli-Watt of RF power [198].

Acknowledgements The authors would like to thank Michael Blaskiewicz, Paul Derwent, Christina Dimopoulou, Wolfgang Hofle, Fritz Nolden, Rolf Stassen and Junxia Wu for their contribution to this chapter and Jens Vigen for reading and correcting the manuscript.

References

1. E. Wilson, B.J. Holzer: *Beam Optics*, this handbook, Chapter 2.
2. B.J. Holzer: *Lattice Design in High Energy Accelerators*, Proc. CERN Accelerator School, CERN-2006-002.
3. LHC Design Report, CERN-2004-003.
4. W. Herr, B. Holzer, B. Muratori: *Concept of Luminosity*, this handbook, Sect. 6.4.
5. K. Steffen: *Periodic dispersion suppressors II*, DESY-HERA 83/02, 1983.
6. R. Brinkmann: *Insertions*, Proc. CERN Accelerator School, CERN-1987-010.
7. Proceedings of CERN Accelerator School: Beam Injection, Extraction and Transfer, 2017, Erice, CERN-2018-008-SP, 2018
8. V. Mertens. Kicker and septa technology, Chapter 5.5.
9. Y. Yamazaki ed., "Accelerator Technical Design Report for J-PARC", KEK Report 2002-13; JAERI-Tech 2003-044, 2003.
10. J. Wei et al., Evolution of the Spallation Neutron Source Ring Lattice, Proc 20th ICFA Advanced beam dynamics workshop, Chicago IL, 2002,

11. W. Bartmann, B. Goddard and C. Hessler, a doublet-based injection-extraction straight for PS2, Proceedings of IPAC'10, Kyoto, Japan, pp. 3945-3947, 2010.
12. O.D. Kazachkovskii, Particle injection into a ring accelerator, Atomic Energy Vol. 84, Number 5, 345-350, 1998.
13. A. Hilaire, V. Mertens and E. Weiss, Beam transfer to and Injection into LHC, Proc. 6th European Particle Accelerator Conference, Stockholm, Sweden, pp. 2117-2119, 1998.
14. J. Uythoven et al., Fast injection into the PS2, Proceedings of PAC09, Vancouver, BC, Canada, 2009.
15. 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, pp. 66-68, 2007.
16. G.I. Budker et al., Experiments on charge-exchange injection of protons into a storage ring, Atomnaya Énergiya, Vol. 19, No. 6, pp. 507-510, December, 1965.
17. J.W.G. Thomason et al., Injection Upgrades for the ISIS Synchrotron, Proceedings of IPAC'10, Kyoto, Japan, pp 705-707, 2010.
18. J. Beebe-Wang et al., Transverse phase space painting for SNS accumulator ring injection, Proc. 1999 Particle Accelerator Conference, New York, pp. 1743-1745, 1999.
19. V. Danilov et al., Three-step H- charge exchange injection with a narrow-band laser, Phys. Rev. ST Accel. Beams 6, 2003.
20. T.S. Ueng et al., Topping Up Experiments at SRRC, Proceedings of the 5th European Particle Accelerator Conference, 10 to 14 June 1996.
21. S. Myers, A Possible New Injection and Accumulation Scheme for LEP, CERN LEP Note 334, 1981.
22. P. Collier, Synchrotron phase space injection into LEP, Proc. 16th IEEE Particle Accelerator Conference, Dallas, TX, USA, pp. 551-553, 1995.
23. B. Goddard and V. Mertens, Moving the beam into and out of the LHC, Chapter 4.4, in 'The Large Hadron Collider: a marvel of technology', L. Evans (ed), EPFL press, Laussane, p 114-133, 1999.
24. M.Q. Barton, Beam extraction from synchrotrons, Proc. VIIIth Int. Conference on High Energy Accelerators, CERN, Geneva, p. 85-8, 1971.
25. 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, pp 1892-1894, 2005.
26. L. Badano and S. Rossi, Characteristics of a betatron core for extraction in a proton-ion medical synchrotron, CERN-PS-97-019-DI, 1997.
27. M. Fraser et al., Improvements to the SPS Slow Extraction for High Intensity Operation, CERN-ACC-NOTE-2019-0010, 2019.
28. S. van der Meer, Stochastic extraction, a low-ripple version of resonant extraction, CERN, CERN/PS/AA 78-6, 1978.
29. M. Fraser et al., SPS Slow Extraction Losses and Activation: Challenges and Possibilities for Improvement, CERN-ACC-2017-123, 10.18429/JACoW-IPAC2017-MOPIK045 2017.
30. 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, p. 438, 1973.
31. R. Cappi and M. Giovannozzi, Multiturn extraction and injection by means of adiabatic capture in stable islands of phase space, Phys. Rev. ST Accel. Beams 7, 2004.
32. K. Johnsen, CERN Intersecting Storage Rings (ISR), Proc. Nat. Acad. Sci. USA, Vol. 70, No. 2, pp. 619-626, 1973.
33. D. Neuffer, Design of muon storage rings for neutrino oscillation experiments, IEEE Transactions on Nuclear Science, Vol. NS-28, No. 3, 1981.
34. S. van der Meer, Stochastic cooling in the CERN antiproton accumulators, IEEE Transactions on Nuclear Science, Vol. NS-28, No. 3, 1981.
35. 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.

36. V. Biryukov, Computer Simulation of Crystal Extraction of Protons from a Large-Hadron-Collider Beam, Phys. Rev. Lett. 74, 2471–2474, 1995.
37. W. Herr, *Concept of Luminosity*, Proceedings of CERN Accelerator School, Zeuthen 2003, CERN-2006-002 (2006).
38. C. Moller, K. Danske Vidensk. Selsk. Mat.-Fys. Medd., **23**, 1 (1945).
39. A. Chao and M. Tigner, **Handbook of Accelerator Physics and Engineering**, World Scientific (1998).
40. W. Herr, *Beam-beam interactions*, Proceedings of CERN Accelerator School, Zeuthen 2003, CERN-2006-002 (2006).
41. J.E. Augustin, *Space charge effects in e^+e^- storage rings with beams crossing at an angle*, Orsay, Note interne 35-69 (1969).
42. P. Raimondi, 2nd SuperB Workshop, Frascati, Italy, (2006).
43. Review of Particle Physics, Vol. **15**, Number 1-4, 213 (2000).
44. C. Augier et al., Physics Letters, **B 315**, 503 (1993).
45. C. Augier et al., Physics Letters, **B 316**, 448 (1993).
46. C. Augier et al., Physics Letters, **B 344**, 451 (1993).
47. J. D. Jackson: Classical Electrodynamics, John Wiley & Sons, 1998.
48. A. Hofmann: The Physics of Synchrotron Radiation, Cambridge, 2004.
49. K.W. Robinson: Phys. Rev. 111 (1958) 373.
50. H. Wiedemann: Particle Accelerator Physics, Springer, 2007.
51. Los Alamos Accelerator Code Group, <http://laacg1.lanl.gov/laacg/componl.html>
52. ASTEC, <https://projects.astec.ac.uk/Plone/Codes/optics/>
53. A. Chao and M. Tigner, **Handbook of Accelerator Physics and Engineering**, World Scientific (1998).
54. W. Herr and E. Forest, **Non-linear dynamics**, This handbook.
55. E. Forest, **Beam Dynamics**, Harwood Academic publishers (1998).
56. A. Dragt and J. Finn, J. Math. Phys., **17**, 2215 (1976); A. Dragt et al., Ann. Rev. Nucl. Part. Sci., **38**, 455 (1988).
57. M. Sands: SLAC Report 121 (1969).
58. J.M. Jowett: AIP Conf. Proc. 153 (1985) 934.
59. J.D. Jackson: Rev. Mod. Phys. 48 (1976) 417.
60. A. Chao: AIP Conf. Proc. 153 (1985) 103.
61. A.W. Chao: J. Appl. Phys. 50 (1979) 595; DOI:10.1063/1.326070.
62. P. Raimondi: *Status on SuperB effort*, Second SuperB Workshop, LNF, Frascati (Italy), March 2006.
63. P. Raimondi, D. Shatilov, M. Zobov: *Beam-beam issues for colliding schemes with large Piwinski angle and crabbed waist*, Preprint physics/0702033.
64. M. Zobov et al: *Test of crab-waist collisions at DAΦNE Φ-Factory*, Phys. Rev. Lett. 104 (2010).
65. D. A. Edwards, M. J. Syphers, An introduction to the physics of high energy accelerators, John Wiley & Sons, Inc. New York, NY, 1993
66. J. Rossbach, P. Schmüser, Basic course on accelerator optics, CERN accelerator school, fifth general accelerator physics course, CERN 94-01 vol. 1, 1994
67. B. de Raad, The CERN SPS proton-antiproton collider, IEEE Transactions on Nuclear Science, Vol. NS-32, No. 5, October 1985
68. R. Bailey, E. Brouzet, K. Cornelis, L. Evans. A. Faugier, V. Hatton, J. Miles, R. Schmidt, D. Thomas, High luminosity performance of the SPS proton-antiproton collider, conf. proc. Particle Accelerator Conference 1989, Chicago, 1989
69. R. Schmidt, Beam-beam observations in the SPS proton antiproton collider, Particle Accelerators, 1995, Vol. 50, pp. 47-60
70. LHC design report, CERN 2004-003 vol. 1, 2004
71. A. Chance, D. Boutin, B. Dalena, B. Holzer, A. S. Langner, D. Schulte, Updates on the Optics of the Future Hadron-Hadron Collider FCC-hh, conf. proc. IPAC2017, Copenhagen, 2017.
72. HERA Design Report, DESY HERA-1981-010

73. N. Auerbach and S. Chlomo, *Phys. Rev. Lett.* 103, 172501 (2009).
74. B. V. Jacak and B. Müller, *Science* 337, pp. 310-314 (2012).
75. The STAR collaboration, *Nature*, Vol. 548, pp. 62-65 (2017).
76. The STAR collaboration, *Nature* 473, pp. 353-35 (2011).
77. B. Dönigus, *Nucl. Phys. A*, 904-905, pp. 547c-550c (2013).
78. M. Wilde et al., *Nucl. Phys. A* 904-905, pp. 547c-550c (2013).
79. P. Asbo-Hansen, et al.: *IEEE Trans. Nucl. Sci.* 24 (1997) 1557.
80. M. Boutheon, et al.: *IEEE Trans. Nucl. Sci.* 28 (1981) 2049.
81. A.W. Chao, M. Tigner (eds.): *Handbook of accelerator physics*, World Scientific (1999).
82. M. Okamura, *HIAT2015*, pp. 274-276 (2015).
83. J.G. Alessi et al., *PAC2011*, pp. 1966-1968 (2011).
84. W. Fischer et al., *Phys. Rev. ST Accel. Beams* 11, 041002 (2008).
85. M. Blaskiewicz et al. *PAC2001*, pp. 3026-3028 (2003).
86. M. Benedict, P. Collier, V. Mertens, J. Poole and K. Schindl (eds.), *CERN-2004-003-V3* (2004).
87. D. Manglunki et al., *IPAC 2012*, pp. 3752-3754 (2012).
88. M. Chanel, *Nucl. Instrum. Methods A* 532, pp. 137-143 (2014).
89. D. Boussard, J.M. Brennan, T. Linnecar, *PAC1995*, pp. 1506-1508 (1995).
90. M. Blaskiewicz, J.M. Brennan, K. Mernick: *Phys. Rev. Lett.* 105 (2010) 094801.
91. C. Montag, et al.: *Phys. Rev. ST Accel. Beams* 5 (2002) 084401.
92. O. Brüning, et al. (eds.): *LHC Design Report*, CERN-2004-003 (2004).
93. J.M. Jowett, et al.: *Proc. EPAC2004*, Luzern, Switzerland (2004) 578.
94. R. Bruce, et al.: *Phys. Rev. ST Accel. Beams* 12 (2009) 071002.
95. R. Bruce, et al.: *Phys. Rev. ST Accel. Beams* 12 (2009) 011001.
96. R. Bruce, et al.: *Phys. Rev. Lett.* 99 (2007) 144801.
97. J.M. Jowett, M. Schaumann et al; *IPAC2016*, Busan, Korea (2016) 1493.
98. J. M. Jowett, *IPAC2018*, Vancouver, Canada (2018) 584.
99. J.M. Jowett, C. Carli: *Proc. EPAC2006*, Edinburgh, Scotland (2006) 550.
100. A. Sessler, "Comment on the word "cooling" as used in Beam Physics," in *International Workshop on Beam Cooling and Related Topics - COOL05*, Galena, Illinois (USA), 2006.
101. A. A. Kolomenski and A. N. Lebedev, "The effect of radiation on the motion of relativistic electrons in synchrotrons," in *Proc. CERN Symposium*, Geneva, 1956.
102. P. Bramham et al., "Stochastic cooling of a stored proton beam," *Nucl. Instr. Meth.*, vol. 125, pp. 201-202, 1975.
103. G. Carron et al., "Experiments on stochastic cooling in ICE (Initial Cooling Experiment)," *IEEE Trans. Nucl. Sci. NS-26*, p. 3456, 1979.
104. S. van der Meer, "Stochastic cooling and the accumulation of antiprotons," *Rev. Mod. Phys.*, vol. 57, p. 689, 1985.
105. D. Möhl, "Stochastic cooling," in *Proc. of CERN Accelerator School*, Rhodes, 1993.
106. F. Caspers, "Techniques of stochastic cooling," in *Workshop on Beam Cooling and Related Topics*, Bad Honnef, 2001.
107. D. Möhl, "The status of stochastic cooling," *Nucl. Instrum. Meth. A*, vol. 391, p. 164, 1997.
108. R. Giannini, P. Lefevre and D. Möhl, "Super LEAR, conceptual machine design," *Nucl. Phys. A*, vol. 558, p. 519, 1993.
109. T. Katayama and H. Tsutsui, "Electron-RI collider and internal target operation of RIKEN storage ring project," *Nucl. Instrum. Methods Phys. Res., A*, vol. 532, pp. 157-171, 2004.
110. FAIR Technical Design Team, "Fair Baseline Technical Report, vols. 1-6," GSI, Darmstadt, 2007.
111. S. Wunderlich et al., "Progress of the RF-system developments for stochastic cooling at the FAIR Collector Ring," in *International Workshop on Beam Cooling and Related Topics (COOL'15)*, Newport News, 2015.
112. C. Dimopoulou, "Stochastic cooling for FAIR," *ICFA Beam Dynamics Newsletter*, no. 64, pp. 106-120, 2014.

113. F. Nolden et al., "Radioactive beam accumulation for a storage ring experiment with an internal target," in *4th Int. Particle Accelerator Conf. (IPAC'13)*, Shanghai, China, 2013.
114. J. M. Brennan and M. Blaskiewicz, "Bunched beam stochastic cooling project for RHIC," *AIP Conf. Proc.*, vol. 821, pp. 185-189, 2006.
115. R. J. Pasquinelli, "Stochastic cooling technology at Fermilab," *Nucl. Instrum. Methods Phys. Res., A*, vol. 532, pp. 313-320, 2004.
116. R. Stassen et al., "Recent developments for the HESR stochastic cooling system," in *International Workshop on Beam Cooling and Related Topics (COOL'07)*, Bad Kreuznach, 2007.
117. C. Peschke et al., "Prototype pick-up module for CR stochastic cooling at FAIR," in *Workshop, COOL'09*, Lanzhou, 2009.
118. J. Marriner, "Stochastic cooling overview," *Nucl. Instrum. Methods Phys. Res., A*, vol. 532, pp. 11-18, 2004.
119. F. Caspers, "Planar slotline pick-ups and kickers for stochastic cooling," in *1987 IEEE Particle Accelerator Conference (PAC 87)*, Washington, 1987.
120. J. Petter, D. McGinnis and J. Marriner, "Novel stochastic cooling pickups/kickers," in *IEEE Particle Accelerator Conference (PAC 89)*, Chicago, 1989.
121. L. Faltin, "Slot type pickup and kicker for stochastic beam cooling," *Nucl. Instrum. Meth.*, vol. 148, pp. 449-455, 1978.
122. D. McGinnis, "The 4-GHz to 8-GHz stochastic cooling upgrade for the Fermilab debuncher," in *Particle Accelerator Conference (PAC 99)*, New York, 1999.
123. G. Carron, F. Caspers and L. Thorndahl, "Development of power amplifier modules for the ACOL stochastic cooling systems," CERN, Geneva, 1985.
124. D. Boussard, "Frontiers of particle beams," in *Joint US - CERN School on Particle Accelerators: Topical Course on Frontiers of Particle Beams*, Vols. Lecture Notes in Physics, vol 296, M. Month and S. Turner, Eds., South Padre Island, Springer, 1986, pp. 269-296.
125. M. Blaskiewicz, J. M. Brennan and K. Mernick, "Three-dimensional stochastic cooling in the Relativistic Heavy Ion Collider," *Phys. Rev. Lett.*, vol. 105, p. 094801, 2010.
126. F. Caspers, "A novel type of forward coupler slotted stripline pickup electrode for non-relativistic particle beams," CERN, Geneva, 2011.
127. J. X. Wu et al., "A novel type of forward coupler slotted stripline pickup electrode for CSRE stochastic cooling," in *4th International Particle Accelerator Conference*, Shanghai, 2013.
128. H. Stockhorst et al., "Status of stochastic cooling predictions at the HESR," in *2nd International Particle Accelerator Conference (IPAC 2011)*, San Sebastian, 2011.
129. C. Carli and F. Caspers, "Stochastic cooling at the CERN Antiproton Decelerator," in *7th European Particle Accelerator Conference (EPAC 2000)*, 2000.
130. T. Eriksson et al., "AD status and consolidation plans," in *International Workshop on Beam Cooling and Related Topics (COOL13)*, Mürren, 2013.
131. R. J. Pasquinelli, "Bulk acoustic wave (BAW) devices for stochastic cooling notch filters," in *IEEE Particle Accelerator Conference (PAC 91)*, San Francisco, 1991.
132. W. Maier et al., "The novel optical notch filter for stochastic cooling at the ESR," in *International Workshop on Beam Cooling and Related Topics (COOL'13)*, Mürren, 2013.
133. R. Stassen et al., "The HESR stochastic cooling system, design, construction and test experiments in COSY," in *11th International Workshop on Beam Cooling and Related Topics (COOL2017)*, Bonn, 2017.
134. D. R. Broemmelsiek et al., "Bunched beam stochastic cooling in the Fermilab Recycler Ring," in *Particle Accelerator Conference (PAC 05)*, Knoxville, 2005.
135. P. F. Derwent et al., "Performance and upgrades of the Fermilab Accumulator Stacktail Stochastic Cooling," in *International Workshop on Beam Cooling and Related Topics (COOL 05)*, 2005.
136. V. A. Lebedev, "Improvements to the stacktail and debuncher cooling systems," in *International workshop on beam cooling and related topics (COOL 09)*, Lanzhou, 2009.
137. H. Herr and D. Möhl, "Bunched beam stochastic cooling," in *Workshop on the Cooling of High-Energy Beams*, Madison, 1978.

138. R. J. Pasquinelli, "Twenty-five years of stochastic cooling experience at Fermilab," in *International Workshop on Beam Cooling and Related Topics (COOL09)*, Lanzhou, 2009.
139. G. Jackson et al., "Bunched beam stochastic cooling in the Fermilab Tevatron Collider," in *International Conference on Particle Accelerators (PAC'93)*, Washington, 1993.
140. F. Caspers and D. Möhl, "Stochastic cooling in hadron colliders," in *17th International Conference on High-Energy Accelerators*, Dubna, 1998.
141. M. Blaskiewicz et al., "Longitudinal solitons in RHIC," in *2003 Particle Accelerator Conference (PAC 2003)*, 2003.
142. J. Bosser et al., "LEAR MD report: bunched beam Schottky spectrum," CERN, Geneva, 1994.
143. J. Murrin et al., "Bunched beam cooling in the FNAL Antiproton Accumulator," in *2nd European Particle Accelerator Conference*, Nice, 1990.
144. A. A. Mikhailichenko and M. S. Zolotarev, "Optical stochastic cooling," *Phys. Rev. Lett.*, vol. 71, p. 4146–4149, 1993.
145. A. Zholents and M. Zolotarev, "Transit time method of optical stochastic cooling," *Phys. Rev. E*, vol. 50, p. 3087, 1994.
146. W. A. Franklin et al., "Optical stochastic cooling experiment at the MIT-Bates South Hall Ring," in *Workshop on Beam Cooling and Related Topics (COOL 07)*, Bad Kreuzbach, 2007.
147. M. Blaskiewicz, J. M. Brennan and F. Severino, "Operational stochastic cooling in the Relativistic Heavy-Ion Collider," *Phys. Rev. Lett.*, vol. 100, p. 174802, 2008.
148. J. M. Brennan and M. Blaskiewicz, "Stochastic cooling in RHIC," in *Particle Accelerator Conference (PAC 09)*, Geneva, 2009.
149. S. Antipov et al., "IOTA (Integrable Optics Test Accelerator): facility and experimental beam physics program," *JINST*, vol. 12, p. T03002, 2017.
150. S. van der Meer, "Stochastic stacking in the Antiproton Accumulator," CERN, Geneva, 1978.
151. B. Autin et al., "Design study of a proton-antiproton colliding beam facility," CERN, Geneva, 1978.
152. E. J. N. Wilson, "Design study of an antiproton collector for the antiproton accumulator (ACOL)," 1983.
153. V. Lebedev and V. Shiltsev, *Accelerator physics at the Tevatron Collider*, New York: Springer, 2014.
154. R. J. Pasquinelli, "Implementation of stochastic cooling hardware at Fermilab's Tevatron Collider," *JINST*, vol. 6, p. T08002, 2011.
155. S. Nagaitsev, L. Prost and A. Shemyakin, "Fermilab 4.3 MeV electron cooler," *JINST 10*, vol. 10, no. 01, p. T01001, 2015.
156. G. L. Budker, "Experiences sur l'obtention d'un faisceau intense de protons par la methode d'injection par echange de charges," in *Symposium International sur les anneaux de collisions a electrons et positrons, Saclay, 26–30 septembre 1966*, Saclay, 1967.
157. J. D. Jackson, *Classical electrodynamics*, New York: John Wiley & Sons, 1975.
158. H. Poth, "Electron Cooling," in *CAS - CERN Accelerator School: Accelerator Physics, Oxford, UK, 16 - 27 Sep 1985*, Vols. CERN-1987-003-V-2, Geneva, CERN, 1987, pp. 534-569.
159. Y. Derbenev and I. Meshkov, "Studies on electron cooling of heavy particle beams made by the VAPP-NAP group at the Nuclear Physics Institute of the Siberian branch of the USSR Academy of Science at Novosibirsk," CERN, Geneva, 1977.
160. G. Budker et al., "Experimental studies of electron cooling," *Part. Accel.*, vol. 7, pp. 197-211, 1976.
161. H. Poth, "Electron cooling: theory, experiment, application," *Phys. Rep.*, vol. 196, p. 135, 1990.
162. A. Y. Larentev and I. N. Meshkov, "The computation of electron cooling process in a storage ring," in *International Conference on Crystal Beams "Ettore Majorana"*, Erice, 1996.
163. M. Chanel, "Ion accumulation for LHC," in *Workshop on Beam Cooling and Related Topics*, Bad Honnef, 2001.

164. J. Bosser et al., "Experimental investigation of electron cooling and stacking of lead ions in a low-energy accumulation ring," *Part. Acc.*, vol. 63, p. 171, 1999.
165. G. Tranquille, "Optimum parameters for electron cooling," CERN, Geneva, 2001.
166. J. Bosser et al., "Stability of cooled beams," *Nucl. Instrum. Methods Phys. Res. A*, vol. 441, pp. 1-8, 2000.
167. S. Nagaitsev et al., "Antiproton cooling in the Fermilab Recycler Ring," in *International Workshop on Beam Cooling and Related Topics (COOL05)*, 2005.
168. A. V. Fedotov, "Progress of high-energy electron cooling for RHIC," in *International Workshop on Beam Cooling and Related Topics (COOL07)*, Bad Kreuznach, 2007.
169. D. Reistad, "Calculations on high-energy electron cooling in the HESR," in *International Workshop on Beam Cooling and Related Topics (COOL07)*, Bad Kreuznach, 2007.
170. I. Ben-Zvi, "R&D towards cooling of the RHIC Collider," *Nucl. Instrum. Meth. A*, vol. 532, pp. 177-183, 2004.
171. Y. Derbenev, "Coherent electron cooling," *Phys. Rev. Lett.*, vol. 102, p. 114801, 2009.
172. W. Petrich et al., "Laser cooling at the Heidelberg Test Storage Ring (TSR)," in *Workshop on Electron Cooling and new Cooling Techniques*, Legnaro, 1990.
173. N. Kjærgaard et al., "Recent results from laser cooling experiments in ASTRID – real-time imaging of ion beams," *Nucl. Instrum. Meth. A*, vol. 441, p. 196, 2000.
174. E. Bonderup, "Laser cooling," in *CAS - CERN Accelerator School: 5th Advanced Accelerator Physics Course, Rhodes, Greece, 20 Sep - 1 Oct 1993*, Vols. CERN Report 95-06, S. Turner, Ed., Rhodes, CERN, 1993, pp. 731-747.
175. H. J. Miesner et al., "Efficient, indirect transverse laser cooling of a fast stored ion beam," *Phys. Rev. Lett.*, vol. 77, p. 623, 1996.
176. I. Lauer et al., "Transverse laser cooling of a fast stored ion beam through dispersive coupling," *Phys. Rev. Lett.*, vol. 81, p. 2052, 1998.
177. H. Okamoto, A. M. Sessler and D. Möhl, "Three-Dimensional laser cooling of stored and circulating ion beams by means of a coupling cavity," *Phys. Rev. Lett.*, vol. 72, p. 3977, 1994.
178. D. Maletic and A. Ruggiero, "Crystalline beams and related issues," in *31st INFN ELOISATRON Workshop*, Erice, 1996.
179. D. Habs and R. Grimm, "Crystalline Ion Beams," *Ann. Rev. Nucl. Sci.*, vol. 45, pp. 391-428, 1995.
180. N. Madsen and J. S. Nielsen, "Laser-cooling for light ion accumulation," in *7th European Particle Accelerator Conference (EPAC 2000)*, Vienna, 2000.
181. G. Plass, "The status of the HIDIF study," *Nucl. Instrum. Meth. A*, vol. 415, p. 204, 1998.
182. A. Noda et al., "Present status and recent activity on laser cooling at S-LSR," in *International Workshop On Beam Cooling And Related Topics (COOL07)*, Bad Kreuznach, 2007.
183. A. Noda et al., "Laser cooling for 3-D crystalline state at S-LSR," in *International Workshop on Beam Cooling and Related Topics (COOL05)*, 2005.
184. M. Ikegami et al., "Heavy ion storage ring without linear dispersion," *Phys. Rev. Accel. Beams*, vol. 7, p. 120101, 2004.
185. A. N. Skrinsky, "Ionization cooling and muon collider," *Nucl. Instrum. Meth. A*, vol. 391, pp. 188-195, 1997.
186. D. N. D. Neuffer, "Introduction to muon cooling," *Nucl. Instrum. Meth. A*, vol. 532, pp. 26-31, 2004.
187. D. Kaplan, "Muon-cooling research and development," *Nucl. Instrum. Meth. A*, vol. 532, pp. 241-248, 2004.
188. A. Blondel and G. Hanson, "MICE status report September 2010," 2010.
189. W. M. Itano et al., "Cooling methods in ion traps," *Phys. Scr.*, vol. T59, pp. 106-120, 1995.
190. F. Herfurth et al., "Highly charged ions at rest: The HITRAP project at GSI," Wako, 2005.
191. N. Beverini et al., "Stochastic cooling in Penning traps," *Phys. Rev. A*, vol. 38, no. 1, pp. 107-114, 1988.
192. D. S. Hall and G. Gabrielse, "Electron cooling of protons in a nested Penning trap," *Phys. Rev. Lett.*, vol. 77, no. 10, p. 1962, 1997.

193. S. Chu, "Nobel Lecture: The manipulation of neutral particles," *Rev. Mod. Phys.*, vol. 70, p. 685, 1996.
194. E. A. Cornell and Wieman, "Nobel lecture: Bose-Einstein condensation in a dilute gas: the first 70 years and some recent experiments," *Int. J. Mod. Phys. B*, vol. 16, no. 30, pp. 4503-4536, 2002.
195. H. G. Dehmelt, "Radiofrequency spectroscopy of stored ions I: Storage," *Adv. At. Mol. Phys.*, vol. 3, pp. 53-72, 1968.
196. D. J. Wineland and H. G. Demelt, "Principles of the stored ion calorimeter," *J. Appl. Phys.*, vol. 46, p. 919, 1975.
197. W. J. Shockley, "Currents to conductors induced by a moving point charge," *Appl. Phys.*, vol. 9, p. 635, 1938.
198. K. R. Brown et al., "Passive cooling of a micromechanical oscillator with a resonant electric circuit," *Phys. Rev. Lett.*, vol. 99, p. 137205, 2007.
199. S. van der Meer, "Stochastic damping of betatron oscillations," CERN, Geneva, 1972.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

