Accelerator Physics and Technology

V. Ziemann, Uppsala University

November 14, 2006

Abstract

Class notes for the Accelerator Physics Course at Uppsala University during the winter 2006.

1 Introduction

The goal of the course is to discuss the concepts and the tools needed to design a modern particle accelerator. This knowledge is useful both for the future accelerator physicist and the user of accelerators, such as a synchrotron radiation user, a high-energy or a nuclear physicist. The participation in experiments conducted at accelerators and the survival in accelerator control rooms of the experimentalist is much facilitated by a basic knowledge of the concepts and the jargon used in the accelerator physics community. Being exposed to some simple design tools such as the beam optics code MAD or the magnet design code MaxwellSV will help to enhance this understanding. At the end of the course, I hope that the participants have acquired this basic understanding and are able to roughly sketch and design an accelerator for a tropical island, if stranded on it with their laptop.

To summarize the two goals of this course:

- Be able to do a zeroth-order design of a particle. accelerator
- Survival in an accelerator control room.

I will start the course by a quick historical tour that briefly discusses the first accelerators that emerged during the first half of the 20th century, followed by a discussion of the different types of accelerators that were built during the second half. In particular we will briefly illuminate the pro and cons of different types of accelerators.

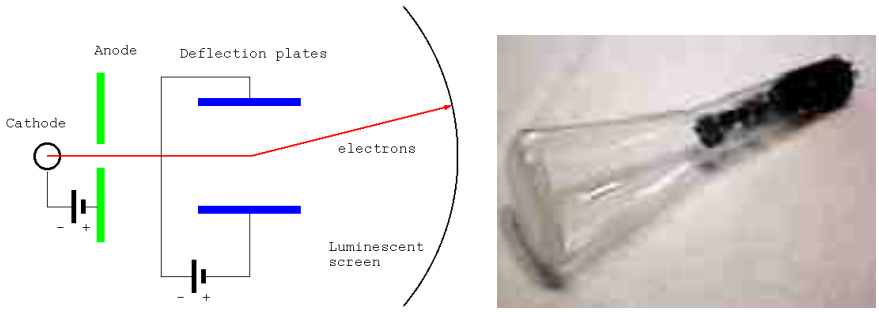

Figure 1: Schematic and picture of a cathode ray tube.

1.1 The early Accelerators

On of the simplest device where accelerated particles are used in the analysis of the physical world is the cathode ray tube, which is a close relative of the oscilloscope or the television. Many of the pertinent features of accelerators are already present in this long-known device.

1.1.1 Cathode Ray Tube

The cathode ray tube was used by Karl Ferdinand Braun already in 1897 to make electrical oscillations visible by directing accelerated electrons in a vacuum tube onto a luminescent screen with the help of electric and magnetic fields. On the left of fig. 1 a photo taken from [6] of such a tube is shown. On the right we show the schematics of a cathode ray tube. The electrons are created in a thermionic cathode, which is basically a heated wire and are extracted by the electric field generated by the voltage between the cathode and the anode. Another voltage, which can be time-varying, and is applied to deflection plates directs the beam to the luminescent screen where the electrons are detected. The entire setup is embedded in a vacuum tube to avoid collisions of the electrons with the rest gas.

As mentioned before, we already have the essential ingredient of an accelerator present: particle creation, acceleration, guide field, detection or diagnostic, and vacuum. Note that in a TV tube the deflection plates are replaced by magnetic coils and the deflection is done both horizontally and vertically.

Exercise 1.1.1: Calculate the electron trajectory. Assume reasonable voltages and distances for a small tabletop experiment.

1.1.2 Van-de-Graaff Electrostatic Accelerator

From 1929 on R. van-de-Graaf developed an electrostatic accelerator in which charge is deposited on a belt made of silk or any other isolating elastic material such as rubber by a corona- or a triboelectric discharge. The belt carrying the charge is moved by pulleys to the upper platform where the charge is deposited

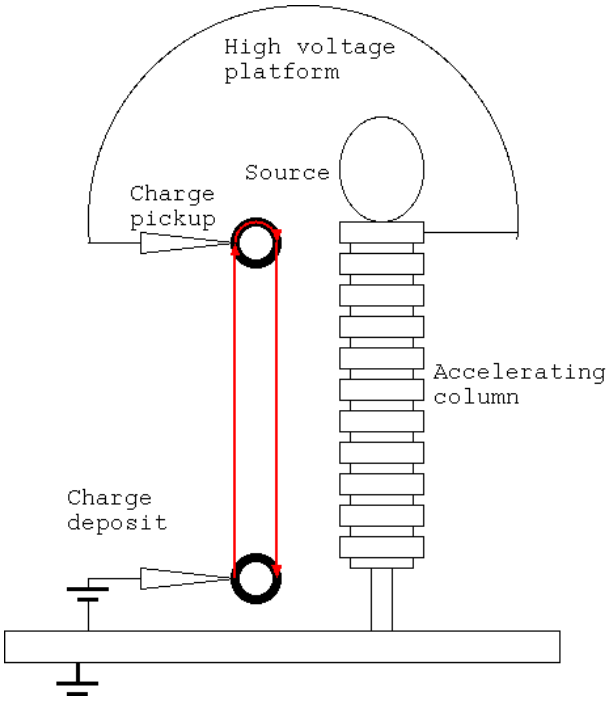

Figure 2: Schematics of a van de Graaff accelerator.

at a higher potential. If an ion source is positioned at the higher potential and the beam travels inside a column where the local potential is determined by a resistor cascade, the beam can be accelerated. A modern version of this device is called a Pelletron and is e.g. used in high-energy electron-coolers. A tandem accelerator uses the high voltage generated by a van de Graaff to accelerate negatively charged ions and strip several electron off of the ion to use the same acceleration potential twice.

1.1.3 Cockroft-Walton

A method to multiply alternating voltages to very high voltages by using a network made of capacitors and diodes, was first found by Greinacher a cascaded version reaching much higher accelerating voltages was later used by Cockroft and Walton in nuclear physics experiments. The basic principle of the voltage multiplier can be explained by looking at Fig. 3. The negative voltage of the AC voltage U_s that is applied at the left-hand side charges capacitor C_1 to U_s and the positive voltage half an oscillation period later charges C_2 to $2U_s$. Cascading these units allows to reach voltages into the MV range.

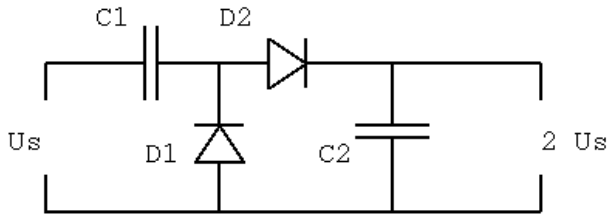

Figure 3: The Greinacher voltage multiplier that is the base of a Cockroft-Walton accelerator.

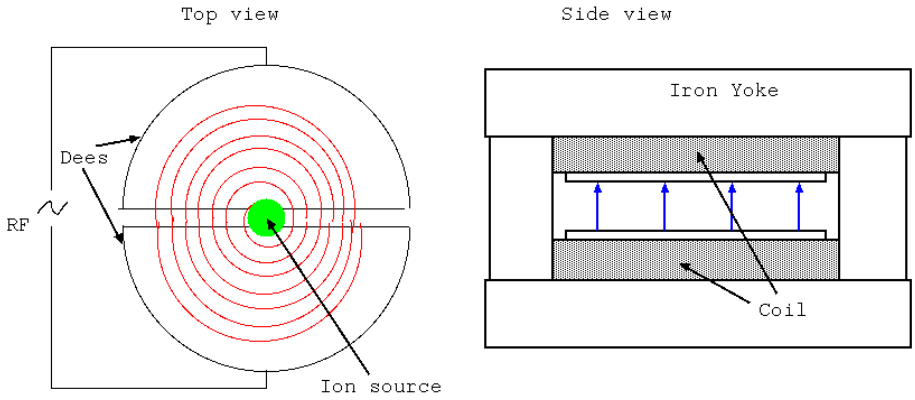

Figure 4: Schematics of a cyclotron.

1.1.4 Cyclotron

The working principle of the cyclotron in which charged particles are accelerated in a static magnetic field by repeated change of polarity of two accelerating 'dees', see Fig.4 was found by Lawrence in 1929. The particles are created in a ion source in the center of the cyclotron and accelerated in the gap between the dees. While the particle travels inside the dee the polarity between the dees changes and the particles can be accelerated when they traverse the next gap again. This process repeats over and over again. Since the magnetic field is constant the radius of the orbit increases as is indicated by the spiral drawn in Fig. 4. Synchronism is guaranteed as long as the particles move a non-relativistic speeds up to energies of about 100 MeV. Specially shaped magnetic fields and variable RF generators allow, however, to reach moderately relativistic energies up to about 600 MeV (PSI, TRIUMF).

1.1.5 Betatron

The induction voltage generated by a temporally varying magnetic field is used to accelerate electrons up to about 200 MeV. In this way the betatron resembles

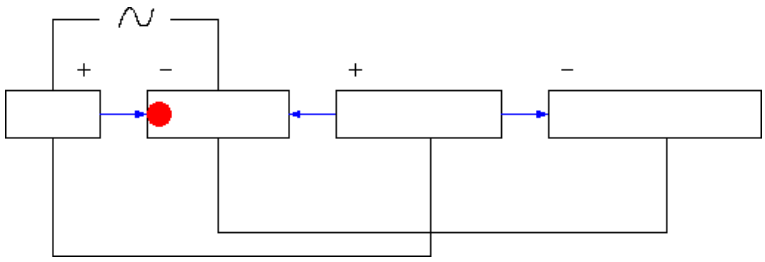

Figure 5: Schematics of a Drift-tube linac also called an Alvarez structure.

a transformer, where the magnet coil is the primary winding and the accelerated electron beam the secondary winding. The principle behind the betatron was invented by Kerst around 1940 and these machines are often used for the generation of X-rays by impinging the accelerated electrons on a target made of, for example, tungsten.

1.1.6 Drift-tube Linac, Alvarez Structure

In the cyclotron the particles hide inside the dees while the electric field that drives the dees reverses polarity. The same principle can also be used in a particular type of linear accelerator, called drift-tube linear accelerator. The basic principle is shown in Fig. 5 where hollow tubes are connected to a RF-generator. A charge, here assumed positive is accelerated, if it arrives at the gap between at the correct time. While the electric field reverses polarity the particle is hidden inside the drift tube and it finds again an accelerating field at the next gap, if the length of the drift tube is chosen properly. This type of accelerator is often used in proton linacs and protons at non-relativistic energies still change their speed as they are accelerated and the length of the drift tubes must be matched to the speed and has to get longer as the speed and thereby the energy of the particles increases.

Instead of connecting an RF generator directly to the drift tubes the entire structure can be embedded in a large resonator and fed by microwaves of typically a few 100 MHz. This is particularly attractive for ultra-relativistic particles.

Exercise: calculate the length of the tubes for a RF system of 10 MHz and a particle energy from 10 MeV to 500 MeV if the particles are protons.

1.1.7 Disc-loaded waveguide linac

The idea of directly using microwaves to accelerate particles in free space is impeded by the fact that microwaves always travel at the speed of light, whereas particles always travel at speeds less of that. We therefore have to make the electromagnetic waves, the microwaves travel slower. One way, the most popular

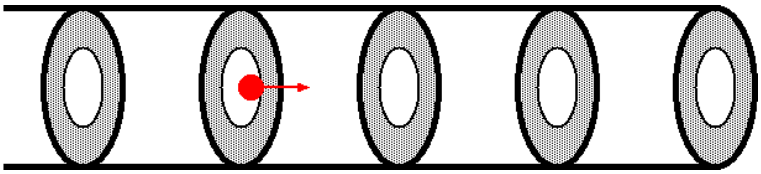

Figure 6: Schematics of a disk-loaded waveguide.

nowadays, to slow down microwaves is to pass them through a disk loaded waveguide which is a vacuum pipe with washer disks included, as is shown schematically in Fig. 6 where the particle moves to the right and the grey washer disks are used to slow down the electro-magnetic wave that is fed into the structure from the left. The 3 km long linac at SLAC, for example, uses these structures with a frequency of about 3 GHz.

1.2 Choosing an Accelerator

The design and construction of particle accelerators are driven by requirements of the experiments that will be done at the respective machines. Here we briefly discuss the possible demands.

1.2.1 Why high energies?

Accelerators for particles of higher and higher energies are built for two reasons. First, resolving small structures of a given size requires a probe or radiation of a wavelength that is of comparable size, or preferably smaller. The de'Broglie wavelength of a particle is inversely proportional to the particle's momentum and therefore high energy particles allow to resolve smaller structures. Second, the energy stored in the accelerated particles can be used to produce new particles of high mass. The higher the beam energy, the heavier the newly generated particles can be.

1.2.2 Electrons versus Protons

In order to reach the highest possible energies protons or heavy ions are advantageous, but these particles have a substructure. The proton, for example is an agglomerate of three quarks accompanied by gluons and the collision between two protons or a proton and an anti-protons is actually a collision between a number of their constituents. The events recorded in the particle detectors are consequently difficult to interpret. Electrons, on the other hand are, as far as we know today, point-like and the events are easier to interpret, but electrons are lighter and emit synchrotron radiation. Furthermore, it is more difficult to reach the highest possible energies.

Often new physics at high energies is explored by proton accelerators and later an electron-positron accelerator is used in precision measurements.

1.2.3 Circular versus Linear

The design of a circular accelerator is driven by the desire to achieve a high collision rate in the detector in order to investigate events with high statistics. The high revolution rate of circular accelerators makes this attractive. In a circular machine, however, a large number of magnets is needed in order to recirculate the beam. For protons the magnet technology actually sets the limit, because magnetic fields in excess of 2 T cannot be reached with electro-magnetic technology and about 10 T is the limit for super-conducting technology which limits the energy if the size of the accelerator is limited. Circular electron machines are limited by the RF-system needed to replenish the emitted synchrotron radiation.

Reaching extremely high energies in electron machines another approach is needed, namely that of a linear collider, which avoids the humongous synchrotron radiation losses, but poses very high demands on the acceleration technology to reach the desired energies within the available space. Another important feature of electron accelerators is that the beam quality is determined by the emission of synchrotron radiation and certain limits can not be circumvented.

Other applications for linear accelerators are high intensity proton linacs in neutron spallation sources, or medical applications. If there are extreme demands on the beam quality, as for example in free-electron lasers, linear accelerators are needed, because the beam quality is defined in the electron gun and transported to the experiment.

1.2.4 Collider versus Fixed Target

In nuclear or high-energy physics experiments high count rates, but also a large energy density in the experiment are of prime concern. In a fixed target experiment the target can be made very thick in order to increase the count rate, but the interaction products will carry a certain amount of kinetic energy, dictated by momentum conservation, which is unavailable for the creation of new particles. In a collider on the other hand, all the kinetic energy of the initial particles is available for the creation of new particles, but the target density is limited by the intensities of the colliding beams and the constantly, at high rate, colliding beams perturb each other up to a limit where an increase in the intensity is detrimental to the count rate.

1.2.5 Synchrotron Light Sources

The purpose of synchrotron light sources is the generation of photons with the high intensities and beam quality, such as spatial and temporal coherence. There are two classes of light sources, those based on rings where the primary sources of

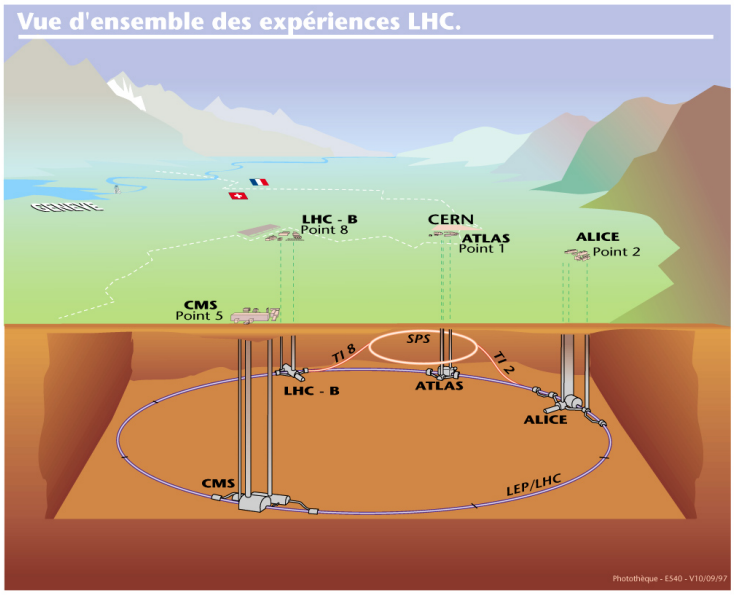

Figure 7: LHC and its detectors. (picture CERN)

synchrotron radiation are the main dipole magnets, but nowadays this is complemented by undulators and wigglers, where an array with sinusoidally alternating magnetic field generates radiation of higher intensity and brightness compared to that emitted in dipoles. The second class of synchrotron radiation sources are linac-based free-electron lasers, where a bunch with extremely high (peak) current and small transverse size is passed through a very long undulator. In the long undulator an instability between the photons and the electrons develops that leads to an exponential growth of the number of photons and therfore intensity. Furthermore the emitted light has a high degree of coherence.

1.3 Examples of Modern Accelerators

In this section we discuss a small group of accelerators in order to illustrate the features and questions mentioned before and how they were answered in a engineering design.

1.3.1 SPEAR at SLAC in Stanford

The 3.5 GeV storage ring SPEAR with a circumference of 270 m was built in the late 1960s as a e^+e^- collider and the first meson containing charm quarks, the J/Ψ particle, was co-discovered in SPEAR. A discovery which was subsequently rewarded the Nobel prize in 1974. As higher-energy e^+e^- colliders became available throughout the 1980 the activities at SPEAR turned towards synchrotron

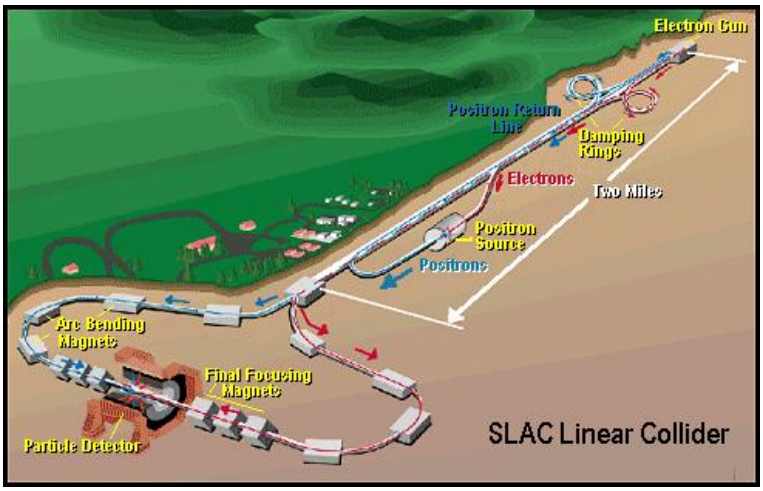

Figure 8: Aerial view of the Slac Linear Collider. (picture SLAC)

radiation research, which used mainly the radiation from the bending dipoles. It was thus one of the early 1st-generation light sources. As undulator and wiggler magnets were added in the straight sections between the dipoles, it morphed into a 2nd-generation light source. During a major rebuild in the early 21th century SPEAR was turned into a modern 3rd-generation synchrotron light source, aptly named SPEAR3.

1.3.2 LEP and LHC at CERN in Geneva

During the 1980s a huge e^+e^- collider name LEP with a circumference of 27 km and a maximum energy of 100 GeV per beam was built at CERN in Geneva with the intent to experimentally investigate the electro-weak theory, especially the Z_0 and W-bosons, which were discovered earlier in the SPS, a proton-antiproton collider at CERN. The physics program in LEP was very successful and experimentally showed, for example, that there are only three families of fundamental particles. Operating LEP at the highest beam energies was limited by the emission of synchrotron radiation. At 100 GeV beam energy about 3 % of the beam energy is lost *every turn* and had to be replenished by a large number of superconducting RF-cavities. Despite the experimental successes LEP was dismantled in 2002 in order to reuse the existing 27 km tunnel to house a collider, called the Large Hadron Collider (LHC) which will bring protons and heavy ions into collisions. The reachable energy is limited by the magnet technology which employs super-conducting magnets cooled by supra-liquid helium at 1.8 K. The peak energy for protons is 7.7 TeV and the more than 1200 dipole magnets with a peak field of 8 T are needed 'to keep the beam in the pipe.' Figure 7 shows the LHC tunnel with the SPS used as injector and the four major experiments ALICE,

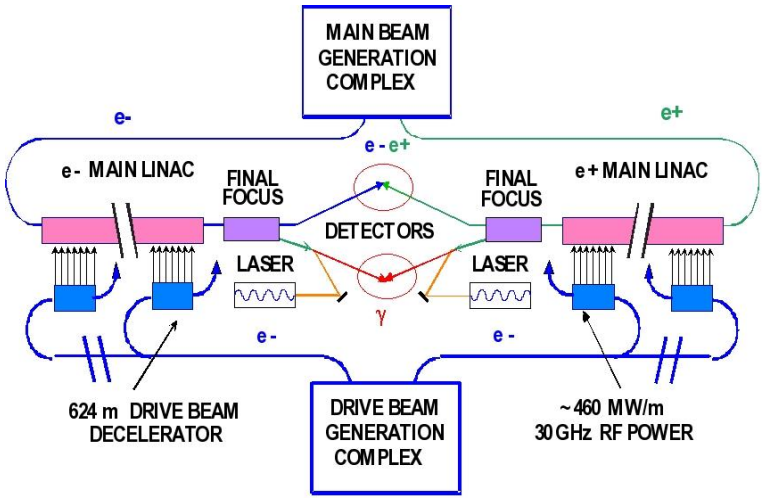

Figure 9: CLIC layout. (picture CERN)

CMS, and LHC-B, and ATLAS (with active Uppsala participation).

1.3.3 SLC at SLAC in Stanford

The improvements in vacuum and microwave technology during the 50s made the generation of high power S-band (3 GHz) possible, which was used to feed the acceleration structures of the 3 km long linear accelerator built at SLAC in Stanford during the 1960s. During the 80s schemes to increase the achievable energy from the initial 20 GeV to 50 GeV were implemented which made the investigation of the Z_0 bosons possible. Furthermore, a positron source, damping rings, and arcs at the end of the linac were added. This led to the revolutionary concept of a linear collider. In order to reach high luminosities the relatively low repetition frequency of 120 Hz needs to be compensated by squeezing the beams to micron size at the collision point which led to a big development effort for suitable beam diagnostic and control. The SLC was in operation until 200?, now part of the linac is used as an injector for the PEP2 B-factory which was the second incarnation of the PEP ring at SLAC. From 2009 part of it will be used in the LCLS free-electron laser.

1.3.4 ILC and CLIC, planned

The experience from LEP showed that reaching 100 GeV in a circular e^+e^- accelerator requires a circumference of 27 km which indicates that the emission of synchrotron radiation will prevent reaching higher energies in a *circular* collider. Going towards TeV energies in en electron machine thus requires a linear accelerator, or, rather two accelerators pointing at each other. In order to keep the length

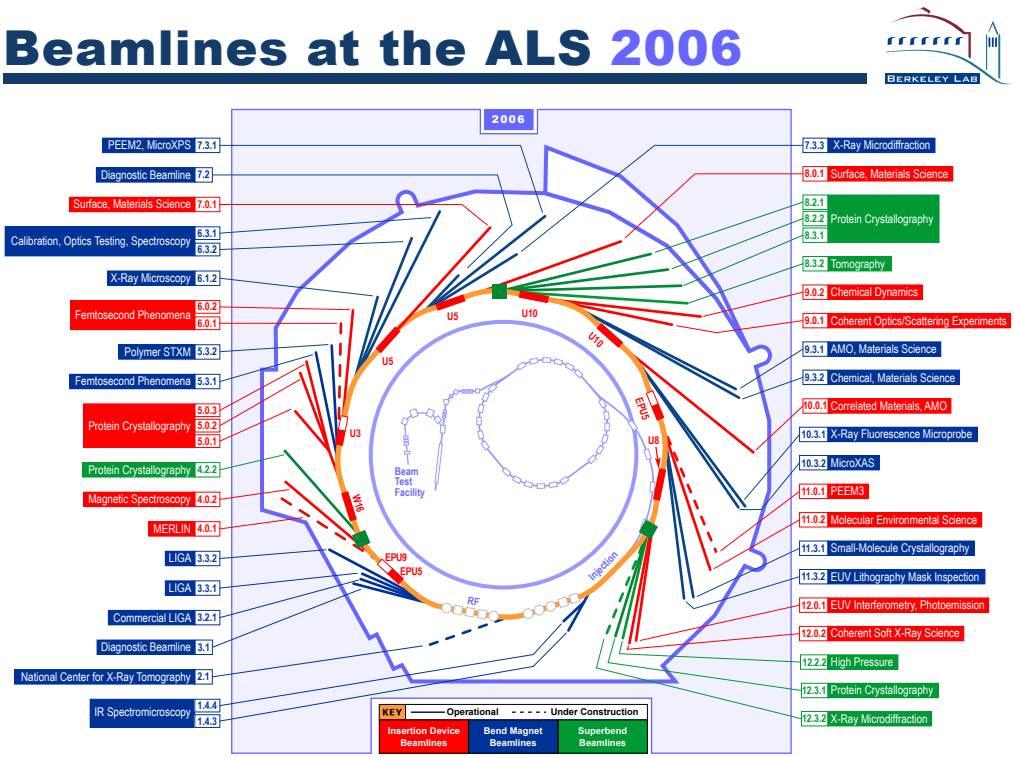

Figure 10: The synchrotron radiation beam lines in the ALS (figure ALS).

of this accelerator within reasonable limits (30 km?) a high accelerating gradient in the RF structures is mandatory. Super-conducting RF structures can reach around 30 MeV/m if powered by 1.3 GHz microwave klystrons. This is the way that the ILC collaboration is going but the peak beam energy is limited to about 0.5 TeV and in order to reach the Multi-TeV another approach, the two-beam acceleration scheme, on which CLIC is based, is needed. Here a low-energy beam at very high intensity is decelerated in specially shaped RF structures and generates 30 GHz microwaves that are used to accelerate the lower-intensity main beam to very high energies. This scheme has shown to reach acceleration gradients in the 150 MeV/m range. Figure 9 shows a schematic of CLIC. The final focus system for both ILC and CLIC is very demanding, because here the intended beam sizes at the collision point will be several orders of magnitude smaller than at SLC, namely nano-meter size beams are required. This will pose extreme constraints on the alignment and beam stabilization system, among other things.

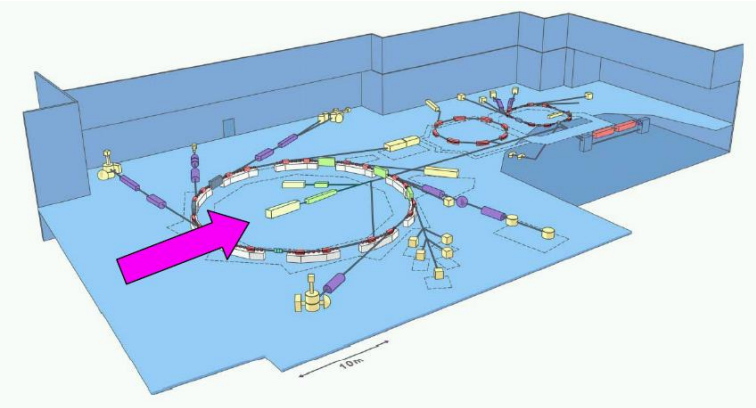

Figure 11: MAX IV layout. (picture from FEL04 conference)

1.3.5 ALS at LBNL in Berkeley

Since 1993 the building that previously housed the 184-inch cyclotron in Berkeley, Ca. is the home of the Advanced Light Source (ALS), a third generation synchrotron light source, i.e. deliberately built and designed to produce synchrotron radiation in 1.9 GeV electron storage ring with a circumference of 2xx m storing up to 400 mA of particles. The ALS is equipped with 28 beam lines that provide user groups with radiation from bending magnets and 15 beam lines from undulators and wigglers. The spectral range of the ALS extends into the X-ray region allowing for example microscopy of biological samples and protein crystallography.

1.3.6 MAXLAB in Lund

In Lund a Swedish national synchrotron light source, called MAXLAB is in operation since the 1980s. The first electron storage ring was MAX I with a maximum energy of 550 MeV. It serves both nuclear physics and synchrotron users. Since 1995 a third generation light source, the 1.5 GeV ring MAX II is operational and serves users with VUV and soft X-ray photons. During 200x the new 700 MeV ring MAX III was taken into operation and it will provide high quality beams and brilliant radiation to users of lower photon energies in the VUV and IR range of the spectrum, liberating the straight sections in MAX II for insertion devices (undulators, wigglers) in the X-ray region. A special feature of MAX III is the integration of support and magnet structure, permitting an extremely compact and inexpensive design. In the future the MAX IV complex is planned, which consists of two rings at 1.5 and 3 GeV on top of each other with a full energy normal-conducting linear accelerator. Included in the design is a cascaded optical klystron free-electron laser.

1.3.7 Other accelerators in Scandinavia

Apart from accelerators for medical applications such as tumor irradiation or isotope production there are a number of accelerators for basic research in Scandinavia. In Jyväskylä, Finland there is a 130 MeV cyclotron that provides a large variety of ions to experimenters in nuclear physics and biology. The ASTRID ring in Århus, Denmark has the unique capability of being able to accelerate both electrons to provide synchrotron radiation and heavy particles such as highly charged ions or even large molecules such as Buckminster-Fullerenes C_{60}. An completely electrostatic accelerator, called ELISE was also conceived at ASTRID. In Stockholm the heavy ion storage ring CRYRING is operated at the Manne Siegbahn Laboratory is equipped with an electron cooler and is mostly used for atomic physics experiments such as dissociative recombination of carbohydrates. In Uppsala the Gustav-Werner cyclotron is in operation since the late 1940s and nowadays provides ions and protons up to a kinetic energy of 180 MeV. Until 2005 the beams were used to fill the 82 m nuclear physics storage ring CELSIUS, but are today used to generate neutrons for material testing and irradiate patients in collaboration with the University hospital. These activities will expand greatly if the Swedish Proton Therapy Center (SPTC) will be built in Uppsala, as is currently discussed.

2 Course Overview

Here we give a short synopsis of the issues that will be discussed and motivate the relevance of the respective concepts. It is intended to whet the appetite of the student to learn more and to give an overview over the material covered. We also discuss prerequisites and expected student participation for the course.

2.1 The Particles

When describing particles in an accelerator we first have to choose a suitable coordinate system to describe the state of a single particle. In an accelerator we typically have to consider a large number (10^6 to 10^{11}) of charged particles, which makes a statistical description suitable.

2.1.1 Their mathematical Description ...

In classical mechanics the state of point-like particles can uniquely be described by their *phase-space coordinates*, which are the positions in x, y, z and the corresponding momenta p_x, p_y, p_z with a suitably chosen origin and zero momenta. In an accelerator we chose a particle that moves on a design trajectory as reference and describe the state of other particles in the beam by their deviation from the

reference particle. Moreover, we use positions and angles of the particle in the direction perpendicular to the reference particle's motion as transverse coordinates and the relative energy difference with as well as the arrival time at a particular position with respect to the reference particle. Of course this only describes a single particle.

The large number of particles in an accelerator warrants the use of statistical methods and we describe an ensembles of many particles by distribution functions in the phase space variables. Sometimes a distribution or a projection can be measured directly, but often averaged quantities that describe a distribution are needed. The most obvious examples are the *moments*, where the zeroth moment describes the total number of particles in the distribution, the first moments describe the center-of-mass quantities, such as average position, and the second moments describe the size of the distribution. Mixed moments or correlations sometimes show physical significance and will be discussed at the appropriate places. Often the moments that are most easily measured experimentally, as is, for example, the beam position.

2.1.2 . . . and Generation

Normal particles used in accelerators are electrons, protons, but also heavy ions, anti-protons, and positrons and they have to be generated in suitable devices to provide the required intensities with beam sizes that are needed in the experiments. Electrons can, for example, be generated in thermionic devices, where a filament is heated and the electrons are extracted by high voltage. Heavy ions can be generated in an electron-cyclotron resonance (ECR) ion source where a plasma is sustained by a microwave and magnetic fields. The positive ions are extracted by high voltage. Positrons are generated by impinging high-energy electrons on a target and separating the positrons in the developing electromagnetic shower of gammas, and electron-positron pairs. Since all particles generated in a given source have the same charge, they repel each other and reaching high intensities normally imposes limits on the achievable beam quality.

In the course we will discuss the generation mechanisms and the engineering of sources for commonly used particles.

2.2 Beam Transport

Once the particles are generated they need to be guided to the experiments by magnetic fields. The dynamics of the particles is determined by the Lorentz-equation and the reference orbit is normally determined by the position and the excitation of dipole magnets.

The beam particles do not necessarily follow the reference trajectory, partially because they all carry the same charge and tend to diverge, partially because of finite alignment tolerances. Furthermore the transverse beam size often needs to

be adjusted to suit experimental demands. This implies the need for focusing forces that kick the particles back towards their reference or design trajectory, similar to what lenses do in optical systems where they allow shaping the transverse beam size of e.g. a laser beam. In a beam transport system the function of lenses is carried by quadrupole magnets and the positioning and excitation of the quadrupoles determines the stability of a beam transport system as well as the transverse beam size throughout the accelerator. The design of simple beam optical systems can be done by analytical means, but more complex systems are analyzed by computer codes like MAD [8] as we will do during the course.

We will discuss the concepts used in beam optics and explore different accelerators using codes such as MAD.

2.3 Magnets

The dipole and quadrupole magnets alluded to before and other magnets need to be designed and built and their parameters can be quickly estimated by analytical means such as Ampere's law. Depending on the accelerator the magnets can be normal- or super-conducting or made of permanent magnets. For a detailed design two- or three-dimensional codes based for example on finite-element methods are required.

In the course we will use the free student version of a commercial electromagnetic design code [9].

2.4 Getting up to speed: Acceleration

The particles in an accelerator are accelerated by radio-frequency (RF) fields and we will discuss the generation of RF-power and how it is applied to the beam in resonating RF-structures which we briefly discuss. At first sight it may be surprising that all particles in an ensemble receive the same accelerating fields which appears as a remarkable feat, but can be explained by the concept of *phase focusing* which keeps all particles in an ensemble, often called bunch, synchronized.

2.5 Where are the they: Diagnostics and Correction

Despite transverse focusing due to quadrupoles and phase focusing due to the RF-system them beam can go badly wrong, especially when starting up a new accelerator. In order to find out what is wrong we need diagnostic devices that tell us the state of the beam in the machine. Especially important are the current, the position and the size of the beam, the moments so-to-speak. We will discuss the devices that measure these quantities and then apply the knowledge about the state of the beam to correct it. As an example one might consider the beam

position, which can be corrected by small extra dipole magnets that are placed in the accelerator.

We will cover the basic beam instrumentation devices and a few correction methods are are frequently encountered in control rooms.

2.6 Oops, too much: Instabilities

In experiments often a large intensity beam is desired, but a large number of particles can become unstable, because the particles start interacting among themselves and create a feedback loop, that can become unstable. One such self-interaction is the space-charge field due to the same charge of all particles, another instability is driven by RF-fields deposited by the leading particles in resonant structures that later parts of the bunch see. In a circular accelerator this may lead to an instability.

We will discuss a small number of mechanisms that lead to instabilities and point at cures.

2.7 Take care: Radiation protection

Depending on the beam energy and intensity an accelerator can be a high-radiation environment and proper care needs to be taken for access and monitoring the radioactive dose to which personnel and equipment is exposed.

We will briefly discuss the units and effects of ionizing radiation and how the radiation is detected.

2.8 Synchrotron Radiation

It is well known that accelerated charges radiate electromagnetic waves, similar to what happens in a normal dipole antenna. Under normal conditions only the light particles electrons or positrons emit radiation. Initially this effect was undesired, because the energy lost needs to be replenished in RF-structures, but later attractive applications of the so-called synchrotron radiation appeared, mostly due to the short wavelength and attractive spectral properties of the radiation. Nowadays electron accelerators dedicated to the generation of synchrotron light are build, where free-electron lasers are the brightest sources.

We discuss the power and spectral characteristics of the synchrotron radiation emitted.

2.9 Prerequisites

Knowledge of the concepts listed in important in order to be able to properly follow the course.

- Classical Mechanics
- Relativistic Kinematics
- Lorenz-force
- Multipole expansion
- Probability distributions
- Linear Algebra
- Ordinary differential equations

2.10 Student Participation

Active student participation in the seminars is a prerequisite for passing the course, because problem solving will deepen the understanding of the subject matter and will enable you to solve other real-world problems if they cross your path.

During the course we will design a fictitious accelerator, called ERPL, which is short for Electron-Ring-Proton-Linac, which should be a collider between a proton-linac in the 200 MeV range and an electron ring with a design energy of 3 GeV. The latter should also be able to provide synchrotron radiation user with radiation. Typical problems will deal with

- Discussion of technical choices for parameters in various subsystems,
- Magnet lattice and beam optics,
- Magnet design,
- Orbit, tune, and chromaticity correction,
- Estimating instability limits

Most, but not all, problems will relate to this accelerator.

Especially in the early phases of the course I will assign the discussion of special subjects to students, which is typically based on a publication that the student presents to the class. This will be similar to a classic seminar. Examples in the first class will be the discussion of

- the cyclotron,
- the betatron,
- and the Cockroft-Walton

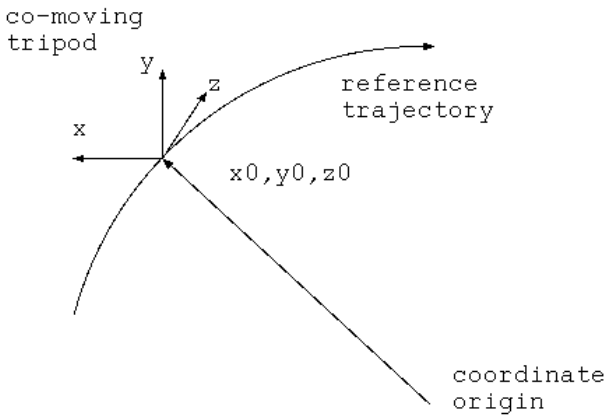

Figure 12: Reference trajectory.

accelerators.

There will be a final exam or a 1/2 hour oral exam depending on the number of participants.

We will visit the CRYRING at the Manne Siegbahn Laboratory in Stockholm.

Having covered the administrative details, we are ready to jump into the subject matter of the course.

3 The Particles and their Description

The particles that are used in accelerators need to be charged, which implies that some normally neutral atom or molecule needs to be ionized in order to be accelerated and guided by electro-magnetic fields. The generation process and the technology of particle sources we postpone to a later chapter and discuss the mathematical description first of a single particle and later an ensemble of many particles in this section. For the purpose of describing the motion of particles in an accelerator, we can treat electron, protons or other ions as essentially point-like, and ignore their internal structure. As mentioned in the introduction are such point-like particles in classical mechanics uniquely described by their phase space coordinates, which consist of their positions and velocities, or equivalently, their momenta.

When we design an accelerator for a particular purpose – such as collider or light source – we envision where the particles that constitute the beam should show up at what time (such as two beams at the same time in the same place in a collider). This basic requirement is then translated into a description of the trajectory, or orbit, of a reference particle that has all the desired properties. Of course such a particle has the right energy at the desired place which means it has to arrive in RF cavities at just the right time to receive the correct energy

increase. Knowing the energy we can place dipole bending magnets in order to guide the reference particle to where we want it to be. Here the magnets are assumed to be ideal and provide just the correct bending angle to guide the reference particle and the distances between the magnet are chosen to correctly determine the arrival time at the desired locations such as the interaction point in a collider. For a given reference trajectory we can find a co-moving coordinate system in the form of a Frenet-tripod that describes the dynamics of the particles that are not quite so perfect as the reference particle in that co-moving coordinate system. It must be noted that the rather heuristic arguments used her can be made rigorous by a proper sequence of canonical transformations to transform the dynamics represented by a Hamiltonian into the co-moving frame. This procedure is, for example, described in detail in Ref. [2].

Rather than giving the the positions and momenta with respect to the reference particle we use the following quantities

x , the horizontal distance to the reference particle;

x' , the horizontal angle with respect to the trajectory of the reference particle;

y , the vertical distance to the reference particle;

y' , the vertical angle with respect to the trajectory of the reference particle;

τ , the arrival time with respect to the reference particle at a fixed point;

δ , the relative momentum difference with respect to the reference particle;

to describe the state of a single particle. It should be noted that some programs such as MAD use slightly different variables, which, however, in the ultra-relativistic limit agree with those mentioned here. An advantage of these quantities is that they describe geometric concepts like distances and angles. The arrival time is relevant if time-varying electro-magnetic fields affect the particle such as RF-structures. The relative momentum difference δ is convenient, because it describes the relative variation of the angles inflicted by magnetic fields with momentum.

Now we have a way to describe a single particle and where it is at a specific time. In a real accelerator we have to deal, however, with a large ensemble of particles, typically on the order of 10^6 to 10^{11} individual particles. If dealing with such a large number of particles that constitute a *beam* we have to use statistical means to describe their multi-variate distribution in the six variables $x, x', y, y', \tau, \delta$. A suitable tool is a distribution function $\psi(x, x', y, y', \tau, \delta)$ which describes the number of particles in an infinitesimal phase-space volume N_{dV}

$$N_{dV} = \psi(x, x', y, y', \tau, \delta) dx dx' dy dy' d\tau d\delta \ . \tag{1}$$

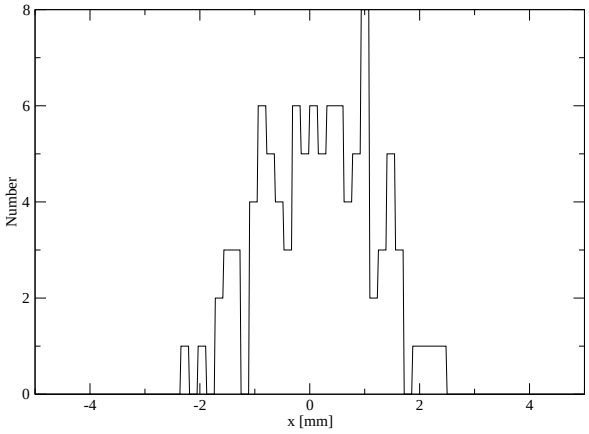

Figure 13: A histogram.

Often projections onto one of two variables are used. Mathematically, they are described by integrating over the unconsidered variables. For example

$$\Psi(x,y) = \int \psi(x,x',y,y',\tau,\delta)dx'dy'd\tau d\delta \tag{2}$$

describes the distribution in the two transverse spatial coordinates x and y which can be measured in an accelerator, by placing a luminescent screen into the path of the beam and recording the image by photographic means. The intensity of the image on the screen is dependent on the number of particles that hit a particular location on the screen. Note, however, that the screen does not discriminate the angles x', y' or the arrival time τ (within limits) or the energy δ of the particles and we therefore sum or integrate over these coordinates. Note that the local intensity on the screen is obtained by summing over all intercepting particles which is the experimental equivalent of creating a histogram. In fig. 13 we show an artificially generated one-dimensional histogram of particle positions.

A distribution function or even a histogram as its experimental approximation contains a lot of information in the sense that many numbers are needed to describe it and a more efficient way is clearly desirable. We can estimate the number of particles represented in the histogram in Fig. 13 by counting the number of particles in all bins. By careful observation we see that the distribution is well centered and has the bulk of particles near $x = 0$. Furthermore it has a width of around $\pm 2\,\mathrm{mm}$. By characterizing distribution by the total number of particles, its center and its width we effectively describe it by its *moments*. The zero-th moment is the total number of particles which can be determined by integrating a distribution function over all space, the first moment $\langle x \rangle$ determines

the center-of-mass and can be calculated by evaluating the integral

$$\langle x^n \rangle = \int x^n \psi(x) dx \tag{3}$$

for $n = 1$. Here we assume here that the distribution function $\psi(x)$ only depends on the variable x. If it depends on other variables as well, the integral has to extend over all other variables as well. The second moment $\langle x^2 \rangle$ is related to the rms width of the distribution σ by

$$\sigma^2 = \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2 \ . \tag{4}$$

Note that often the moments or the width are the quantities that can most easily determined experimentally. Take for example the screen image discussed in the previous paragraphs. By turning the picture into a two-dimensional histogram, calculate the integrals numerically by observing that the integral can be rewritten as a sum and the integration measure dx is the bin width of the histogram. Thus we can determine the moments and thereby the center and width of the distribution. If we denote averaging over the distribution function by angle brackets

$$\langle g() \rangle = \int g() \psi() dV \tag{5}$$

where the empty brackets can filled by any number of suitable arguments we can calculate averages over any function.

Since the particle distributions are often determined by the repeated application of random forces (which are often balanced by damping) the central limit theorem implies that the resulting distribution function is Gaussian. In one dimension a normalized and centered Gaussian is given by

$$\phi(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-x^2/2\sigma^2\right) \tag{6}$$

and the multi-variable version in n dimensions with $\vec{x} = (x_1, \ldots, x_n)$ is given by

$$\phi(\vec{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \sigma}} \exp\left(-\frac{1}{2} \sum_{i,j=1}^{n} (\sigma^{-1})_{ij} x_i x_j\right) \tag{7}$$

where we have introduced the symmetric sigma-matrix σ that is given by

$$\sigma_{ij} = \int x_i x_j \phi(\vec{x}) d^n x \ . \tag{8}$$

It is possible with some effort to prove that eq. 8 is indeed correct by explicitely evaluating the integral at least for the case $n = 2$.

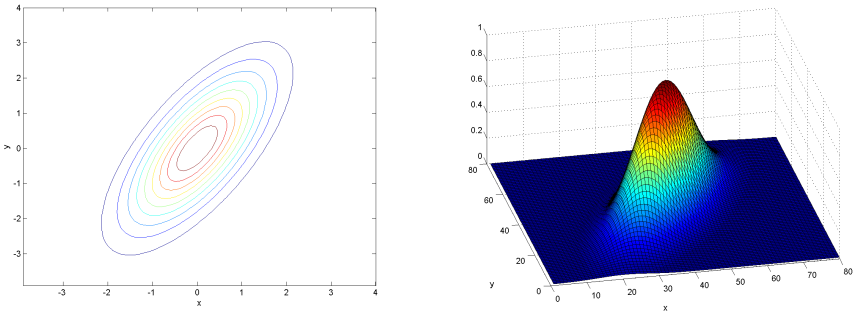

Figure 14: A two dimensional Gaussian distribution.

An example of a two-dimensional distribution is in order now to illustrate the concept of correlation. In Fig. 14 we display a contour plot and a surface plot of a Gaussian distribution with the sigma matrix

$$\sigma = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} \langle x^2 \rangle & \langle xy \rangle \\ \langle xy \rangle & \langle y^2 \rangle \end{pmatrix} \tag{9}$$

which exhibits a strong correlation between the two directions, as described by the off-diagonal elements of the sigma-matrix. The definition of the sigma-matrix in terms of second moments in Eq. 9 is possible, because the distribution is centered and the first moments $\langle x \rangle = \langle y \rangle$ are zero.

We now have a found means to describe several aspects of the particles and beams that move through the accelerator, namely

- individual particles,
- distributions of particles,
- and moments of distributions.

It is possible to mathematically describe how the three representations of the state of the beam changes as it passes through the accelerator. The dynamics of the individual particles can be described by Hamiltonian mechanics, the distribution function by the Vlasov- or Fokker-Planck equations and the dynamics of the moments can be described by maps, which in the simplest case reduce to matrix operations, as we will see in the next sections.

4 Transverse Beam Optics

In the previous section we found that the information about the first and second moments gives a fairly well description of most distributions that are reasonably

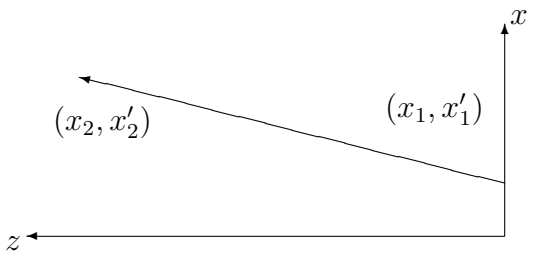

Figure 15: Particle trajectory in a drift space.

well-behaved. Here we will describe methods of how to find out how the moments change from one place in the beam line to the next. We restrict ourselves to the transverse degrees of freedom x, x', y, y'.

4.1 Drift

We start by considering the simplest elements in a beam line and describe their influence on the motion of a single particle and deduce the maps for the moments from that. We start by considering what happens in a drift space which is a fancy word for an empty beam pipe without magnetic or electric fields. Obviously the particles travel on a straight line through a drift space and the transverse coordinates x_2, x_2' at the end of the drift space of length L are related to those at the beginningx_1, x_1' by

$$\begin{aligned} x_2 &= x_1 + Lx_1' \\ x_2' &= x_1' \end{aligned} \tag{10}$$

which is easy to understand when looking at Fig. 15. The particle comes from the right and travels towards the left. Initially is has a distance x_1 to the axis and moves away from the axis with an angle x_1'. During the passage the angle does not change, but the distance to the $z-$axis increases, according to the first of Eq. 10. We observe that the final coordinates with subscript 2 are linear combinations of the initial coordinates with subscript 1 and we can consequently write Eq. 10 in matrix form

$$\begin{pmatrix} x_2 \\ x_2' \end{pmatrix} = \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_1' \end{pmatrix} . \tag{11}$$

It is easy to verify that the matrix for two consecutive drift spaces of length L_1 and L_2 can be obtained by multiplying the matrix for $L1$ with that of L_2.

Observe that in all pictures the particles propagate from the right to the left which makes writing down the equivalent matrix equations easier, because matrices are usually multiplied from the left to a column vector that represents the particle.

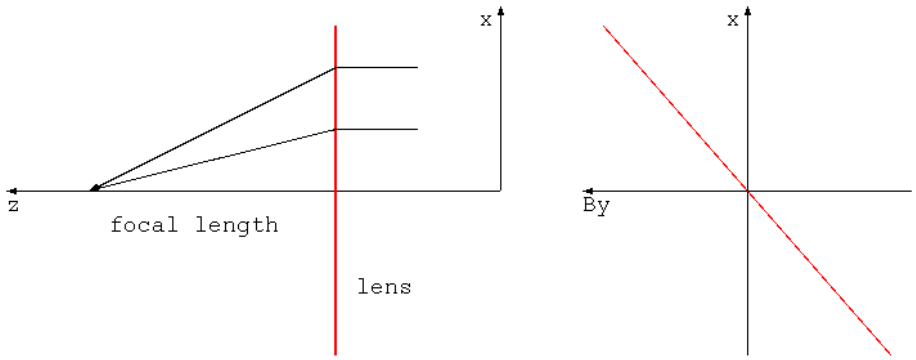

Figure 16: Thin quad.

4.2 Thin Quadrupole

The next element is a thin quadrupole which behaves in the same way as a thin lens known from light optics. We will discuss real quadrupoles later, but fill for the time being focus on thin quadrupoles, because they are very useful for crude estimates of beam lines. The defining property of a thin lens is that it applies a kick (change in angle x') to the particle that is proportional to the distance from the center of the lens or the quadrupole. This behavior is illustrated in Fig. 16 where two particles coming from the right traveling parallel to the axis receive a downward kick proportional to their distance from the axis. This causes both particles to cross the axis at a downstream location. The distance from the lens to the crossing point is, of course, the focal length of the lens. On the right of Fig. 16 the linearly increasing magnetic field that causes the particle deflection is shown.

We now have the task to build a matrix that reflects the behavior we just described, namely a kick proportional to the distance to the axis. It is easy to verify that the matrix

$$\begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix} , \tag{12}$$

where f is the focal length, has this property by calculating the composite matrix of a thin quadrupole with focal length f and a drift space of length f and applying it to an incoming particle with state $(x_1, x_1' = 0)$. The resulting vector should have the form $(x_2 = 0, x_2')$.

Similar to light lenses there are both focusing and defocusing lenses and they differ by the sign of the focal length. A defocusing lens thus has negative sign which is intuitively satisfying, because the intersection with axis lies before then lens (at negative z) for a defocusing lens and after the lens (at positive z) for a focusing lens.

Optical lenses are often round and the focal length in the horizontal and

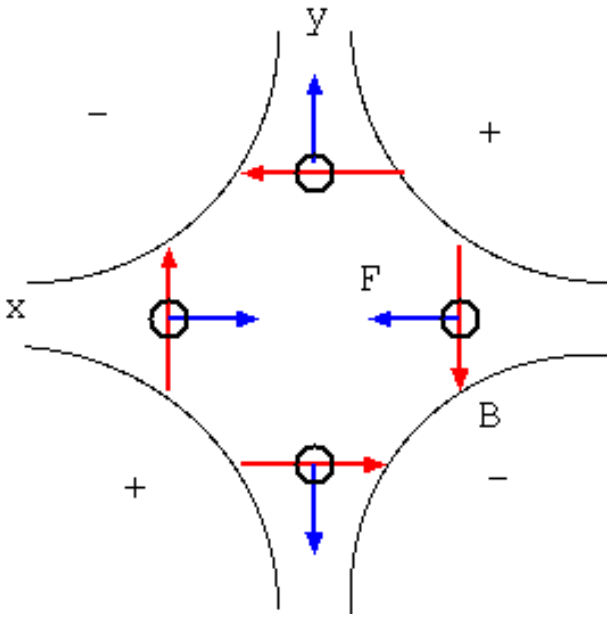

Figure 17: Forces in a quadrupole, if a positively charged particle moves into the plane.

vertical plane are equal, even though there are examples, such as cylindrical lenses that generate a line focus. In magnetic quadrupole lenses the deflection is generated by the magnetic field which has to obey Maxwells equations, especially $\nabla \times \vec{B} = 0$ especially inside the quadrupole we have $dB_y/dx = dB_x/dy$. This causes a magnetic field that rises linearly along the positive $x-$axis to decrease along the positive $y-$axis. Another way of visualizing this behavior is by looking at the field lines (red) in a quadrupole in Fig. 17 and calculate the Lorenz force (blue). A particle on the horizontal axis is kicked towards the center of the quadrupole (focusing), whereas a particle on the vertical axis is deflected away from the quadrupole center (defocusing). In summary, a quadrupole that focuses in one plane defocuses in the other plane. The 4×4 matrix for the two transverse planes that reflects this is given by

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ -\frac{1}{f} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \frac{1}{f} & 1 \end{pmatrix} . \tag{13}$$

The matrix operates on a vector $\vec{x} = (x, x', y, y')^T$. Note that the 4×4 matrix in Eq. 13 contains two 2×2 blocks along the diagonal and 2×2 zero-matrices in the off-diagonal positions. The zeros in the off-diagonals imply that a normal quadrupole does not couple the horizontal motion in x and the vertical motion

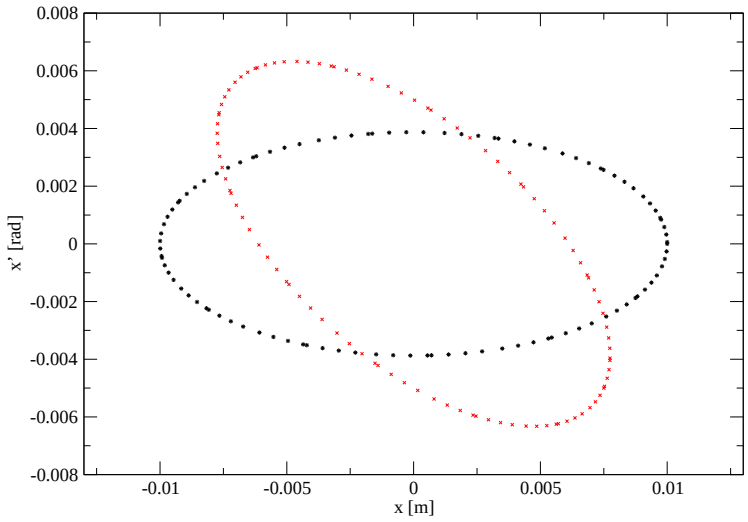

Figure 18: Phase space portrait of the motion in the FODO sequence defined in Eq. 14.

in y.

We now have the basic ingredients to design simple beam lines in a rough way to investigate their general properties. We will later return to other beam line elements such as real quadrupoles, dipoles, and RF cavities.

4.3 FODO Lattice

One of the simplest sequence of magnets is a so-called FODO structure which consists of a sequence of quadrupoles and drift spaces. It can be schematically depicted by

$$(QF/2)\ (L)\ (QD)\ (L)\ (QF/2) \tag{14}$$

where (QD) and (L) denote the matrix of a defocusing quadrupole and a drift of length (L), respectively. The focusing quadrupole is split in half to make the beam line symmetric. If we assume the following values $L = 1\,\mathrm{m}$ for the drift space and $f = \pm 1\,\mathrm{m}$ for the focal length of the quadrupoles we obtain the following matrix M for the segment in Eq. 14

$$M = \begin{pmatrix} 1 - \frac{L^2}{8f^2} & L + \frac{L^2}{4f} \\ -\frac{1}{4f^2} + \frac{L^2}{16f^3} & 1 - \frac{L^2}{8f^2} \end{pmatrix} . \tag{15}$$

Note that we have chosen alternating focusing and defocusing magnets, because in the other plane the sequence were reversed, i.e. the horizontally focusing

magnets become defocusing and vice-versa. In this way the horizontal x-plane and vertical y-plane are effectively exchanged.

We can now visualize the motion of the particles by repeated application of the matrix M on initial conditions $x = 0.01\,\mathrm{m}$ and $x' = 0$ and iterate 100 times. If we plot x' versus x we create a phase portrait of the motion. An example is displayed in Fig. 18, where we plot $x - x'$ at the starting point in black stars and in the center of the first drift space as red crosses. We observe that the motion in both cases is confined to an ellipse, indicating that the dynamics of the particle is similar to that of a harmonic oscillator. Note also that the area of the ellipses appears to be equal which indicates the presence of a conserved quantity or an integral of motion as is the energy in a harmonic oscillator.

4.4 The Lattice Beta-function

The presence of a conserved quantity could be intuitively expected, because the transfer matrices for drift and quadrupole have unit determinant and therefore the product of these matrices also has unit determinant. This property can be exploited to find a useful parameterization of the transfer matrix M which is a 2×2 matrix with unit determinant. It therefore has three independent parameters. Since the motion resembles that of a harmonic oscillator we require that in the center of the parameterization we have a rotation matrix with rotation angle μ. This is the first parameter. Since x and x' have units of meter and radians we have to scale them to equalize the units a quantity with units $\mathrm{m}^{1/2}$. This scaling parameter we call β and has the units of meter. This is the second parameter. Finally we introduce a unit-less parameter α to account for the third parameter needed and arrive at this parameterization

$$\begin{aligned} M &= \begin{pmatrix} \sqrt{\beta} & 0 \\ -\frac{\alpha}{\sqrt{\beta}} & \frac{1}{\sqrt{\beta}} \end{pmatrix} \begin{pmatrix} \cos\mu & \sin\mu \\ -\sin\mu & \cos\mu \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{\beta}} & 0 \\ \frac{\alpha}{\sqrt{\beta}} & \sqrt{\beta} \end{pmatrix} \\ &= \begin{pmatrix} \cos\mu + \alpha\sin\mu & \beta\sin\mu \\ -\frac{1+\alpha^2}{\beta}\sin\mu & \cos\mu - \alpha\sin\mu \end{pmatrix} . \end{aligned} \tag{16}$$

The parameterization shown in the first line of Eq. 16 has a simple interpretation. First we apply a coordinate transformation by the matrix with α and β, then we apply a rotation, followed by the inverse coordinate transformation. The β-matrix is thus just an affine transformation that rescales and changes the angle of the coordinate axis. Since the motion in the new coordinate system is a circle it is common to call this 'transforming into normalized phase space.'

The rotation angle μ is commonly called the phase advance of the oscillations that the particle performs and β is the ubiquitous beta-function. We can now exploit the parameterization from Eq. 16 to determine the phase advance, β, and α for the FODO beam line (or lattice) in Eq. 14 with transfer-matrix given by

Eq. 15 and obtain the following equations

$$\begin{pmatrix} \cos\mu + \alpha\sin\mu & \beta\sin\mu \\ -\frac{1+\alpha^2}{\beta}\sin\mu & \cos\mu - \alpha\sin\mu \end{pmatrix} = \begin{pmatrix} 1 - \frac{L^2}{8f^2} & L + \frac{L^2}{4f} \\ -\frac{1}{4f^2} + \frac{L^2}{16f^3} & 1 - \frac{L^2}{8f^2} \end{pmatrix} . \tag{17}$$

Adding the diagonal elements and dividing by 2 we find

$$\cos\mu = 1 - \frac{L^2}{8f^2} \tag{18}$$

and using the trigonometric equation $1 - \cos\mu = 2\sin^2(\mu/2)$ we obtain

$$\sin(\mu/2) = \frac{L}{4f} . \tag{19}$$

From the 12-matrix element in Eq. 17 we can determine the beta function

$$\beta = L\frac{1 + L/4f}{\sin\mu} = L\frac{1 + \sin(\mu/2)}{\sin\mu} \tag{20}$$

where we assume that we have calculated μ already in Eq. 19. Finally, from the difference of the diagonal elements we determine

$$\alpha = 0 . \tag{21}$$

Note that here we have performed the explicit calculations to determine the parameters μ, β, and α from the matrix elements for a matrix explicitely given by Eq. 15. The same procedure, however, can be done for any matrix M whose elements are given, for example, in numeric form.

Now we need to discuss the assumptions that went into the choice of parameterization in Eq. 16. First note that the beta function and phase advance are determined by the magnet lattice, i.e. the positioning and excitation of the quadrupoles. No information about the particle- or beam-distribution entered into the discussion. Therefore the beta function we talk about here is that of the lattice. later we will encounter another beta function, that is related to the beam distribution. In some cases these two beta functions are equal, normally in circular accelerators, but they need not be equal.

Second, we required that there is a rotation matrix at the bottom because the phase space in Fig 18 looks like that of a harmonic oscillator. We chose the beta-matrices to the left and right of the rotation matrix to be inverses of each other. This choice implies that we tacitly assume that the transfer matrix describes a subsection of the accelerator that has *periodic boundary conditions.* This is often the case, because accelerators are build of a number of similar modules, such as the FODO cell discussed above. By choosing equal beta matrices on the left and right guaranteed that there is the same number of parameters in the parameterization as there are free parameters in the original 2×2-matrix M.

The restriction to three parameters is not really required. If we choose the following parameterization of an arbitrary transfer matrix R

$$R = \begin{pmatrix} \sqrt{\beta_2} & 0 \\ -\frac{\alpha_2}{\sqrt{\beta_2}} & \frac{1}{\sqrt{\beta_2}} \end{pmatrix} \begin{pmatrix} \cos\mu & \sin\mu \\ -\sin\mu & \cos\mu \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{\beta_1}} & 0 \\ \frac{\alpha_1}{\sqrt{\beta_1}} & \sqrt{\beta_1} \end{pmatrix} \tag{22}$$

with different α, β on either side of the rotation matrix we can no longer uniquely determine the five parameters $\alpha_1, \beta_1, \alpha_2, \beta_2$, and μ from the transfer matrix R alone. In practical examples often the beta functions α_1, β_1 at the start of a beam line are given from someone who worked on the section of beam line before the one with transfer matrix R and we are left with three parameters α_2, β_2, and μ that we need to determine from the three independent transfer matrix elements in R.

Of course, we can reverse the interpretation and argue that the transfer matrix R propagates the beta functions from the start point α_1, β_1 to the end point α_2, β_2. This transformation is done by the matrix

$$\begin{pmatrix} \beta_2 \\ \alpha_2 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} R_{11}^2 & -2R_{11}R_{12} & R_{12}^2 \\ -R_{11}R_{21} & R_{11}R_{22} + R_{12}R_{21} & -R_{12}R_{22} \\ R_{21}^2 & -2R_{21}R_{22} & R_{22}^2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \alpha_1 \\ \gamma_1 \end{pmatrix} \tag{23}$$

with $\gamma = (1 + \alpha^2)/\beta$ and R_{ij} are the matrix elements of the transfer matrix R. The correctness of Eq. 23 can be established by inserting the parameterization of R from Eq. 22 simplifying the expression.

The parameterization in terms of beta functions and phase advance is very widely used and also very convenient. If we assume that we have calculated a table of beta functions and phase advances for a beam line we can immediately calculate for example the sensitivity of the particle orbit x at a position labeled 2 to dipole-errors where the beam gets and small kick θ at another position labeled 1. The 12-element of the matrix in Eq. 22 provides just this correlation $x = R_{12}\theta$ with

$$R_{12} = \sqrt{\beta_1\beta_2}\sin\mu \ . \tag{24}$$

We immediately see that the amplitude of the particle oscillation is determined by the beta function and that it propagates in sinusoidally with phase advance μ. Obviously this is a quasi-harmonic oscillation with varying amplitude.

4.5 Tune and Stability

If we consider a circular accelerator that consists of elements such as drifts, quadrupoles and other that can be described by transfer matrices, we can calculate the so-called full-turn matrix which maps the coordinates of a particle at the starting point of the matrix to those after one turn through the accelerator.

Since the start- and end-point of the map are the same the phase ellipses before and after the turn are identical this implies that the beta functions at the start- and end-point must be identical. We can consequently use the parameterization in Eq. 16 for both the horizontal and the vertical 2×2 full-turn matrices [1]. We have seen in the above example that the phase advance μ and beta functions can be determined from the transfer matrix if the beta functions at the start and end-point are equal. In the case that we consider a full-turn matrix the phase advance μ that appears in the argument of the trigonometric functions is usually denoted by

$$\mu = 2\pi Q \tag{25}$$

where Q is called the *tune* of the circular accelerator. There is usually a tune for the horizontal and the vertical motion, conventionally denoted by Q_x and Q_y, respectively. Sometimes the tunes are also denoted by the symbols $\nu_{x/y}$.

The tune has a very simple interpretation: it is the number of oscillations that a particle performs during one turn in the circular accelerator.

Given the full-turn matrix we saw in the previous section that we can find the phase advance μ and thereby the tune $Q = \mu/2\pi$ by calculating the sum of the diagonal elements – the trace – of the full-turn transfer matrix, as should be obvious from the second line of Eq. 16 and we obtain

$$2\cos(2\pi Q) = M_{11} + M_{22} = \mathrm{Tr}(M) \ . \tag{26}$$

Observe that in this way we only obtain the fractional part of the tune due to the periodicity or multi-valuedness of the cosine function.

Obviously we can only calculate the tune in argument of the cosine if the trace of the matrix has magnitude less than 2, because the cosine on the left-hand-side is limited between ± 1 the right hand side can not exceed ± 2. This constitutes a *stability criterion* for stable oscillations in a circular accelerator.

Of course we can physically build an accelerator where we arrange the magnets in such a way that the trace of the transfer matrix exceed 2, but in that case we cannot use the parameterization in Eq. 16 and find beta-functions and tunes.

4.6 Courant-Snyder Invariant or Action variable

In the discussion in the previous sections the beta function was entirely determined by the transfer matrix. No information about the particle coordinates was used in the discussion. We know, however, that the particle coordinates are located on an ellipse in phase space. The coordinate transformation that appears as the beta-matrix in Eq. 16 in the far right converts the real particle coordinates

[1] provided there is no $XY-$coupling between the transverse planes.

(x, x') to those in normalized phase space where the motion is described by circles

$$\begin{pmatrix} \tilde{x} \\ \tilde{x}' \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{\beta}} & 0 \\ \frac{\alpha}{\sqrt{\beta}} & \sqrt{\beta} \end{pmatrix} \begin{pmatrix} x \\ x' \end{pmatrix} . \tag{27}$$

In normalized phase space the phase portrait of different particles is described by circles and the motion is described by a phase advance that makes the particle hop from point to point on the circle determined by the phase advance μ. Particles with different initial coordinates in real space will determine circles with different radii and we can characterize these circles by their radius that is historically denoted by $2J$,

$$2J = \tilde{x}^2 + \tilde{x}'^2 \tag{28}$$

where J is the Courant-Snyder or action variable. It is often somewhat incorrectly called emittance. We will use the term Courant-Snyder or action invariant for the quantity that describes the amplitude of a single particle and emittance when talking about the average of the action over an ensemble of particles.

The circle in normalized phase space given by Eq. 28 that is characterized by $2J$ can be expressed in real space coordinates (x, x') by using Eq. 27. A little algebra yields

$$\frac{1+\alpha^2}{\beta} x^2 + 2\alpha x x' + \beta x'^2 = 2J \tag{29}$$

which is clearly the description of an ellipse as a quadratic equation as should, because it is a conic section.

So far we dealt with the beam line, its magnet lattice and the description of the motion of a single particle. Of course it is possible to describe the motion of a large ensemble of particles through their individual constituents, but that is very inefficient. Using moments is much better and that is what we turn to in the next section.

4.7 The Sigma- or Beam-matrix

In section 3 we found that well-behaved distribution functions can be efficiently described by their first few moments, namely zeroth moment, the particle number; the first moments, the centroids; and the second moments, the beam sizes. Here we discuss how the moments of the beam distribution propagate through a beam line. Once we can do this we have reasonably complete information about the behavior of the beam everywhere in the accelerator.

We start by considering how a single particle propagates due to the effect of a transfer matrix R and then calculate the moments after the beam line represented by R by averaging the final coordinates over the initial distribution. To make this approach clearer we work this out in detail. We describe the initial vector

by $\vec{x} = (x_1, \cdots, x_n)$ where n can be any number, but] most often it will be 2, 4, or 6. The individual particle is assumed to propagate according to

$$\bar{x}_i = \sum_{j=1}^{n} R_{ij} x_j \tag{30}$$

if written in component form. We denote the particle coordinates at the end of the beam line that is represented by R with a bar. For averaging over the initial distribution function we use the Eq. 5 on page 21 and also use the notation that angle brackets denote averaging. Clearly averaging Eq. 30 we obtain

$$\begin{aligned} \bar{X}_i &= \langle \bar{x}_i \rangle = \langle \sum_{j=1}^{n} R_{ij} x_j \rangle = \sum_{j=1}^{n} R_{ij} \langle x_j \rangle \\ &= \sum_{j=1}^{n} R_{ij} X_j \end{aligned} \tag{31}$$

which states that the centroids X_i propagate in the same way the individual particles do, which is convenient, because we can use the single particle dynamics to describe the behavior of a large ensemble of particles. In particular beam position monitors that we will discuss further below in section 7 are sensitive to the centroid of the beam motion, but we can model these measurements with single particle dynamics.

We now turn to the second moments and how they propagate in a beam line defined by transfer matrix R. The sigma matrix is in general defined as the *central second moments* of the distribution. 'Central' in this context means that the centroid motion is subtracted. The sigma matrix is then defined as

$$\sigma_{ij} = \langle (x_i - X_i)(x_j - X_j) \rangle \ . \tag{32}$$

In the remainder of the present section we will for the sake of simplifying the equations assume that the centroid of the distribution is located on the beam axis, i.e. $X_i = 0$. The sigma matrix at the end of the beam line $\bar{\sigma}$ is then given by

$$\begin{aligned} \bar{\sigma}_{ij} &= \langle \bar{x}_i \bar{x}_j \rangle = \langle \sum_{k=1}^{n} R_{ik} x_k \sum_{l=1}^{n} R_{jl} x_l \rangle = \sum_{k=1}^{n} R_{ik} \sum_{l=1}^{n} R_{jl} \langle x_k x_l \rangle \\ &= \sum_{k=1}^{n} \sum_{l=1}^{n} R_{ik} R_{jl} \sigma_{kl} \end{aligned} \tag{33}$$

in terms of the initial sigma matrix σ and the transfer matrix R. Note that Eq. 33 is given in component form. Written explicitely in matrix form we obtain

$$\bar{\sigma} = R \sigma R^T \tag{34}$$

where R^T denotes the transpose of matrix R. Strictly speaking we have only shown that the sigma matrix propagates with Eq. 34 if $\vec{X} = 0$ but with a little more effort it is straightforward to show that Eq. 34 also holds if the full definition of the sigma matrix in Eq. 32 is used.

Equation 34 now enables us to propagate a beam through a beam line, if we know all transfer matrices along the way. This method is in fact implemented in many beam transport codes, starting from TRANSPORT [12], MAD [8] and many others. I want to stress the relevance of the sigma matrix, because it carries all the information about the beam properties such as beam size $\sigma_{11} = \sigma_x^2$ or angular divergence $\sigma_{22} = \sigma_{x'}^2$ throughout the accelerator.

4.8 Emittance and Beam Beta-function

In Eq. 34 we see how the sigma matrix that contains the information of the beam sizes propagates through the beam line. Previously we also realized that the transfer matrices R all have unit determinant and calculating the determinant of both sides of Eq. 34

$$\det \bar{\sigma} = \det(R\sigma R^T) = \det(R)\det(\sigma)\det(R^T) = \det \sigma \tag{35}$$

we observe that the determinant of the sigma matrix remains constant as the beam propagates through the beam line. This is a property of a sigma-matrix of any dimension, provided that the transfer matrix has unit determinant, as is the case for conservative motion, i.e. where the motion can be derived from a Hamiltonian, where the transfer matrices are symplectic.

In the following we will, however restrict ourselves to the horizontal or vertical motion, for example x, x' or y, y'. Here the conservation of a quantity should not come as a surprise, because we already saw that the Courant-Snyder invariant is a quantity that is preserved for the individual particle as a particle propagates down a beam line. That a similar quantity is preserved for a distribution that consists of many individual particles is only natural. The quantity related to the sigma-matrix that stays constant is commonly called the *emittance* ε and is defined by

$$\varepsilon^2 = \det \sigma \ . \tag{36}$$

The square in the definition appears to be consistent with the commonly used definition of the emittance.

Now we will proceed to find a convenient parameterization of the sigma-matrix. Note that the 2×2 sigma-matrix is symmetric and therefore has three independent parameters. The existence of a conserved quantity, the emittance ε, implies that there are two more independent parameters. As we will verify in a moment, a convenient parameterization of the sigma-matrix is

$$\sigma = \varepsilon \begin{pmatrix} \beta & -\alpha \\ -\alpha & \gamma \end{pmatrix} \tag{37}$$

with the already encountered definition $\gamma = (1+\alpha^2)/\beta$ which guarantees that the determinant of the matrix in Eq. 37 is unity. At first sight it may appear arbitrary to choose the parameters α and β for the parameterization of the sigma matrix, because we had already used them in the parameterization of the transfer matrix R in Eq. 22. The usefulness is, however, immediately established if we propagate the parameterization of the sigma-matrix from Eq. 37 using the parameterization of the transfer-matrix R from Eq. 22

$$\begin{aligned}
R\sigma(1)R^T &= \begin{pmatrix} \sqrt{\beta_2} & 0 \\ -\frac{\alpha_2}{\sqrt{\beta_2}} & \frac{1}{\sqrt{\beta_2}} \end{pmatrix} \begin{pmatrix} \cos\mu & \sin\mu \\ -\sin\mu & \cos\mu \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{\beta_1}} & 0 \\ \frac{\alpha_1}{\sqrt{\beta_1}} & \sqrt{\beta_1} \end{pmatrix} \\
&\quad \varepsilon \begin{pmatrix} \beta_1 & -\alpha_1 \\ -\alpha_1 & \gamma_1 \end{pmatrix} \\
&\quad \left[\begin{pmatrix} \sqrt{\beta_2} & 0 \\ -\frac{\alpha_2}{\sqrt{\beta_2}} & \frac{1}{\sqrt{\beta_2}} \end{pmatrix} \begin{pmatrix} \cos\mu & \sin\mu \\ -\sin\mu & \cos\mu \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{\beta_1}} & 0 \\ \frac{\alpha_1}{\sqrt{\beta_1}} & \sqrt{\beta_1} \end{pmatrix} \right]^T \\
&= \varepsilon \begin{pmatrix} \sqrt{\beta_2} & 0 \\ -\frac{\alpha_2}{\sqrt{\beta_2}} & \frac{1}{\sqrt{\beta_2}} \end{pmatrix} \begin{pmatrix} \cos\mu & \sin\mu \\ -\sin\mu & \cos\mu \end{pmatrix} \\
&\quad \times \begin{pmatrix} \cos\mu & \sin\mu \\ -\sin\mu & \cos\mu \end{pmatrix}^T \begin{pmatrix} \sqrt{\beta_2} & -\frac{\alpha_2}{\sqrt{\beta_2}} \\ 0 & \frac{1}{\sqrt{\beta_2}} \end{pmatrix} \\
&= \varepsilon \begin{pmatrix} \sqrt{\beta_2} & 0 \\ -\frac{\alpha_2}{\sqrt{\beta_2}} & \frac{1}{\sqrt{\beta_2}} \end{pmatrix} \begin{pmatrix} \sqrt{\beta_2} & -\frac{\alpha_2}{\sqrt{\beta_2}} \\ 0 & \frac{1}{\sqrt{\beta_2}} \end{pmatrix} \\
&= \varepsilon \begin{pmatrix} \beta_2 & -\alpha_2 \\ -\alpha_2 & \gamma_2 \end{pmatrix} \\
&= \sigma(2) \ .
\end{aligned} \tag{38}$$

Here we observe how well the parameterization of the transfer-matrix R in Eq. 22 works together with the parameterization of the sigma-matrix for Eq. 37. We see that R just transforms the sigma-matrix with beta-functions subscripted with 1 to those subscripted with 2.

From the parameterization in Eq. 37 we also find one of the interpretations of the emittance and beta function. Remember that the 11-element of the sigma matrix is the square of the rms beam size which is often denoted by σ_x^2, but we can also write it as

$$\sigma_{11} = \sigma_x^2 = \varepsilon\beta \tag{39}$$

or by taking the root we get the well-known expression

$$\sigma_x = \sqrt{\varepsilon\beta} \tag{40}$$

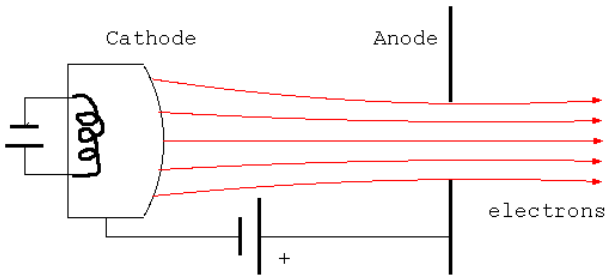

Figure 19: Schematic of a thermionic electron gun.

where the beam size is conveniently described in a way that describes a constant part, the emittance, that is unchanged in a beam line and therefore describes the scale or magnitude of the beam size. The second part, the beta function, that describes the variation of the beam size and is determined by the excitation and placement of the magnetic elements.

4.9 Emittance Growth and Equilibrium

In this section we learned that the emittance stays constant in a beam line with static magnetic fields, but the question remains what determines the value of the emittance to start with. The primary determining factor is the particle source that we discuss in a later chapter and just name a thermionic electron gun as an example. In a thermionic gun the electrons are generated by an electrically heated cathode and extracted by a positive voltage as is schematically shown in Fif. 19. The transverse beam properties of the electrons at the right hand side is defined by the temperature of the cathode and the geometry of the gun, such as the shape of the cathode and anode and possibly other control electrodes that are not shown in Fig. 19. Another strong influence comes from space charge that will cause the equally charged electrons to repel each other, thereby blowing up the beam and increasing the emittance. In this way the initial emittance is primarily determined in other particle sources, but even after leaving the source and accelerated to higher energies the emittance can change due to stochastic and friction forces. The former include the emission of synchrotron radiation, which is an essentially stochastic quantum mechanical process that we will consider more carefully later. Another stochastic process that will increase the beam size is the interaction with a target or scattering of the particles off of each other in high intensity beam, which is called intrabeam scattering. Friction or damping forces are also caused by synchrotron radiation, but also by explicit cooling devices, such as electron coolers or a stochastic cooling system.

The quantitative behaviour of such systems can be nicely visualized with a simple one dimensional example where a particle on the x-axis receives a sequence

of random kicks of rms magnitude θ which is a called Wiener process. The individual particle experiences the dynamics

$$x_{n+1} = x_n + \theta \hat{P} \tag{41}$$

which called a Langevin equation of this stochastic process, where $\hat{P}$ is a 'function' that produces gaussian distributed random numbers with zero average and unity rms. If averaged over many realizations, either in time or over ensemble, where angle brackets denote the apprpriate averaging, we can write $\langle P \rangle = 0$ and $\langle P^2 \rangle = 1$. If the particle is part of an ensemble (think beam) we can calculate how the average of the ensemble $X = \langle x \rangle$ changes from time n to time $n+1$, and just do the ensemble average of the previous equation

$$X_{n+1} = \langle x_{n+1} \rangle = \langle x_n + \theta \hat{P} \rangle = X_n + \theta \langle P \rangle = X_n \tag{42}$$

which means that random kicks with mean zero do not change the center or average position of the distribution. Intuitively one would expect that these random kicks will make the distribution wider rather than move its center. The width, however, is described by the centered second moment of the distribution, or the rms, which we now calculate

$$\sigma_{n+1}^2 = \langle (x_{n+1} - X_{n+1})^2 \rangle = \langle (x_n + \theta \hat{P} - X_n)^2 \rangle = \sigma_n^2 + \theta^2 \tag{43}$$

which means the square of the rms with increases linearly with turn number. Assume that initially we start with all particles at zero and have $\sigma_0 = 0$. The rms width σ_n after n random kicks of the distribution is

$$\sigma_n^2 = \sigma_0^2 + n\theta^2 = n\theta^2 \tag{44}$$

which implies that the rms width itself increases with the root of the number of kicks

$$\sigma_n = \theta \sqrt{n} = \theta \sqrt{t/T} \tag{45}$$

where we introduced the elapsed time t and the time between random kicks T to replace the kick number n. This temporal behaviour is characteristic for a diffusion process and can also be derived from a Fokker-Planck equation with diffusion constant $D = \theta^2/T$.

Including damping in the above analysis amounts to adding a damping term to eq 41 which turns it to

$$x_{n+1} = e^{-\alpha} x_n + \theta \hat{P} \tag{46}$$

where $\alpha = T/\tau$ is the damping decrement related to the 'revolution time' and the damping time τ. Performing a similar analysis we can find the equilibrium rms width σ_∞ from the requirement $\sigma_{n+1} = \sigma_n$ which results in

$$\sigma_\infty^2 = \frac{\theta^2}{1 - e^{-2\alpha}} \approx \frac{\theta^2}{2\alpha} = \frac{\tau \theta^2}{2T} \tag{47}$$

provided that the damping decrement α is small compared to unity.

Turning back to the accelerator physical problem with randomly excited betatron oscillations. We can quantitatively analyze the emittance increase due to the random transverse kicks as is common for example in a traget and to a lesser extent in the interaction of the beam with the residual gas in the accelerator. Let's consider a simple ring defined by a transfer matrix R with tune Q and beta function β

$$R = \begin{pmatrix} \cos(2\pi Q) & \beta \sin(2\pi Q) \\ -\sin(2\pi Q)/\beta & \cos(2\pi Q) \end{pmatrix} \tag{48}$$

and the beam is the given by its sigma matrix in eq. 37. We now want to calculate the change in emittance if we subject the beam to a sequence of random kicks the rms kick angle θ_{rms} which changes the x' coordinate to $x' + \theta_{rms}\hat{P}$ where $\hat{P}$ is the random process we encountered earlier. The equivalent Langevin equation now reads

$$\begin{pmatrix} x_{n+1} \\ x'_{n+1} \end{pmatrix} = R \begin{pmatrix} x_n \\ x'_n + \theta_{rms}\hat{P} \end{pmatrix} \tag{49}$$

and we can calculate the dynamics of the averages $X = \langle x \rangle$ and $X' = \langle x' \rangle$ and the rms width $\sigma_x^2 = \langle (x - X)^2 \rangle$, angular divergence $\sigma_{x'}^2 = \langle (x' - X')^2 \rangle$ and correlation $\sigma_{xx'} = \langle (x - X)(x' - X') \rangle$ which constitute the sigma matrix and calculate the emittance as the determinant of the sigma matrix and find that the emittance growth rate for a single turn is

$$\frac{d\varepsilon}{dn} = \frac{\beta}{2}\theta_{rms}^2 \quad \text{or} \quad \frac{d\varepsilon}{dT} = \frac{\beta}{2T}\theta_{rms}^2 \tag{50}$$

where n is turn number and T is the revolution period. This expression describes the emittance growth rate due to for instance the transverse kicks the beam experiences in a target on with scattering off of the residual gas in the beam pipe, but also due to an exit window where the beam leaves the vacuum system in order to interact with an external target. Later we will estimate the rms scattering angles for the different processes as a function of material type and thickness.

Note that we dealt with a process that explicitely changes the transverse angular coordinate of the particles in the beam. There are other processes that randomly change the momentum of the beam. The emission of synchrotron radiation is the prime example, but also phase jitter of an RF system will randomly change the energy of particles. This can have a an influence on the emittance if there is a correlation between energy and transverse position at the location where the energy change occurs. We will discuss this process after having introduced the dispersion function in one of the next sections.

4.10 Beta functions, the classical approach

In the previous sections we discussed the transport of particles and beams in magnet lattices. From the basic quantities, the transfer matrices and the beam matrix we deduced parameterizations in terms of emittances and beta functions that are commonly used when describing accelerators, but are just convenient parameters. In the next section we will connect this approach with that normally found in textbooks on accelerator physics, such as Ref. [1] or [3].

From the transfer-matrices for drift and thin quadrupole we see that the forces that act on the particle are a function of the position s in the accelerator and we can write

$$x''(s) - k(s)x = 0 \tag{51}$$

where $k(s) = (dB_y/dx)/pc$ is the integrated quadrupole strength normalize to the momentum $pc = B\rho$. We can interpret Eq. 51 in the following way: either there is no quadrupole at position s in which case we have $k(s) = 0$ or there is one with constant gradient. If $k > 0$ the quadrupole is focusing and for $k < 0$ it is defocusing. If the accelerator is circular, Eq. 51 is a Hill-equation and $k(s) = k(s+C)$ where C is the circumference or the periodicity of the accelerator.

We can solve Eq. 51 with the following Ansatz for the trajectory of the particle

$$x(s) = Aw(s)\cos(\psi(s) + \psi_0) \tag{52}$$

with a periodic amplitude function $w(s)$ and a phase function $\psi(s)$. Of course this ansatz is motivated by the observation that the motion of a particle in an accelerator (at least when it is stable) resembles that of a harmonica oscillator, albeit with variable amplitude and frequency. Inserting Eq. 52 in Eq. 51 and sorting according to cos and sin we obtain

$$A\left[w'' - w\psi'^2 - kw\right]\cos(\psi + \psi_0) - A\left[2w'\psi' + w\psi''\right]\sin(\psi + \psi_0) = 0\ . \tag{53}$$

In order to be valid for non-zero amplitude A and for any starting phase ψ_0 the expressions in the square brackets must vanish identically and we get

$$\begin{aligned} 0 &= w'' - w\psi'^2 - kw \\ 0 &= 2w'\psi' + w\psi''\ . \end{aligned} \tag{54}$$

Multiplying the second equation with w we can write it as $(w^2\psi')' = 0$. Integrating once we find $\psi' = 1/w^2$. Integrating once more we find

$$\psi(s) = \int_0^s \frac{dr}{w(r)^2} \tag{55}$$

and inserting this into the first of Eq. 54 we obtain the following expression for the amplitude function w

$$w'' - \frac{1}{w^3} - k(s)w = 0\ . \tag{56}$$

4.11 Quadrupole

We consider a quadrupole with constant gradient $k = (dB_y/dx)/B\rho$ that has length l and calculate the transfer matrix R that maps the input coordinates x_0, x_0' to those at the end of the quadrupole. Here it is convenient to start from the equation of motion

$$x'' + kx = 0 \ . \tag{57}$$

For $k > 0$ this equation is easily solved by $\cos(\sqrt{k}s)$ and the corresponding cosine function

$$x(s) = A_1 \cos(\sqrt{k}s) + A_2 \sin(\sqrt{k}s) \ . \tag{58}$$

The coefficients can be determined by matching to the initial values x_0, x_0' and we obtain

$$x(s) = x_0 \cos(\sqrt{k}s) + \frac{x_0'}{\sqrt{k}} \sin(\sqrt{k}s) \ . \tag{59}$$

At the end of the quadrupole we have $s = l$ and can write for the 2×2 horizontal transfer-matrix

$$R = \begin{pmatrix} \cos(\sqrt{k}l) & \frac{1}{\sqrt{k}} \sin(\sqrt{k}l) \\ -\sqrt{k} \sin(\sqrt{k}l) & \cos(\sqrt{k}l) \end{pmatrix} \tag{60}$$

that maps the initial coordinates x_0, x_0' to those at the end of the quadrupole. Note that in the limit of a thin quadrupole with $l \to 0$, while keeping $\sqrt{k}l$ constant the transfer matrix approaches that of a thin focusing quadrupole where kl equals the inverse focal length $1/f$.

In case the quadrupole is defocusing and $k < 0$ we can solve Eq. 57 in terms of hyperbolic sines and cosines. the corresponding transfer matrix is then given by

$$R = \begin{pmatrix} \cosh(\sqrt{-k}l) & \frac{1}{\sqrt{-k}} \sinh(\sqrt{-k}l) \\ \sqrt{-k} \sinh(\sqrt{-k}l) & \cosh(\sqrt{-k}l) \end{pmatrix} \tag{61}$$

and the 4×4 matrix can be built by placing the 2×2 matrices on the diagonal and 2×2 zero-matrices on the off-diagonal places.

4.12 Dipole

The primary task of the dipole magnets in an accelerator is to define the reference orbit. Intuitively there are two parameters that define the effect of a dipole magnet on a particle with momentum $\vec{p}$, the magnetic field $\vec{B}$ and the length of the dipole. The trajectory in the dipole is determined by the balance of the centrifugal force and the Lorenz force

$$\frac{\gamma m v^2}{\rho} = e\vec{v} \times \vec{B} \tag{62}$$

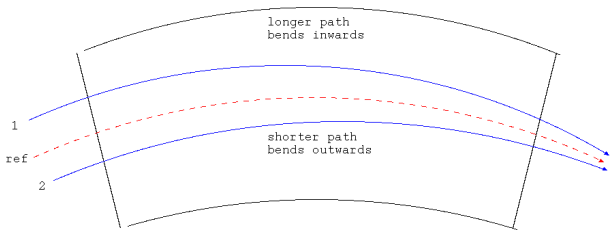

Figure 20: Top-view on a sector bending magnet explaining dipole focusing.

where γm is the relativistic mass of the particle with mass m and speed v. Solving for the bending radius $1/\rho$ we find

$$\frac{1}{\rho} = \frac{eB}{pc} \tag{63}$$

where we introduced the momentum $p = \gamma mv$. Note that this equation implies a proportionality between the momentum p and the quantity $B\rho$ and often the latter is used to quantify the momentum of the particle, because it conveniently relates the the magnetic field to a geometric quantity, ρ. The 1.5 TeV electron beams in CLIC have a $B\rho$ of 5000 Tm which immediately implies that a magnet with 1 T would provide a bending radius of 5 km. This makes relating engineering quantities (Tesla) to beam related quantities (bending radius) very easy. From the radius of curvature ρ we can deduce the bending angle ϕ by simple geometric reasoning

$$\phi = \frac{l}{\rho} = \frac{eBl}{pc} \ . \tag{64}$$

which increases proportionally with the magnetic field and inversely proportional with the momentum. Higher energy beams are 'stiffer' and refuse to bend easily in a given magnetic field.

We now have to consider two different type of dipoles which often appear in accelerators, rectangular bends and sector bends and start by considering the latter. A sector bend is defined as a bending magnet into which the reference trajectory enters and exits perpendicular to the entrance and exit face of the magnet. This implies that the entrance and exit faces are *not* parallel to each other and that the magnet looks like a piece of a slice of a pie, as is indicated in Fig. 20 which shows a top view of a horizontal sector bending magnet with the dashed reference trajectory labeled by 'ref'. The trajectory, labeled '1', that enters the bend further on the outside – at larger radius vector – of the the reference trajectory experiences a longer magnet and is therfore bent more, as is shown in Fig. 20. The converse is true for the trajectory on the inside, labeled '2'. The effect of the dipole is that a particle off the reference trajectory experiences a kick $\Delta x'$ proportional to its distance from the reference trajectory, which is similar to a focusing quadrupole and this effect is indeed called 'weak focusing'.

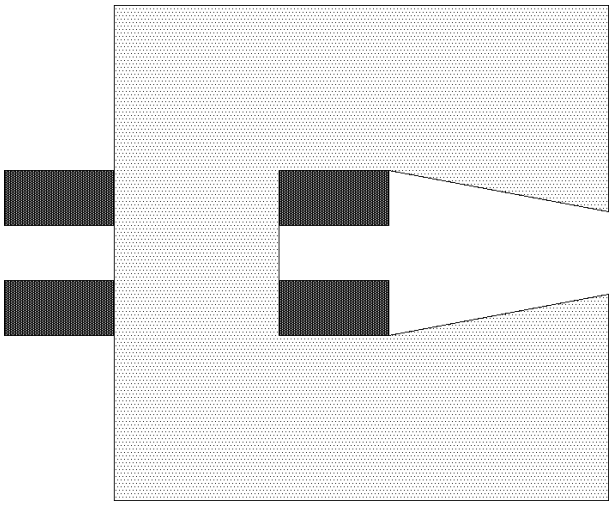

Figure 21: Side view of a combined function dipole.

In order to calculate this focusing effect we consider the difference in bending angle that a particle with distance x from the reference trajectory experiences in an infinitesimally short dipole of length ds

$$\Delta x' = \phi(0) - \phi(x) = \frac{ds}{\rho} - \frac{ds}{\rho}\frac{\rho + x}{\rho} = -\frac{ds}{\rho^2}x \tag{65}$$

which can be rewritten as

$$x''(s) + \frac{1}{\rho^2}x = 0 \tag{66}$$

resulting in an equation similar to that describing a quadrupole, except that the strength-parameter k is replaced by $1/\rho^2$. Note that the focusing effect is proportional ρ^2 and is therefore independent of the sign of the magnetic field B. This is particularly important when considering the weak focusing of wigglers with alternating magnetic field, which *does not* cancel on average.

Apart from the horizontal weak focusing we just encountered, sector bends can have a quadrupole-like gradient added. This gradient can, for example be generated by tilted pole faces, such as those shown in Fig. 21 which shows an example of a combined function magnet. It is called that way, because it serves two purposes, it bends and it focuses similar to a quadrupole, thus combined function. The differential equations that govern the dynamics in a combined dipole is thus described by the following equations

$$\begin{aligned} x''(s) - \left(k + \frac{1}{\rho^2}\right)x &= 0 \\ y''(s) + ky &= 0\ . \end{aligned} \tag{67}$$

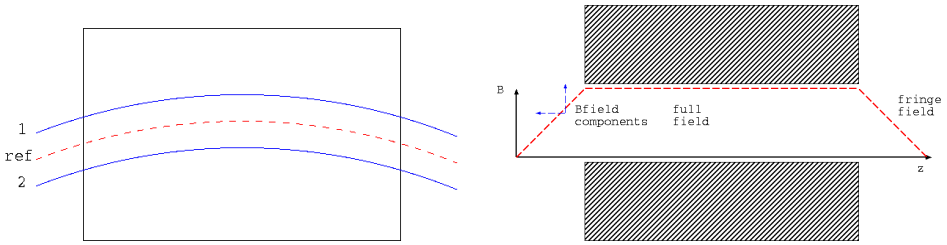

Figure 22: (Left) Top-view of a rectangular bending magnet and three horizontal trajectories. (Right) Side-view of a rectangular bending magnet and the vertical magnetic field profile. Note the fringe field.

The 2×2 transfer matrices that describes focusing in the respective planes of such a bend are given by Eq. 60 and Eq. 61, depending on the sign of $k + 1/\rho^2$ and k.

Apart from sector bending magnets such as that shown in the top-view in Fig. 20 there are also rectangular bends, so-called RBENDs which have parallel entrance and exit faces. In such a magnet the length of the trajectory does not depend on the horizontal offset, as was the case for sector bends, discussed earlier in this section. Therefore, there is no horizontal focusing in a rectangular bend. Of course, a quadrupole gradient can be added by shaping the pole face (see Fig. 21) or other means such as additional coils.

In a rectangular bend the particles enter the fringe field with a horizontal angle and can interact with the longitudinal component of the $B-$field which is present in the fringe-field region, because the vertical component B_y varies with z and due to $dB_y/dz = dB_z/dy$ also the longitudinal component B_z varies with the vertical distance to the center of the magnet y. The Lorenz-force resulting from a longitudinal field component with a horizontal velocity component of a particle will be vertical. Consequently we have a vertical force, that depends on the vertical position.

We will now approximately calculate the magnitude of this effect by first observing that

$$\frac{dB_z}{dy} = \frac{dB_y}{dz} \approx \frac{B}{g} \tag{68}$$

where g is the full gap height. We thus assume that the full vertical field inside the magnet decays to zero over the longitudinal distance of one gap height. Moreover, here we assume that this decay is linear, which is only a crude approximation that is convenient in the calculation. Inside the fringe-field region the longitudinal component therefore can be approximated by

$$B_z = \frac{B}{g} y \tag{69}$$

with an integration constant that we ignored for the time being. The vertical force

that a particle experiences is given by the vertical component of the Lorenz-force equation

$$\frac{dp_y}{dt} = ev_x B_z \approx \frac{\phi}{2} c \frac{B}{g} y \tag{70}$$

and changing the time derivative dt to the a derivative along the beam line by $dz = c\,dt$ we obtain after integrating over the longitudinal extent of the fringe-field

$$\Delta p_y = \frac{\phi}{2} B y \tag{71}$$

and normalizing by the total momentum $pc = B\rho$ results in

$$\Delta y' = \frac{\Delta p_y}{p} = \frac{\phi}{2} \frac{y}{\rho} \,. \tag{72}$$

Rewriting this in terms of a transfer matrix we find

$$\begin{pmatrix} y_2 \\ y_2' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{\tan(\phi/2)}{2\rho} & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_1' \end{pmatrix} \tag{73}$$

where we replaced the approximate value of the deflection angle $\phi/2$ by $\tan(\phi/2)$ which follows from a more careful treatment that is for example shown in Ref. [4]. The matrix in Eq. 73 is the map from just outside the fringe field to just inside the magnet. The same effect will affect the particle on its way out of the rectangular bend and therefore the combined effect of a rectangular bend with length l in the non-deflecting plane is given by

$$R_y = \begin{pmatrix} 1 & 0 \\ -\frac{\tan(\phi/2)}{2\rho} & 1 \end{pmatrix} \begin{pmatrix} 1 & l \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\frac{\tan(\phi/2)}{2\rho} & 1 \end{pmatrix} \tag{74}$$

which can be used numerically in beam optics codes.

In this section we considered the effect of the dipole on a particle on the reference orbit defining the bending angle and the effect of a transversely offset particle, namely focusing, but in the next section we will discuss the effect of a momentum offset in a dipole magnet and what influence this has on the particle orbit.

4.13 Dispersion

The dipole magnets in an accelerator deflect the beam by an angle given by $\phi_0 = eB/p_0 c$ and the nominal deflection angle is only given for a particle with the reference momentum p_0. If the particle has a momentum that differs by a small amount the angle also varies slightly and we have

$$\phi(\delta) = \frac{eB}{p_0(1+\delta)c} \approx \phi_0(1-\delta) = \phi_0 - \phi_0\delta \tag{75}$$

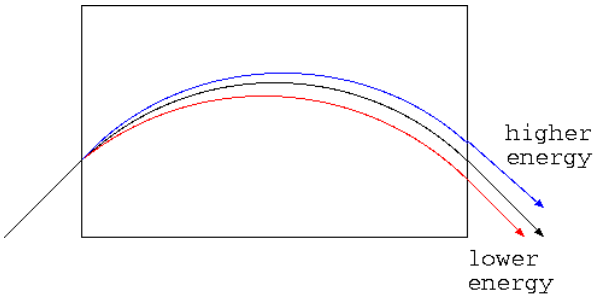

Figure 23: Effect of a dipole on particle with different energies.

which implies that the particle gets a small kick $-\phi_0\delta$ with respect to the reference trajectory, depending on its relative momentum error $\delta = \Delta p/p$. This is just the effect that is desired in a spectrometer, where the particles are transversely sorted according to their energy.

Conceptually the *dispersion* is a particle trajectory of a particle with momentum offset δ. If we consider the start of a beam line and we assume that particles with different energies are located on the reference orbit. As the particles progress along the beam line, they stay together on the reference orbit until they meet a dipole magnet, when they start diverging and are being sorted according to their energy. In the subsequent sections of the beam line the off-energy particle will perform oscillations around the reference trajectory. In every dipole magnet if will, however, receive a small kick that will change these oscillations with respect to the reference trajectory.

Since the dispersion is 'generated in the dipoles' we now have a closer look at the detailed trajectory inside the dipole and will calculate an extended transfer matrix for the dipole that takes energy offset into account. The additional kick that a particle with momentum offset δ receives in an infinitesimally short dipole magnet is

$$\Delta x' = \delta \frac{ds}{\rho} \tag{76}$$

which is the momentum dependent effect that we need to add to the differential equation that describes the motion in the bending-plane of a dipole Eq. 66 and we arrive at

$$x'' + \frac{1}{\rho^2} x = \frac{1}{\rho}\delta \ . \tag{77}$$

Since this is an ordinary linear differential equation we can assume that the solution is proportional to the momentum offset δ and introduce the dispersion function $D(s)$ with $x = D\delta$. The differential equation then simplifies to

$$D'' + \frac{1}{\rho^2} D = \frac{1}{\rho} \tag{78}$$

which, as can be easily verified, is solved by the Ansatz

$$D(s) = A\cos(s/\rho) + B\sin(s/\rho) + \rho \tag{79}$$

with integration constants A and B. If the dispersion function and its derivative at the entrance of the dipole are denoted by D_0 and D_0' we can express the integration constants in terms of the initial values and write the trajectory of the off-momentum particle as $x = D\delta$ in terms of its initial values as

$$D(s) = D_0\cos(s/\rho) + \rho D_0'\sin(s/\rho) + \rho[1-\cos(s/\rho)]\ . \tag{80}$$

and an equation for $D'(s)$ that can be calculated easily by differentiating Eq. 80 with respect to s.

Since the dispersion function $D(s)$ describes the position of off-momentum particles in the accelerator it is important in the design of accelerators and, since it stems from a *linear* differential equation is easily calculated by transfer-matrices. The complete 6×6 transfer matrix for a sector bend is then given by

$$R = \begin{pmatrix} \cos\phi & \rho\sin\phi & 0 & 0 & 0 & \rho(1-\cos\phi) \\ -\sin(\phi)/\rho & \cos\phi & 0 & 0 & 0 & \sin\phi) \\ 0 & 0 & 1 & l & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ -\sin\phi & -\rho(1-\cos\phi) & 0 & 0 & 1 & l/\gamma^2 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{81}$$

with the bending angle $\phi = l/\rho$. So far we have not motivated the entries in the fifth row, which describe the dependence of the arrival time (or equivalently, longitudinal position in the bunch) on the dispersion at the entrance of the dipole. Clearly, if the incoming dispersion is such that the initial angle D_0' is pointing outwards in the dipole, the trajectory inside the dipole is longer and the particle arrives later, or further back in the bunch. A similar argument holds for D_0. The 56-element describes the effect that particles with different energies have slightly different speeds, despite being ultra-relativistic and therefore have different arrival times.

The matrix in Eq. 81 can be used to propagate the dispersion function through an accelerator or a beam line if the initial values are given. In a circular accelerator, we have to require, however, that the dispersion function is periodic with the circumference. We therefore have to calculate the initial values at the starting point of the ring from the full-turn matrix. In this case by full-turn matrix we mean the entire 6×6 matrix. Let's assume that we have done so and denote the full-turn matrix by $\tilde{R}$. If the ring is not coupled we can consider a reduced 3×3 matrix which consists of row and columns 1,2,6 only and denote that reduced matrix by S. This matrix S has the following structure

$$S = \begin{pmatrix} \tilde{R}_{11} & \tilde{R}_{12} & \tilde{R}_{16} \\ \tilde{R}_{21} & \tilde{R}_{22} & \tilde{R}_{26} \\ 0 & 0 & 1 \end{pmatrix} \tag{82}$$

and we will find the equilibrium solution by solving for a vector $\vec{D} = (\tilde{D}, \tilde{D}', 1)$ that satisfies

$$\vec{D} = S\vec{D} \tag{83}$$

which leads to

$$\begin{pmatrix} 1 - R_{11} & R_{12} \\ R_{21} & 1 - R_{22} \end{pmatrix} \begin{pmatrix} \tilde{D} \\ \tilde{D}' \end{pmatrix} = \begin{pmatrix} R_{16} \\ R_{26} \end{pmatrix} \tag{84}$$

and we finally arrive at

$$\begin{pmatrix} \tilde{D} \\ \tilde{D}' \end{pmatrix} = \frac{1}{(1 - R_{11})(1 - R_{22}) - R_{12}R_{21}} \begin{pmatrix} 1 - R_{22} & -R_{12} \\ -R_{21} & 1 - R_{11} \end{pmatrix} \begin{pmatrix} R_{16} \\ R_{26} \end{pmatrix} \tag{85}$$

which denotes the periodic, or equilibrium, dispersion values at the start of the circular beam line that is described by $\tilde{R}$.

The dispersion of an accelerator is, apart from the tunes and the beta functions, one of the most important characterizing functions. In particular is describes the distance an off-momentum particle travels away from the reference trajectory. In diagnostic equipment that tries to measure the momentum of the beam by observing its transverse offset it is therefore advantageous to place it at a location with large dispersion. Furthermore, it is easy to show that it also describes the contribution due to an rms momentum spread σ_p of the beam size of an ensemble of particle as $D\sigma_p$. The total beam size due to finite emittance and finite momentum spread is then given by

$$\sigma_x = \sqrt{\varepsilon_x \beta_x(s) + D^2(s)\sigma_p^2} \tag{86}$$

if we assume that the momentum spread and the emittance are uncorrelated.

4.14 Momentum Compaction Factor

The dispersion trajectory that we calculated in the previous section describes the orbit of a particle with momentum offset δ. If such a particle traverses a dipole magnet it will lie further on the out- or inside, depending on the sign of the dispersion function at that location and the sign of δ. If it lies further outside it will have a longer path to travel and therefore will arrive later at the exit of the magnet. The change in length Δl with momentum variation of the path in a single dipole with bending radius ρ is given by

$$\frac{\Delta l}{\delta} = \int_0^l \frac{D(s)}{\rho} ds \ . \tag{87}$$

In the case of a circular accelerator we can sum the pathlength change over all dipoles and normalizing to the circumference C of the ring to obtain the so-called momentum-compaction factor α defined by

$$\alpha = \frac{\Delta C}{C\delta} = \frac{1}{C} \int_{\text{all dipoles}} \frac{D(s)}{\rho} ds \tag{88}$$

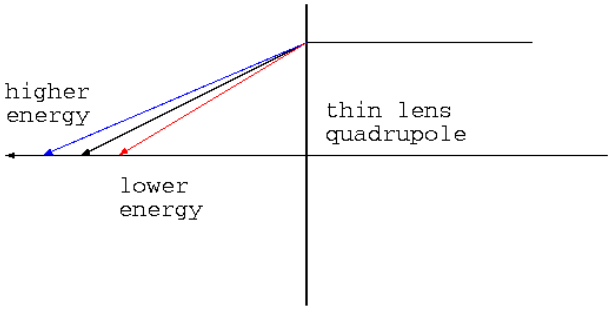

Figure 24: Effect of a quadrupole on particle with different energies. The focal point of the higher energy particle is further downstream than that of the reference and lower energy particle.

which gives the fractional change of the circumference normalized to the energy $(\Delta C/C)/\delta$. This quantity plays a central role in the stability of the longitudinal motion of the particles in a storage ring.

4.15 Chromaticity

In the same way that the dispersion describes the effect of a relative momentum error in dipole magnets is the chromaticity the corresponding effect in of quadrupole magnets. The effect of a quadrupole on the beam is given by the gradient normalized to the particle momentum and if the latter is different from that of the reference particle, the quadrupole does not do to the particles what it is intended to do. As an example we consider a thin quadrupole with focal length f which varies according to

$$f(\delta) = f(1+\delta) \quad \text{or} \quad k_1 \to \frac{k_1}{1+\delta} \approx k_1 - k_1\delta \tag{89}$$

and a quadrupole focuses a high energy particle less than the reference particle and consequently the focal point is further downstream. A lower energy particle is correspondingly stronger focused and its focal point lies closer to the quadrupole.

This effect can be quantified by considering the effect of an error in a single (thin) quadrupole in an otherwise perfect ring. For simplicity, assume that the ring is described by a transfer matrix

$$R = \begin{pmatrix} \cos(2\pi Q)\mu & \beta\sin(2\pi Q) \\ -\sin(2\pi Q)/\beta & \cos(2\pi Q) \end{pmatrix} \tag{90}$$

with beta function β and tune Q. We introduce a 'weak' thin quad with focal length f and obtain the combined transfer matrix by multiplication with the result

$$\begin{pmatrix} \cos(2\pi Q) & \beta\sin(2\pi Q) \\ -\sin(2\pi Q)/\beta - \sin(2\pi Q)/f & \cos(2\pi Q) - \beta\sin(2\pi Q)/f \end{pmatrix} \tag{91}$$

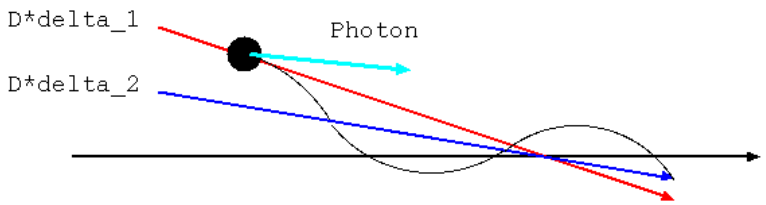

Figure 25: The mechanism that causes the excitation of betatron oscillations by a random energy loss.

and we remember that the trace of the matrix is related to the new tune $Q + \Delta Q$ of the perturbed system

$$\begin{aligned} 2\cos(2\pi Q) - \frac{\beta}{f}\sin(2\pi Q) &= 2\cos(2\pi(Q+\Delta Q)) \\ &= 2\cos(2\pi Q)\cos(2\pi\Delta Q) - 2\sin(2\pi Q)\sin(2\pi\Delta Q) \end{aligned} \tag{92}$$

and assume that the tune change ΔQ is small we can expand the cosine and sine to first order and compare coefficients to obtain

$$\Delta Q = \frac{\beta}{4\pi f} \tag{93}$$

which is a very well-used equation that describes the tune change due to a small quadrupole change. If the quadrupole is not thin but of finite length and with excitation error k_1 and length l we can write the tune change as an integral over the quadrupole length

$$\Delta Q = \frac{1}{4\pi}\int_0^l \beta(s) k_1 ds \ . \tag{94}$$

earlier in this paragraph we discussed that in the presence of finite energy error δ the excitation error of the quadrupole is given by $-k_1\delta$. If the momentum of a particle is different from the reference momentum *all* quadrupoles in the accelerator have the 'wrong' focusing strength. Thus the chromaticity being the change of the tune as a function of energy error is given by

$$\frac{\Delta Q}{\delta} \approx \frac{1}{4\pi}\int_{all\ \ quadrupoles} \beta(s) k_1(s) ds \tag{95}$$

where we assume that the nominal quadrupole excitations $k_1(s)$ are different for different quadrupoles and therefore depnd on the longitudinal position s in the accelerator.

4.16 Emittance growth due to random energy kicks

In the previous sections we found that the dispersion trajectory $D(s)$ is the closed orbit of a particle with momentum offset δ. If we assume that a particle with energy offset δ_1 is initially on its equilibrium orbit, the dispersion trajectory $D\delta_1$ as

is shown in Fig. 25. If that particle looses energy and has the new energy offset $\delta_2 < \delta_1$ at a position with non-zero dispersion it will stay at transverse position $D\delta_1$ but has energy δ_2. The equilibrium orbit of the particle with momentum δ_2 is, however, $D\delta_2$ and the particle finds itself away from its new equilibrium orbit and will therefore start oscillating around the new equilibrium orbit. In summary: initially the particle is on its equilibrium orbit, but through the energy loss the euilibrium orbit has jumped away and the particle starts betatron oscillations around the new equilibrium orbit. The same argument holds for D' the derivative of the dispersion. Note that this process is the dominant mechanism that determines the emittance in electron storage rings and synchrotron radiation sources.

The change in betatron state vector (x, x') that the particle through a relative momentum loss δ at a position with dispersion D receives is given by

$$\begin{pmatrix} x \\ x' \end{pmatrix} = -\delta \begin{pmatrix} D \\ D' \end{pmatrix} \tag{96}$$

or $\vec{x} = -\delta\vec{D}$ In the $m-$th turn the particle experiences the momntum change δ_m with corresponding betatron state change $\vec{x} = (x_m, x'_m)$. After n turns we then have to add all $\vec{x}_m$ over the previous turns with $m \leq n$

$$\begin{pmatrix} x \\ x' \end{pmatrix} = -\sum_{m=1}^{n} \delta_m R^{n-m} \begin{pmatrix} D \\ D' \end{pmatrix} . \tag{97}$$

In Ref. [1] a very elegant method is used to explictely calculate the emittance ε_n after n turns as defined in Eq. 29 as

$$\varepsilon_n = \gamma x_n^2 + 2\alpha x_n x'_n + \beta x'^2_n \tag{98}$$

where α, β, and γ are the twiss parameters at the location of the momentum kicks. Following Ref. [1] one finds for the emittance

$$\varepsilon_n = (\gamma D^2 + 2\alpha DD' + \beta D'^2) \sum_{k,l}^{n} \delta_k \delta_l \cos(2\pi(k-l)Q) . \tag{99}$$

If we assume that the δ from different turns are statistically independent and have rms δ_{rms} we can write the sum as

$$\sum_{k,l}^{n} \delta_k \delta_l \cos(2\pi(k-l)Q) = n\delta_{rms}^2 \delta_{kl} \tag{100}$$

wher δ_{kl} is the Kronecker delta which is unity for $k = l$ and zero else. For the final result we find

$$\frac{d\varepsilon}{dt} = \frac{\gamma D^2 + 2\alpha DD' + \beta D'^2}{T} \delta_{rms}^2 \tag{101}$$

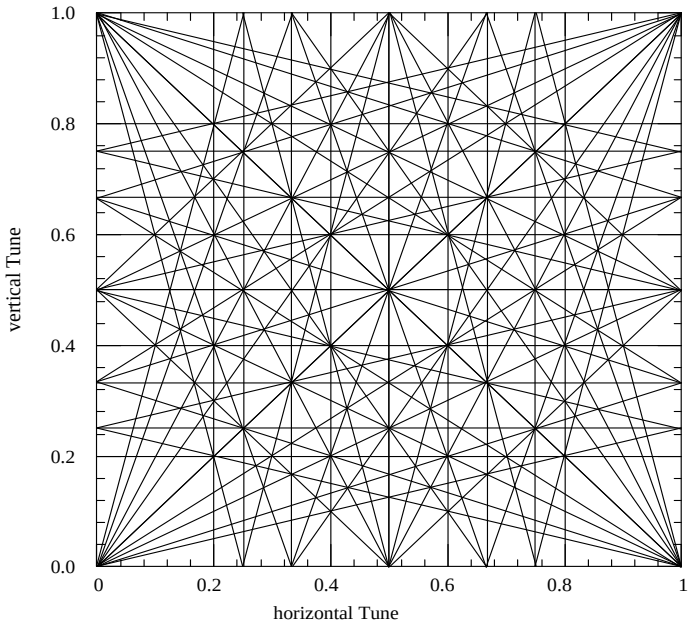

Figure 26: The tune diagram with all resonance up to fifth order.

where T is the revolution time.

We find that the susceptibility of the emittance growth $d\varepsilon/dt$ to the energy variation δ_{rms} is proportional to the factor

$$\mathcal{H} = \gamma D^2 + 2\alpha DD' + \beta D'^2 \tag{102}$$

and we immediately see that zero dispersion is desirable at location where the particles change their momentum, as is the case for RF cavities and, especially important for synchrotron light sources, for dipole magnets, where the beam looses energy due to synchrotron radiation. The design of small emittance lattices is focussed on designing magnet configurations that have small beta functions and small dispersion, in particular in the dipoles.

4.17 Sextupoles, Octupoles and other Multipoles

In an accelerator often higher multipoles are used for various purposes, some of which we will briefly discuss further below. The kicks that the multipoles cause to the beam are given by

$$\Delta x' + i\Delta y' = \frac{k_n L}{n!}(x + iy)^n \tag{103}$$

where we use the definition of the magnetic multipoles we will encounter in the section on magnets normalized to the particle momentum $B\rho$ by

$$k_n L = \frac{L}{B\rho}\frac{\partial^n B_y}{\partial x^n} \tag{104}$$

and for $n = 1$ we recover the integrated quadrupole gradient $k_1 L$.

In the previous section we that the chromaticity describes the momentum dependence of the tune, which means that particle with different energies see a different tune. Since the particles in a circular accelerator behave like a harmonic oscillator or a swing and we know that pushing a swing in the right beat will excite very large oscillations if the perturbation is in phase.

One prominent use of sextupoles is the correction of the chromaticity which we found to be the energy dependence of the tune due to the energy dependence of the focussing of quadrupoles. To compensate this we need other energy dependent focussing elements and a sextupoles located at a position with dispersion provides just this feature as we will see when considering the horizontal kick that particles with momentum offset δ experience in a sextupole. This is given by

$$\Delta x' = k_2 L(x + D\delta)^2 = k_2 L(x^2 + 2D\delta x + D^2\delta^2) \ . \tag{105}$$

The second term then behaves like an energy-dependent quadrupole with the focal length defined by

$$\frac{1}{f} = 2k_2 L D\delta \ . \tag{106}$$

By suitably choosing the locations and excitations of the sextupoles we are then able to adjust the energy dependent part of the tune, preferably such that is small to reduce the spread in the tune diagram in Fig. 26.

In the section about the emittance growth we found that in order to achieve small equilibrium emittances in synchrotron radiation sources the beta functions and the dispersion must be kept small whci requires closely spaced, strong quadrupoles, but many strong quadrupoles will cause the chromaticity to be large according to eq. 95 which in turn requires strong sextupoles to correct the chromaticity, which is a beneficial effect but they also add non-linear transverse kicks proportional to x^2 to the dynamics which might cause stability problems as we will discuss in the next section.

4.18 Non-linear Dynamics

The stability problems alluded to in the previous section can be visualized in simple one-dimensional model of a storage ring with tune Q and a single sextupole of unit strength. The map that describes this system is given by

$$\begin{pmatrix} x_{n+1} \\ x'_{n+1} \end{pmatrix} = \begin{pmatrix} \cos(2\pi Q) & \sin(2\pi Q) \\ -\sin(2\pi Q) & \cos(2\pi Q) \end{pmatrix} \begin{pmatrix} x_n \\ x'_n + x_n^2 \end{pmatrix} \tag{107}$$

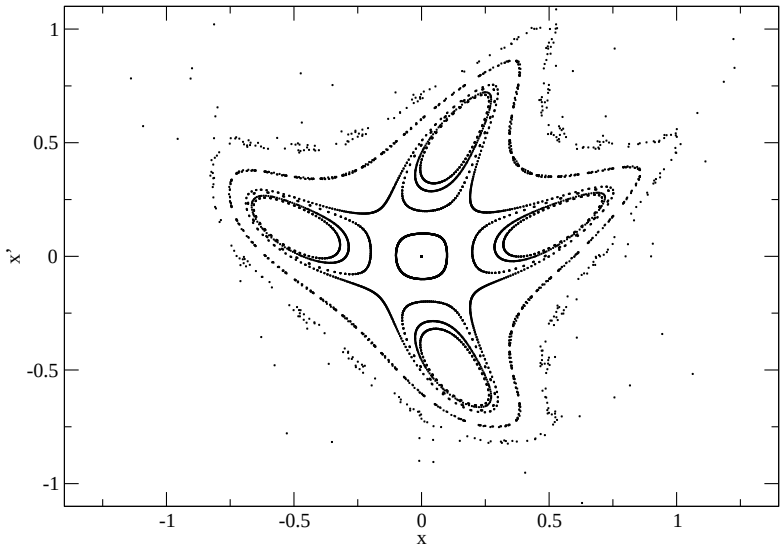

Figure 27: The $x - x'$ phase space for the map shown in Eq. 107 with tune $Q = 0.2516$.

which is often called the Henon-map and is one of the often-used examples when discussing non-linear effects in circular accelerator. The dynamics of a non-linear map is nicely visualized by displaying the phase-space (x_n, x'_n) on consecutive turns which is called a Poincare section of the dynamics. In Fig. 27 we show the phase plot of particles subjected to the map in eq. 107 with tune $Q = 0.2516$. As starting positions we use $x_0 = 0.1, \ldots, 0.9$.

For small amplitudes we find that the phase plot is almost circular as can be expected, because for small amplitudes the quadratic non-linearity is negligible. As the starting amplitude is increased the circles become more square, which is a consquence of the choice of tune that is close to 1/4. For starting amplitude $x_0 = 0.2$ there are four islands that are visited on consecutive turns. If the starting amplitude exceeds about 0.4 we observe a chaotic layer (deterministic chaos!) and eventually for even larger starting amplitudes the motion becomes unstable and the amplitude grows without bounds. The boundary that separates the motion limited to a finite phase space around the origin and the un-bounded motion is called the *dynamical aperture*. Note that this aperture is not a physical boundary or a wall like the vacuum chamber, but it is entirely determined by the dynamics of the system, which in turn is given by the magnets.

correct with octupoles

Hamiltonian dynamics, John Irwin's tricks

4.19 Steering Correctors

Steering dipole correctors are physically the same type of dipoles as the dipoles that define the reference trajectory. In simulation programs, however, they do not change the design trajectory, but only change the angle with respect to the reference trajectory of a particle $\Delta x'$ or $\Delta y'$ by a small amount. In this way they generate betatron oscillations with respect to the design orbit, but do not change the design orbit itself. Note, that this is only a matter of interpretation of the effect of the corrector. Physically such a magnet also produces a magnetic dipole field, just as their bigger companions, the bending dipoles.

In an accelerator steering correctors are usually needed to correct for example the effect of misaligned magnets or external stray fields on the beam. We will discuss their use more carefully in the section on diagnostics and correction.

4.20 Simulation Program MAD

In the previous sections we calculated the properties of the accelerator such as tune and beta functions by analytical means for some simple examples. In the real world, however, the accelerators and beam lines are too complex to analyze in such a way alone and computer programs were written to perform the tedious matrix manipulations and the extraction of the physical relevant quantities from the matrices. One of the first programs was TRANSPORT [12] written by Karl Brown in the late 1960s and one of the most used ones in more recent times is MAD [8], originally written by Christoph Iselin and Hans Grote at CERN. MAD is still under active development and is continuously updated and improved, especially for the design and commissioning of LHC. Apart from these programs there is a plethora of others such as DIMAD [14], TRACE3D [15], PARMILA [16], WINAGILE [17], updated TRANSPORT [13].

Since MAD is one of the most widely used codes and has a very usable input language we will also use it in this course. The executables for Windows and Linux can be downloaded from [8]. On the same web-site also the manual is available and instructions to compile and install MAD on MAC-OSX. Of course, MAD has evolved over the decades and the most recent version is called MADX which is what we will use in the following sections.

We start by reconsidering the FODO cell that we discussed in section 4.3. An input file with name FODO.MADX that we can run by writing

```
madx < FODO.MADX
```

on the command line in either Windows or Linux is shown in Fig 28. Running the file produces a postscript file with name `madx.ps` that can be viewed with

```
// file: FODO.MADX
// MADX Example 1: FODO cell
// Author: V. Ziemann, Uppsala University
// Date: 060910

BEAM, PARTICLE=ELECTRON,PC=3.0;

D: DRIFT,L=1.0;
QF: QUADRUPOLE,L=0.5,K1=0.2;
QD: QUADRUPOLE,L=0.5,K1=-0.2;

FODO: LINE=(QF,5*(D),QD,QD,5*(D),QF);
USE, PERIOD=FODO;

TWISS,SAVE,BETX=15.0,BETY=5.0;
PLOT,HAXIS=S, VAXIS=BETX, BETY;

MATCH, SEQUENCE=FODO;
PLOT,HAXIS=S, VAXIS=BETX, BETY;
```

Figure 28: The input file for a FODO cell.

any postscript interpreter such as Ghostview [18]. The two plots generated by running `FODO.MADX` are shown in Fig. 29 and 30.

We will now briefly discuss the contents of the `FODO.MADX` input file for MADX and go through it line by line. For a detailed description of all the commands, please consult the MADX manual from Ref. [8]. First we note the header information with lines that are preceeded by `//` which signifies a comment that can be added to make the input file more readable for humans. The program simply ignores in a line whatever comes after a `//` for people that like C++ or `!` for people that like Fortran. The first real data line contains the `BEAM` statement which informs MADX about the type of beam and the momentum we are dealing with. More parameters such as the emittances or the current can be defined in a `BEAM` statement. Notice that use of commas and semicolons. MADX is rather picky about missing separators, so be careful. In the next three lines the elements in the beam line are defined, first a drift, labeled `D` and half-length quadrupoles. Note that the needed parameters length and K_1 for the quadrupoles are also defined here. The quadrupoles were chosen to be of half-length to make the FODO cell start and end in the center of a quadrupole for symmetry reasons. In the subsequent line the element sequence is defined to consist of a `QF` element 5 drift spaces of 1 m length, two (half-length) `QD`, 5 drift spaces and the other half of the

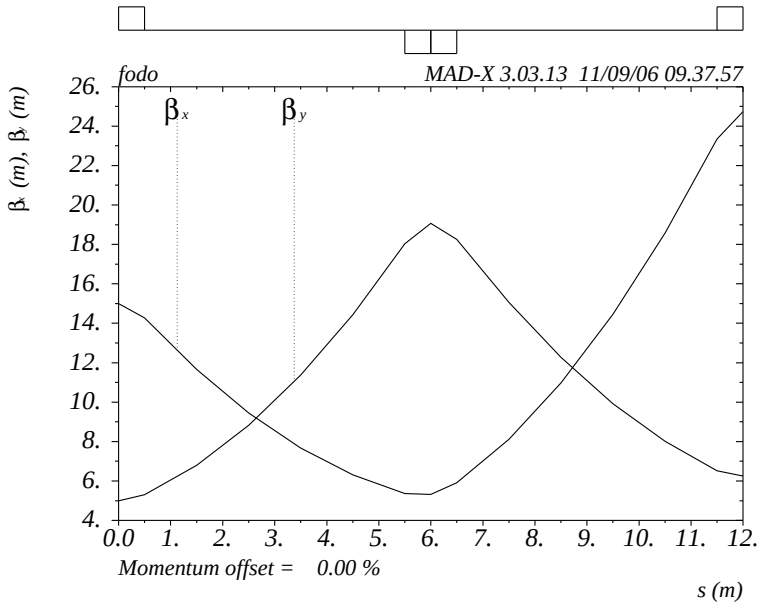

Figure 29: The beta functions from the first plot in `FODO.MADX` with fixed initial beta functions.

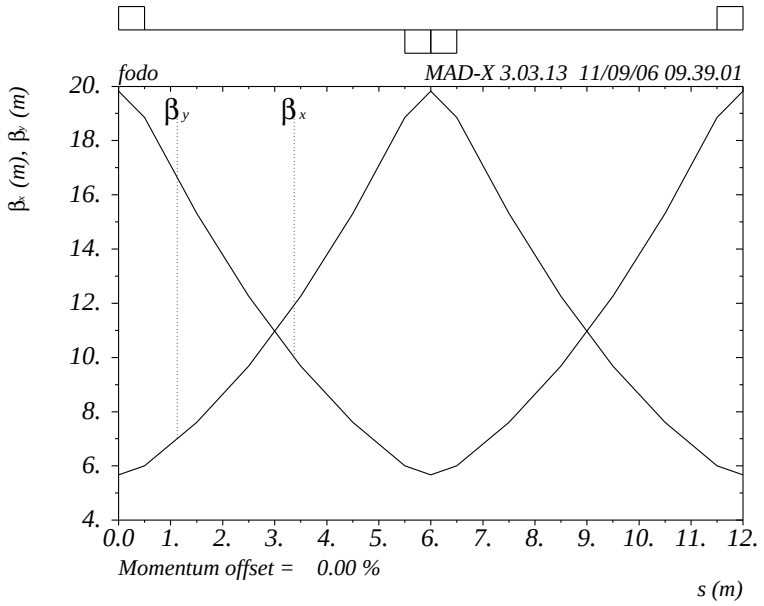

Figure 30: The beta functions after matching the periodic solution.

```
// file: FODO2.MADX
// MADX Example 2: FODO cell with dipoles
// Author: V. Ziemann, Uppsala University
// Date: 060911

BEAM, PARTICLE=ELECTRON,PC=3.0;

DEGREE:=PI/180.0;                  // for readability

QF: QUADRUPOLE,L=0.5,K1=0.2;       // still half-length
QD: QUADRUPOLE,L=1.0,K1=-0.2;      // changed to full length
B: SBEND,L=1.0,ANGLE=15.0*DEGREE;  // added dipole

FODO: SEQUENCE,REFER=ENTRY,L=12.0;
  QF1:   QF,      AT=0.0;
  B1:    B,       AT=2.5;
  QD1:   QD,      AT=5.5;
  B2:    B,       AT=8.5;
  QF2:   QF,      AT=11.5;
ENDSEQUENCE;

USE, PERIOD=FODO;

MATCH, SEQUENCE=FODO;
PLOT,HAXIS=S, VAXIS=BETX, BETY, INTERPOLATE=TRUE;
PLOT,HAXIS=S, VAXIS=DX, INTERPOLATE=TRUE;
```

Figure 31: The input file for a FODO cell with dipole magnets.

QF. We could have used a single drift space of 5 m length but then the output would look less nice with straight lines instead of, at least approximate, parabolas. In the line that starts with a USE command we inform MAD which beam line we will work on from here on.

Now that we have defined the beam line we apply the TWISS command to calculate the beta functions through the beam line with initial values specified in the TWISS command. Then the beta functions β_x, β_y are plotted versus the longitudinal position. The result of this plot is shown in Fig. 29. We note that the beta functions start at the indicated values of 15 and 5 m an then propagate through the FODO cell. We have not specified α_x and α_y which therefore are assumed to be zero. Note that the beta functions at the end of the cell are different from those at the beginning. If we assume that this cell is just one building block of many that constitute a periodic sequence we want to know

what initial beta functions reproduce themselves at the end of the sequence. MADX provides a so-called *matching* module that can vary parameters in order to achieve certain constraints. We will discuss this feature more further down, but the simplest matching is actually to find the periodic solution for the beta functions and this is done in the line that starts with the command `MATCH.` We then plot the resulting beta functions which are shown in Fig. 30 and verify that the solution is periodic.

Note that MADX labels the graphs such that it is easy to identify which beta function is which. Furthermore the magnet sequence is depicted on the top of the graph where focusing quadrupoles ($K_1 > 0$) are shown as boxes above the line and defocusing quadrupoles as boxes below the line. We will see in the second example that dipoles extend above and below the line.

In the second example we add the dipoles in the center of the straight section between the quadrupoles and we use the second format for the description of a beam line as a sequence of elements. At the top of the file that is shown in Fig. 31 we define the variable `DEGREE` which we use to convert degrees to radians. Note the `:=` sign which implies that whenever one of the variables on the right hand side changes anywhere in the file, the variable on the left hand side is recomputed. This is not needed here, but may serve as an example of this feature. Then the elements of the beam line are defined. Note the comments following the `//`. In the lines following the one starting with `FODO:` we define the beam line sequence. Observe that we only specify magnets. The drift spaces are generated by MADX automatically. In the line with `FODO:` we specify the reference position of the elements as `ENTRY` which means that the subsequent `AT=` commands refer to the entrance of the magnets, rather than the `CENTRE` which is the default. The total length of the beam line is specified here as well. The definition of the sequence is completed with the `ENDSEQUENCE` command. Again note all the semicolon! Then we instruct MADX to `USE` the just-defined beam line sequence, `MATCH` for the periodic optical functions and display the results in two plots. Note that the plots here use the `INTERPOLATE` option of the plot command to make them smoother. We see that we can use `DX` as keyword to display the dispersion. Other variables that can be displayed, apart from the beta functions, are explained in the MADX manual in the section about variables.

In the previous sections we showed how to extract the beta functions and the dispersion and now we show how to display other beam optical quantities such as tunes or chromaticities. MADX actually prepares an internal table called `SUMM` with global beam optical quantities as soon as the `TWISS` command or the `MATCH` command is executed. The table values for tunes can be accessed with the following command

```
Value, TABLE(SUMM,Q1);
Value, TABLE(SUMM,Q2);
```

which prints the horizontal and vertical tune to standard output. The chromatic-

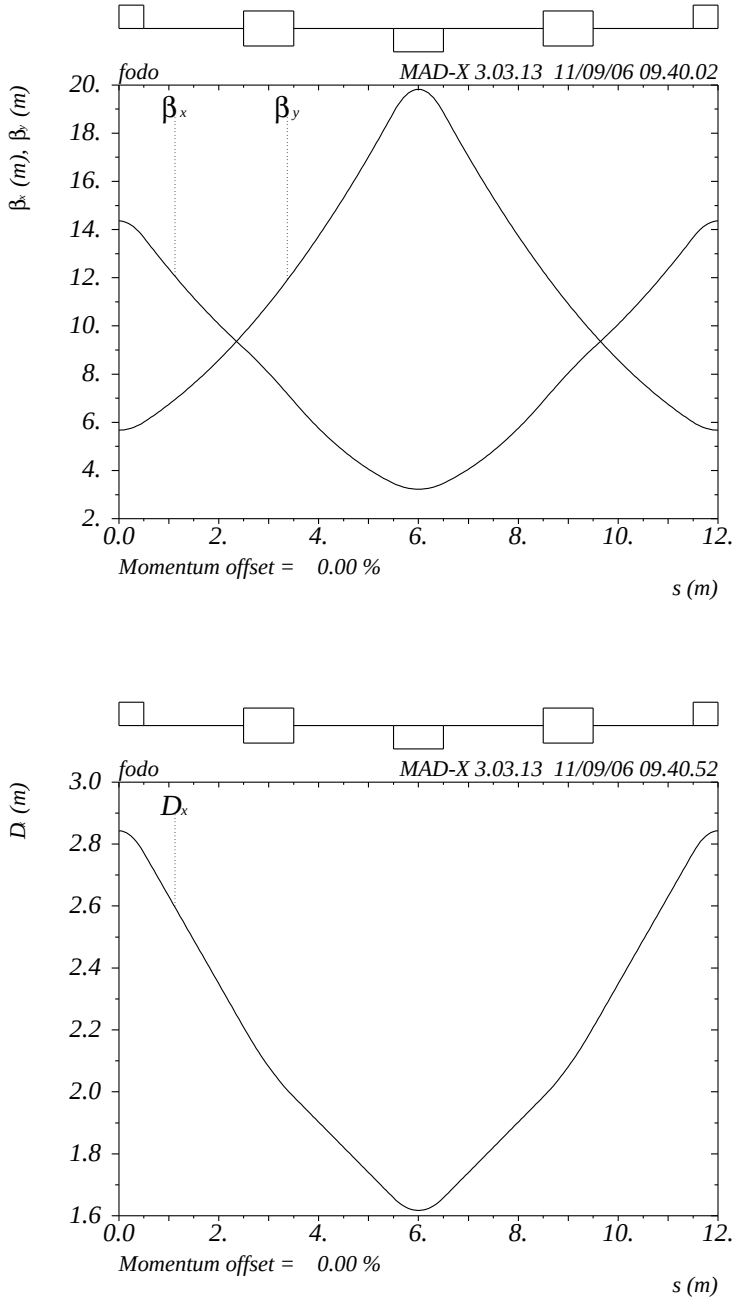

Figure 32: The periodic solution for the FODO cell with dipoles included. The upper plot shows the beta functions and the lower plot shows the horizontal dispersion function.

ities are accessible via

```
Value, TABLE(SUMM,DQ1);
Value, TABLE(SUMM,DQ2);
```

and other variables such as the maximum and rms values of beta functions and dispersions are also available in this table. See the section on *Variables in the SUMM table* in the MADX manual for the available quantities.

Sometimes it is useful to have direct access to the beta functions, phase advances and dispersion in numerical, tabular form instead of graphical output. Moreover, the transfer matrices in a beam line are sometimes also useful to have available in numerical format. The following sequence of commands performs the trick and writes two output files.

```
SELECT,FLAG=SECTORMAP,clear;
SELECT,FLAG=TWISS,column=name,s,betx,bety;
TWISS, file=optics.dat,sectormap;
```

The first line registers the demand to write the sectormap file which contains the transfer matrices. the second command selects the columns that will be written to the `optics.dat` file and `sectormap` directive causes the `TWISS` command to write the sectormap data to file called `sectormap`. Note that all variables from the summary table are also written to `optics.dat`.

An often encountered problem that appears in the disgn of beam optical system is that certain constraints must be fulfilled using a number of magnets that can be varied in order to achieve the desired objective. This process is called *matching* and there is a corresponding command in MADX that allows he user to define the constraints and the magnets to be varied. As an example we will match the phase advance of the FODO cell encountered in example 1. If inserted immediately after the `USE` statement the following sequence of commands

```
MATCH, SEQUENCE=FODO;
CONSTRAINT,SEQUENCE=FODO,RANGE=#E,MUX=0.16666666,MUY=0.25;
VARY,NAME=QF->K1,STEP=1E-6;
VARY,NAME=QD->K1,STEP=1E-6;
LMDIF,CALLS=500,TOLERANCE=1E-20;
ENDMATCH;
```

will vary the quadrupole excitations in example 1 to achieve the phase advance per cell of $0.1666 * 360 = 60$ degree in the horizontal plane and 90 degree in the vertical. Note that the command sequence is started by the keyword `MATCH` with the definition of the sequence `FODO` that will be matched. In the second line the `CONSTRAINT` is defined. The specifier `#E` specifies the end of the sequence and `MUX` and `MUY` specify the phase advance in tune-units, i.e. in radians divided by 2π. More than one line of constraints can be used. The following two lines define

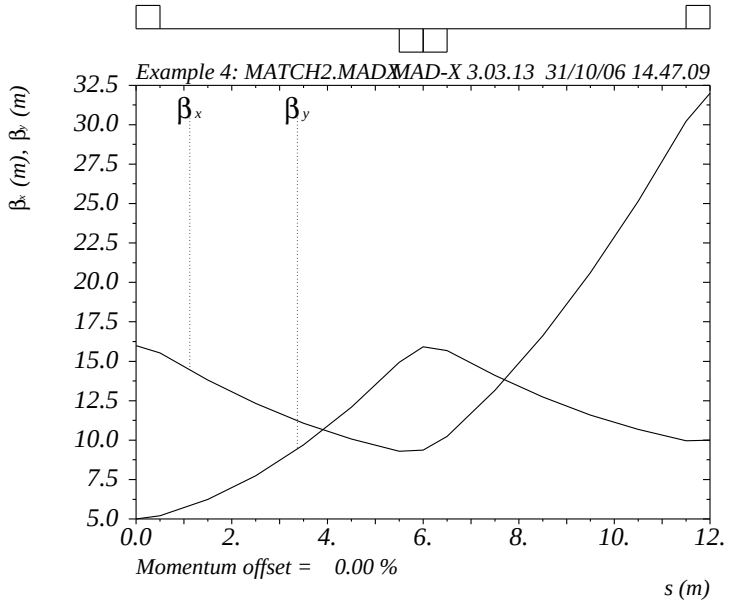

Figure 33: The initial beta functions aer matched to twice their value at the aned of the cell.

the parameters that will be varied and `QF->K1` denotes the normalized strength, the k_1-value of magnet with name `QF`. The step size specifies the approximate initial step size for varying the variable. The next line specifies the minimizer, here `LMDIF` is specified and the maximum number of iterations `CALLS` and the desired tolerance of the cost-function is specified. The `ENDMATCH` command exits matching module. Note that this matching task is usually done when designing a beam line and one wants to find a standard module that is then used repeatedly. The arcs of a collider or a long beam line are examples.

Another often encountered task is finding quadrupole excitations that will transform the beta functions at the start of the beam line to desired beta functions at the end of a beam line. An example is the matching of the beta function at the interaction point to a small value. Here, however, we illustrate the procedure by using the FODO cell from example 1.

```
MATCH, SEQUENCE=FODO,BETX=16,BETY=5;
CONSTRAINT,SEQUENCE=FODO,RANGE=#E,BETX=32,BETY=10;
VARY,NAME=QF->K1;
VARY,NAME=QD->K1;
LMDIF,CALLS=500,TOLERANCE=1E-20;
ENDMATCH;
```

```
PLOT,HAXIS=S, VAXIS=BETX, BETY;
```

Note that we specify the initial beta functions in the first line of the `MATCH` statement and the desired betafunctions at the end of the beam line in the contraint section. The definition of the varied magnets and the optimizer is the same as in the previous example. We added a `PLOT` statement in order to verify that the matching succeeded and generate a display of the result of the match which is shown in Fig. 33.

The matching module of MADX is very powerful and versatile and for details you should consult the manual available at [8]. Most parameters can be varied and constraints can be very general, either fixed values can be defined, or ranges or maximum values for beta functions, dispersion or global values such as the tune or chromaticity.

5 Beam Optics Examples

In this section we discuss several examples of how the beam optics of accelerators is designed. The general strategy is one of 'divide-and-conquer' in the sense that one designs suitable modules that are optimized individually and then used repeatedly. Another guiding principle is that of using optics concepts borrowed from ray light optics.

We start by considering a circular collider such as LHC and the beam optical elements that consitute it and then we will discuss design criteria for synchrotron light sources and finally some special sections such as bunch compressors. But, let's start with the collider.

5.1 FODO arcs

In a large circular collider the interaction regions (IR) are separated by long arcs in which the beam must be transported from IR to another. Therefore a simple lattice, which is what a magnet configuration called, is desireable unless other constraints are important. In a collider the arcs are mostly transport-sections that get the beam around. The FODO lattice that we encountered in the previous section fulfills this requirement. It is easily tunable since the horizontal and vertical beam sizes are large in their alternate quadrupoles which means that the QF mostly affect the horizontal beta function, because the vertical beam size is small in them and we remember that the tune shift due to a quadrupole change is given by $\beta/4\pi f$. This decouples correction of the respective tunes in an almost natural way. If the constraints on the beam size ar moderate the distance between qudrupoles can be made large and the excitation of the quadrupoles rather moderarte, which means few quadrupoles are required and the power consumption is small. The dispersion function will oscillate between its normally non-zero extreme values.

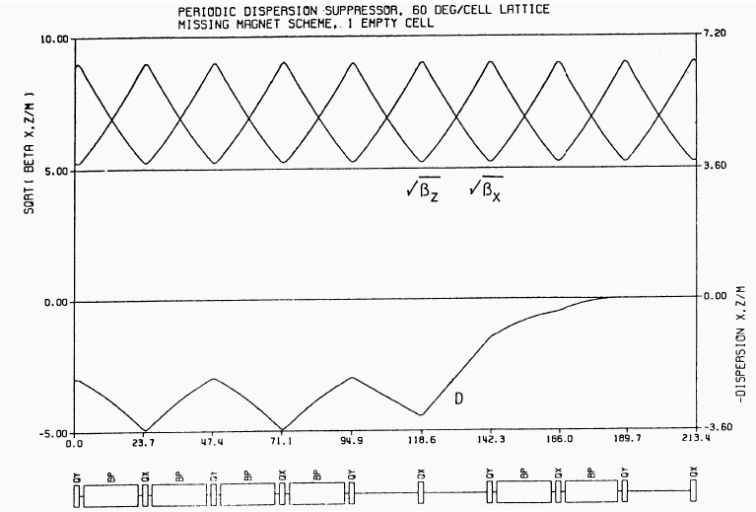

Figure 34: Illustration of the missing magnet dispersion suppressor, borrowed from Ref. [29].

5.2 Dispersion suppressor

At the interaction point (IP) of a collider we want the beam size as small as possible and recalling that it is given by $\sqrt{\varepsilon\beta + (D\Delta p/p)^2}$ we require the dispersion D to be zero at the IP. Since the dispersion can not be zero in the arcs, because it is generated in the bending dipoles, we need to design a piece of beamline at the intersection between the arcs and the interaction region where the dispersion is matched to zero. This section is called a dispersion suppressor.

Several methods are possible to match the dispersion to zero if at least six quadrupoles are available to independently adjust the six parameters β_x, α_x, β_y, α_y, D, D'. One method, called the missing-magnet scheme is illustrated in Fig. 34 where a periodic FODO lattice with bending dipoles between the quadrupoles is matched to $D = D' = 0$ at the right. The dispersion suppressor consists of one cell (between cosecutive QF quadrupoles) and one cell with dipoles. Provided the phase advance of the lattice is chosen suitably the two missing dipoles cause a perturbation that interferes to zero with the normal dispersion oscillations. See Ref. [29] for a more thorough discussion of dispersion suppressors.

5.3 Telescope and mini-beta

After the dispersion suppressor we have $D = D' = 0$ and are faced with the task to make the beams small. This is normally done very similar to light ray optics with a telescope. There is a particular difference due to the quadrupoles . They focus in one plane, but defocus in the other plane as we already discussed earlier. It is, however, possible to assemble quadrupoles in doublets that almost behave like

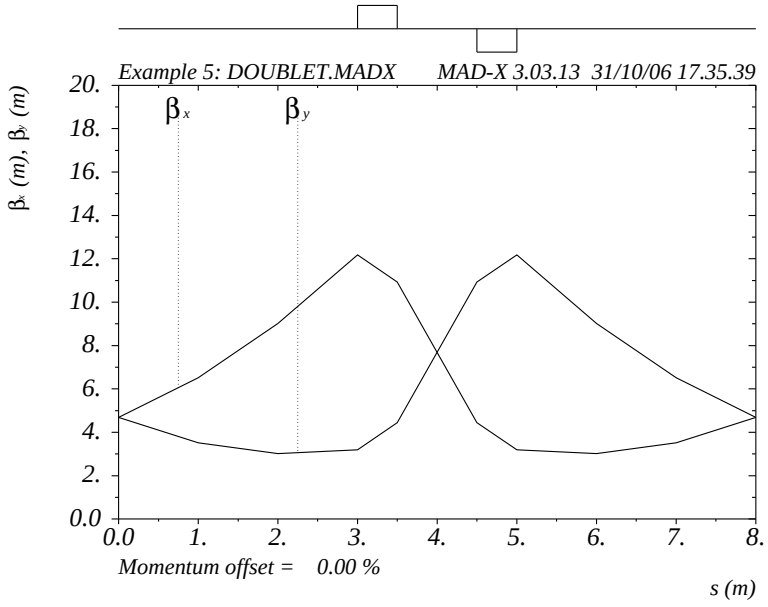

Figure 35: Illustration of a doublet and the mathched beta functions.

light lenses and focus in both planes. One example is a doublet, which consists of two quadrupoles close (compared to their focal length) to each other. The other example is a triplet with three quadrupoles in a row with excitation pattern $(1, -2, 1)$ which means that the center quadrupole has twice the excitation or is twice as long as the outer quadrupoles and the distance between the quadrupoles is short compared to the involved focal lengths.

Using two doublets or triplets with a distance between them it is possible to build an optical telescope that has the optical properties that it does point-to-point image of the image plane (near the arc) to the IP which means that the R_{12} and the R_{34} are made zero. Moreover, we also require that the imaging is parallel-to-parallel, which means that the R_{21} and R_{43} are zero. Consider a simple one-dimensional optical system as indicated in Fig. 36 where the beam comes from the right and first passes through a lens with focal length f_1 and then a distance of length $f_1 + f_2$ whence the beam passes the second lens with focal length f_2 and after another drift space of length f_2 it arrives at the IP. We can easily write down the transfer matrix for the first drift-lens-drift module with index 1 with the result

$$\begin{pmatrix} 1 & l_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f_1 & 1 \end{pmatrix} \begin{pmatrix} 1 & l_1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 - l_1/f_1 & 2l_1 - l_1^2/f_1 \\ -1/f_1 & 1 - l_1/f_1 \end{pmatrix} . \qquad (108)$$

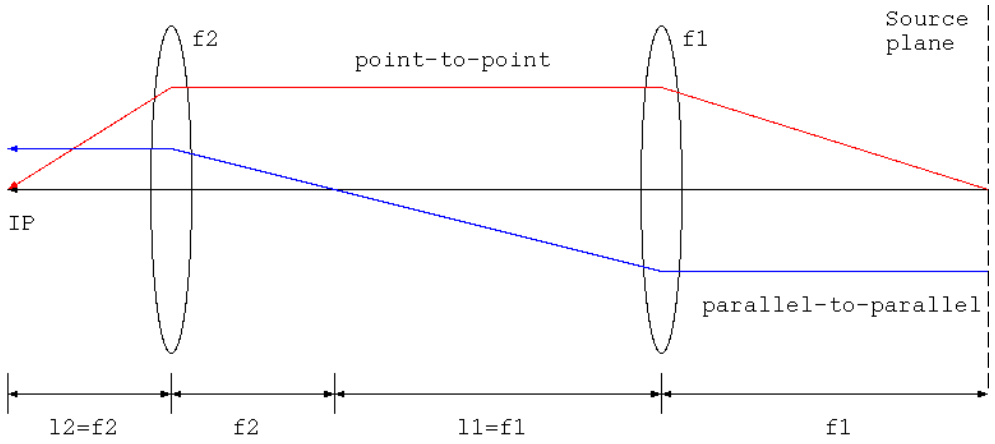

Figure 36: A telescope.

since we have chosen $l_1 = f_1$ the matrix simplifies to

$$\begin{pmatrix} 0 & f_1 \\ -1/f_1 & 0 \end{pmatrix} \tag{109}$$

and we also have a similar matrix for the second drift-quadrupole-drift system with index 2. Multiplying the two matrices with index 1 and 2 we arrive at the following transfer matrix R that represents the beam optical system between the source plane and the IP

$$R = \begin{pmatrix} -f_2/f_1 & 0 \\ 0 & -f_1/f_2 \end{pmatrix} \tag{110}$$

and we see that it describes a system that demagnifies the x coordinate by the factor $M = -f_2/f_1$ which means that the ratio of the focal lengths. The minus sign describes the inversion of a picture that is commonly encountered in normal telescopes. In practice we now have to realize such a system with quadrupole magnets and the matching module discussed in an earlier section is very handy to find the quadrupole values once the geometry, i.e. the magnet lengths and distances are decided upon.

Conceptually we thus find that the interaction region of a collider consists of a dispersion suppressor that often doubles up as a matching section to adjust the beta functions and a telescope that demagnifies the beam to very beam sizes which are required to reach high collision rates and thereby high count rates for the experiments.

5.4 SLC Final focus system

The situation in linear colliders such as the SLC, ILC or CLIC is very similar and we follow Ref. [30] to discuss the SLC final focus as a prototype system. The

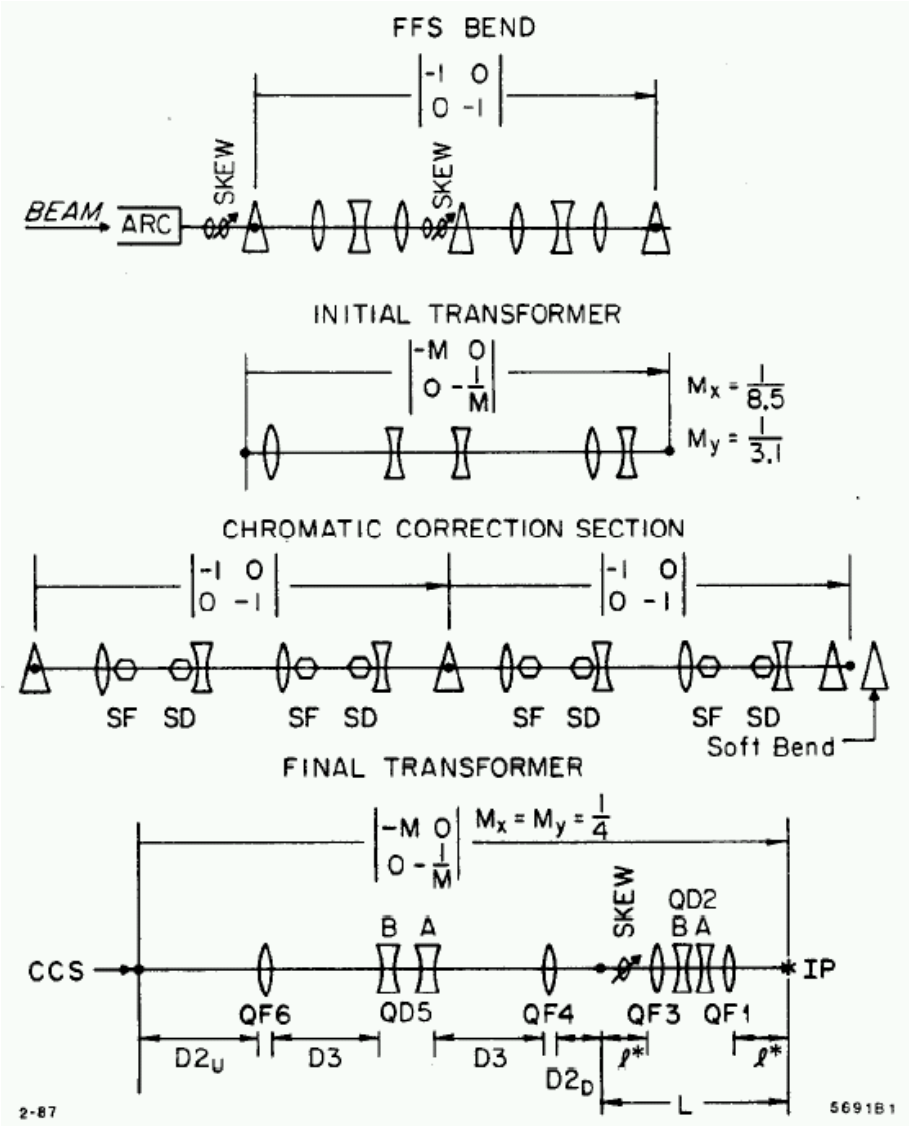

Figure 37: The conceptual layout of the SLC final focus, borrowed from Ref. [30].

conceptual layout is shown in Fig. 37. The first section, called the FFS-BEND, suppresses the dispersion from the arcs and matches the beta functions of the arcs to the telecopes that follow. There are also some skew quadrupoles to adjust the $x - -y$ coupling. The next section, called the INITIAL TRANSFORMER, is a telescope that demagnifies the beam size by a factor $M_x = 8.5$ in the horizontal plane and $M_y = 3.1$ in the vertical plane. The FINAL TRANSFORMER demagnifies by another factor of 4 in both planes. Especially the final triplet quadrupoles just upstream the IP are very strong in order to squeeze the beam to micron size at the IP. As a consequence the beam and the beta functions is very large as well. But strong quadrupoles together with large beta functions imply that the chromaticity is large. In the context of a linear collider this means that the focus for particles with different energies has a different distance from the last quadrupole and the beam size at the IP is diluted and enlarged. In order to compensate this chromatic effect due to the strong final focus quadrupoles we need a chromatic correction section (CCS) that is just upstream of the final transformer. In the CCS a little dispersion is generated by weak dipoles and sextupoles are placed at locations with non-zero dispersion. This allows correcting the chromaticity, because, as we discussed earlier, this produces effective momentum dpendent quadrupoles,which, properly adjusted will cancel the chromaticity of the real quadrupoles.

Sextupoles are -1 transform apart but families are interleaved. This cancels the geometric aberrations of the sextupoles to first order but the sextupoles of the different families mix and cause octupolar aberrations. In a later stage actually real octupoles were installed to compensate this. This I need to explain much better in the next version.

After discussing the interaction regions of colliders we will now turn to synchrotron radiation sources and their design concepts.

5.5 Double Bend Achromat

Storage ring based synchrotron radiation sources are usually designed to produce high intensity synchrotron radiation with a high degree of coherence as possible. This implies that as many electrons as possible should be forced into a small bunch in all three dimensions. This defines the the following criteria

- low transverse emittances,
- much space for undulators and wigglers with zero dispersion,
- high beam current,
- short bunches.

The latter two points are normally achieved by paying careful attention to the beam pipe and make it very smooth to avoid beam instabilities, as we will discuss

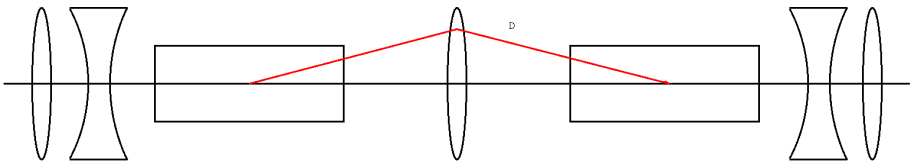

Figure 38: Basic layout of the double bend achromat.

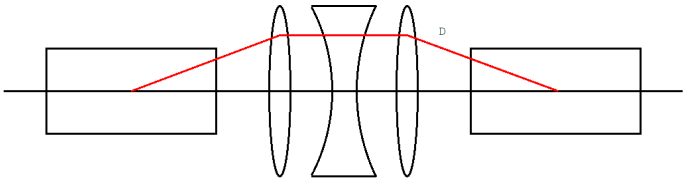

Figure 39: Basic layout of the triplet achromat lattice.

further below. Short bunches can be achieved to some extent by beam optical tricks but the normal way is to increase the RF voltage. This will become clear in the following chapter. Much space for insertion devices, as indulators and wigglers are commonly known is created by clustering the normal magnets dipoles and quadrupoles in groups with lang spaces inbetween.

The first requirement for small emittance defines the transverse beam size σ according to $\sigma^2 = \varepsilon\beta$ and, as we saw in section 4.16, the emittance is generated by the emission of photons at locations with dispersion. This was quantified in eq. 102 by the parameter $\mathcal{H}$ and we obviously want *small beta functions and small dispersion in the dipoles.* There are several ways to achieve this and one lattice type is the *Double Bend Achromat* (DBA) which has the layout shown in Fig. 38. One or several central quadrupoles between two dipoles is used to bend the disperison back before growing too large and the phase advance between the centers of the dipoles is roughly 180 degrees. the remaining quads are used to match the beta functions to small values inside the dipoles. An optimized version of this DBA lattice is also named Chasman-Green lattice. THis type of lattice is for example used in the ESRF in Grenoble.

A very thorough discussion of the emittance optimization of the DBA and other synchrotron radiation source lattice types can be found in Ref. [4].

5.6 Triplet Achromat lattice

A variant of the previous DBA lattice is the Triplet Achromat lattice, where three quadrupoles are placed inbetween the dipoles and all the beat matching is done by these three quadrupoles. These type of lattices can be made very compact and was realized in the small light source ACO in Orsay near Paris, which is,

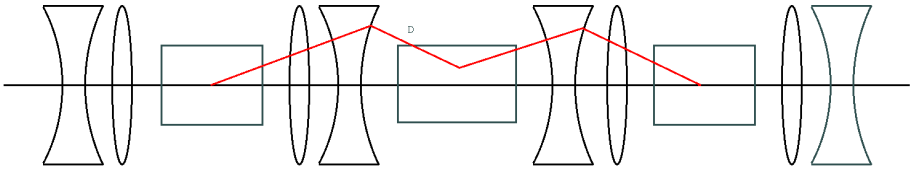

Figure 40: Basic layout of the triple bend achromat lattice.

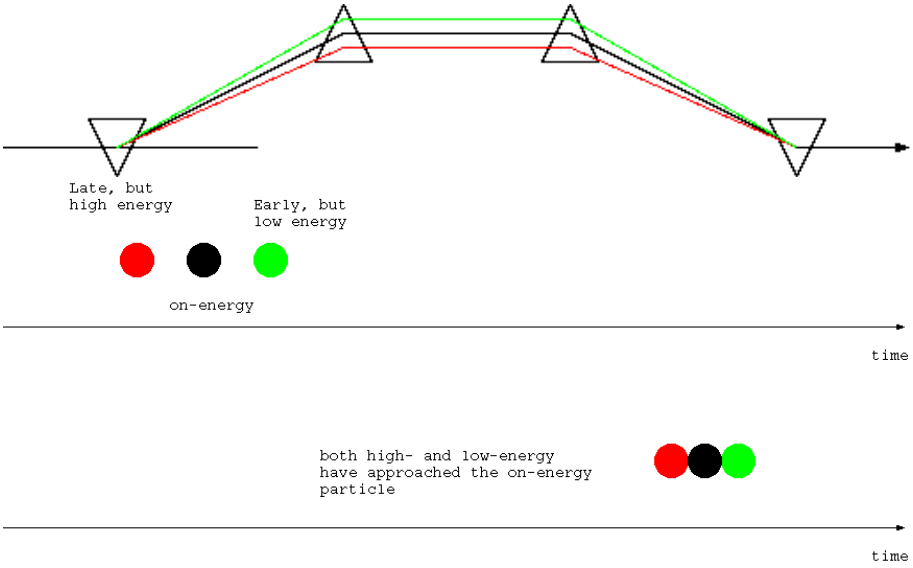

Figure 41: Basic layout of simple bunch compressor.

however, decommisioned.

5.7 Triple Bend Achromat

A further lattice type found in synchrotron light sorces is the so-called triple bend achromat which consists of a section with three dipoles and a number of quadrupoles. The latter make sure that the beta function is small in the dipoles.

All these lattices are optimized to have small beta functions and dispersion in the bending dipoles. We now turn to a special insertion that plays an important role in linear accelerators that desire to achieve very short bunches and corresponding high peak currents, which is especially the case for linacs used for Free-electron lasers such as the FLASH facility at DESY or LCLS at SLAC in Stanford.

5.8 Bunch compressor

In a bunch compressor we seek to reduce the bunch length at the expense of the momentum spread. This is achieved by accelerating the bunch off-crest in the RF-system. In this way the head of the bunch can be made to receive a lower energy than the tail of the bunch. We now have to produce a device that translates momentum difference into arrival time difference. A chicane is such a device where three dipole magnets wih bending angles ϕ, -2ϕ, and ϕ, respectively, are arranged as is shown in the upper plot of Fig. 41. The idea is to give particles with different energies different path lengths. In particular a particle with higher energy will be deflected less in the dipoles and will take a shortcut on the inside of he chicane, resulting in a shorter path length. This effect can be calculated by first considering the length of the unperturbed path

$$l = \frac{2L}{\cos\phi} \approx \frac{2L}{1-\phi^2/2} \approx 2L\left(1+\frac{\phi^2}{2}\right) \tag{111}$$

If the momentum offset is δ the bending angle will be reduced by $\phi \to \phi/(1+\delta)$ and we find

$$l(\delta) = 2L\left(1+\frac{\phi^2}{2(1+\delta)^2}\right) \approx 2L\left(1+\frac{\phi^2}{2}\right) - 2L\phi^2\delta \tag{112}$$

where we observe that the first term is equal to $l(0)$. For the R_{56} which describes the path length change as a function of momntum change we thus find

$$R_{56} = \frac{l(\delta)-l(0)}{\delta} \approx -2L\phi^2 \tag{113}$$

which is proportional to the total length of the chicane and the square of the bending angle ϕ.

5.9 Collimation

To be written

6 Longitudinal Dynamics

In this section we will discuss the motion of particles in the longitudinal phase space which consists of the arrival time $\tau = z/c$ where z is the longitudinal distance to the reference particle and $\delta = \Delta p/p$ the relative momentum deviation. The arrival time is determined by the speed of the particle (a small effect proportional to $1/\gamma^2$) and the path length in the magnetic structure. The latter effect we already discussed in section 4.14 about the momentum compaction factor α in a circular accelerator which describes the relative change in path length

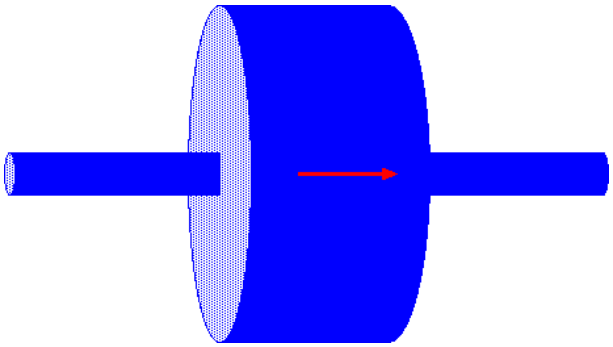

Figure 42: A pillbox cavity

or circumference $\Delta C/C$ with momentum deviation δ. Any effect or device that changes the energy or momentum of the particles will also affect its longitudinal dynamics. Of course RF-cavities that are used to apply longitudinal electric fields to the particle with the intent to accelerate of, as we shall see, focus the beam fall into this category. Other effects that will change the energy is, for example, the emission of synchrotron radiation or the energy loss in a target.

6.1 RF-cavity

The rapidly oscillating field in a radio-frequency cavities will affect the energy change ZeV where Ze is the charge of the particle and $V(t) = E_z(t)l$ is the longitudinal voltage that the particle experiences during the passage through the cavity. As an example of a cavity we show so-called pillbox cavity in Fig. 42. The particle beam will travel from the left to the right and will experience a longitudinal electric field, as indicated by the arrow. Depending on its arrival time it will experience a field pointing either in the direction of travel or anti-parallel. In the latter case the cavity would thus decelerate the particle. The longitudinal electric field component $E_z(t)$ will oscillate with a frequency f according to

$$E_z(t) = \hat{E}_z \cos(2\pi f t) \tag{114}$$

where $\hat{E}_z$ is the peak electric field in the cavity. The frequencies that are possible in a cavity are mostly determined by its geometry and we will in the next sections discuss several aspects.

The first aspect we consider depends on the fact that the cavity has a finite length l and the particle a finite speed βc and that the electric field changes during the passage of the particle. The reduction of the electric field experienced by a particle can be calculated by averaging the electric field in Eq. 114 over the transit time $\tau = l/\beta c$ and normalizing to the peak field with the result

$$\eta_{\rm ttf} = \frac{\sin x}{x} \quad \text{with} \quad x = \frac{\pi f l}{\beta c} \tag{115}$$

where $\eta_{\rm ttf}$ is called the *transit time factor* of a cavity. It is rarely explicitely mentioned when discussing cavities, but usually absorbed in the peak field.

As mentioned above are the electro-magnetic fields determined by the geometry of the cavity and we will consider a very simplified one which is circular with radius R and has length l. For simplicity we ignore the perturbation due to the entrance and exit hole for the beam pipe. Of course the dynamics of the electro-magnetic fields i governed by Maxwell's equations

$$\begin{aligned} \vec{\nabla} \times \vec{H} = \frac{\partial \vec{D}}{\partial t} + \vec{j} \ , \qquad \vec{\nabla} \cdot \vec{B} = 0 \ , \\ \vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \ , \qquad \vec{\nabla} \cdot \vec{D} = \rho \end{aligned} \tag{116}$$

with the additional equations that define the material properties

$$\vec{D} = \varepsilon\varepsilon_0 \vec{E} \quad \text{and} \quad \vec{B} = \mu\mu_0 \vec{H} \tag{117}$$

where ε is the relative permittivity and μ the relative permeability of the material. Moreover, in some materials Ohm's law is valid

$$\vec{j} = \sigma \vec{E} \tag{118}$$

where σ is the conductivity of the material.

Inside the cavity we can assume that we have vacuum and no sources $\rho = 0$ and $\vec{j} = 0$ and also $\mu = \varepsilon = \sigma = 0$. By multiplying the first of Maxwell's equations with $\vec{\nabla}\times$ and using the vector-analysis identity

$$\vec{\nabla} \times (\vec{\nabla} \times \vec{A}) = \vec{\nabla}(\vec{\nabla} \cdot \vec{A}) - \triangle \vec{A} \tag{119}$$

we arrive at the well-known wave equation

$$\triangle \vec{E} - \mu_0 \varepsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2} = 0 \tag{120}$$

and find that the propagation speed in vacuum is given by

$$c^2 = \frac{1}{\mu_0 \varepsilon_0} \ . \tag{121}$$

Since our cavity has cylindrical symmetry we write the Laplace-operator $\triangle$ in cylindrical coordinates (r, ϕ, z) and moreover look for a time-harmonic solution with frequency $\omega = 2\pi f$ which reduces Eq. 120 to

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \vec{E}}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 \vec{E}}{\partial \phi^2} + \frac{\partial^2 \vec{E}}{\partial z^2} + k^2 \vec{E} = 0 \tag{122}$$

with $k = \omega/c$. Note that the previous equation holds for every component of the Electric field vector $\vec{E}$. We now select to investigate the $z-$component $E_z(r, \phi, z)$ further and try to solve Eq. 122 by a separation ansatz $E_z = f(r)g(\phi)h(z)$ which we insert and obtain

$$\frac{1}{rf}\frac{\partial(rf')}{\partial r} + \frac{1}{r^2}\frac{g''}{g} + \frac{h''}{h} + k^2 = 0 \ . \tag{123}$$

Obviously h''/h only depends on z and all the other terms depend on r or ϕ or are constant. Therefore h''/h must be constant which we call $-k_z^2$ and therfore have

$$\frac{h''}{h} + k_z^2 = 0 \quad \text{or} \quad h''(z) + k_z^2 h(z) = 0 \tag{124}$$

which is solved by

$$h(z) = E_0 e^{\pm i k_z z} \ . \tag{125}$$

The remaining equation for f and g is given by

$$\frac{1}{rf}\frac{\partial(rf')}{\partial r} + \frac{1}{r^2}\frac{g''}{g} = -(k^2 - k_z^2) = -k_c^2 \tag{126}$$

where we implicitly define the cutoff wave-number k_c. This equation can be rewritten as

$$\frac{r}{f}\frac{\partial(rf')}{\partial r} + k_c^2 r^2 = \frac{g''}{g} \ . \tag{127}$$

Again, since the left hand side only depends on r and the right hand side only on ϕ each side must be constant independently with constant k_r^2 and we obtain the two equations

$$\begin{aligned} &r^2 f''(r) + rf'(r) + (k_c^2 r^2 - k_r^2) f(r) = 0 \\ &g''(\phi) - k_r^2 g(\phi) = 0 \ . \end{aligned} \tag{128}$$

The second equation is solved by exponentials or sine and cosine functions with argument $k_r\phi$ which, if we require periodicity in ϕ, implies that $2\pi k_r = 2\pi m$ or that k_r needs to be an integer. For $g(\phi)$ we then obtain

$$g(\phi) = e^{\pm i m \phi} \ . \tag{129}$$

We now insert $k_r = m$ into the first equation and make a variable substitution $s = k_c r$ with the result

$$s^2 f'' + sf' + (s^2 - m^2) f = 0 \tag{130}$$

which is just the defining equation for the Bessel functions of integer order [19], denoted by $J_{\pm m}(s)$ and the solution of the first equation is

$$f(r) = J_m(k_c r) \tag{131}$$

where we have substituted back the original variable r.

Collecting the solutions for f, g, and h we find the solution for the longitudinal electric field component

$$E_z(r, \phi, z, t) = E_0 J_m(k_c r) e^{\pm im\phi} e^{\pm ik_z z} e^{i\omega t} \tag{132}$$

in side the cavity. If we assume that the cavity boundaries are perfectly conducting, we have $E_z = 0$ at $r = R$ and this implies that $J_m(k_c R) = 0$ and we conclude that the Bessel-function has to have a zero on the surface of the cavity. If we denote the $n-$th zero of the Bessel function J_m by γ_{mn} we must have $k_c R = \gamma_{mn}$. For example, the mode with $m = 0$ therefore requires $k_c R = 2.405$ where 2.405 is approximately the first zero of J_0. The requirement that the electric field vanishes at $z = 0$ and at $z = l$ can be fulfilled by combining the two exponentials with argument $\pm ik_z z$ to a cosine with the same argument that vanishes at $z = 0$ and l whereby we get $k_z l = p\pi$ with integer p. Collecting these constraints for k_z and k_c we obtain a dispersion relation for the resonance frequencies

$$k^2 = \frac{\omega_{mnp}^2}{c^2} = \left(\frac{\gamma_{mn}}{R}\right)^2 + \left(\frac{p\pi}{l}\right)^2 \tag{133}$$

with integers m, n, and p. Obviously, if the geometry of the cavity is given through its radius R and length l the admissible frequencies are given by $f_{mnp} = \omega_{mnp}/2\pi$ in Eq. 133. Only these frequencies satisfy the boundary conditions that the field must vanish on the metallic boundaries.

We can turn the argument around and use Eq. 133 to design a cavity that has the desired frequency that we deem useful for our accelerator. A value for the frequency that is often chosen is 500 MHz and we now have to choose R and l suitably. We also require that 500 MHz is the fundamental mode, or the lowest possible eigenfrequency. We therfore pick $m = 0$ and $n = 1$, the first zero of the zeroth Bessel function. Solving Eq. 133 for R we obtain

$$R = \frac{2.405\lambda}{2\pi\sqrt{1 - (p\lambda/2l)^2}} = 0.287\,\mathrm{m} \tag{134}$$

where we introduced the wave length $\lambda = c/f = 0.6\,\mathrm{m}$. If we select $p = 0$ we find the relation between the radius and the wavelength

$$R = \frac{2.405\lambda}{2\pi} \tag{135}$$

for the mode characterized by $(mnp) = (010)$ and since it can be shown that this mode cannot have a non-zero longitudinal magnetic field component B_z it is called a TM_{010} (for transverse magnetic) mode.

In the previous paragraph we have focused on the fundamental TM_{010} mode, but there are, of course, other modes at higher frequency that can be excited

either by the beam or externally and perturb the beam at a later time. These higher order modes can provide a means by which the beam can 'talk to itself' and can cause a feedback-like interaction that often leads to an instability. But this aspect we will discuss in a later section on instabilities.

6.2 Cavity losses

In the previous section we worked out the electro-magnetic oscillation modes that are admissible in a resonant cavity under the assumption that the walls of the cavity are perfectly conducting and that no field penetrates into the material. In a real cavity, however, the walls have finite conductance and the fields penetrate into the material where they can dissipate energy. The field intensity therefore diminishes and energy is lost by heating the cavity walls. The losses are proportional to the field intensity and can be modelled by a friction coefficient that can be written by

$$\frac{dU}{dt} = -\frac{\omega}{Q}U \tag{136}$$

with the quality factor Q which intuitively describes how many oscillation cycles it takes to reduce the energy by $1/e$. The detailed calculation of the $Q-$value of a cavity in terms of the geometry and the conductance is discussed in Ref. [21].

The electric field in the cavity that provides the acceleration voltage $U_z = E_z l$ is excited by an external current I via, for exmaple, a small antenna and the electro-magnetic fields inside the cavity behave like a damped resonator circuit we can write $U_z = ZI$ with

$$Z(\omega) = \frac{R_s}{1 + iQ\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)} \tag{137}$$

in resemblance to Ohm's law. A current provides a voltage and the proportionality constant is resembles a generalized impedance $Z(\omega)$ that depends on the frequency. The way the impedance Z is written is just a convenient way to express the response function $U_z(\omega)/I(\omega)$ of forced and damped harmonic oscillator

$$U_z'' + 2\alpha U_z' + \omega_0^2 U_z = \omega_0^2 R_s I \sin(\omega t) \tag{138}$$

with the shunt impedance of the cavity R_s and $\alpha = \omega_0/2Q$. The description of a RF-cavity in terms of impedances will also become very useful in later sections about instabilities.

6.3 Phase Stability and Synchrotron Oscillations

In a circular accelerator it is obvious that a RF-cavity has to operate at a multiple of the revolution frequency in order to avoid that just a sequence of quasi-random kicks is applied to the particle. This integer number is called the *harmonic number*

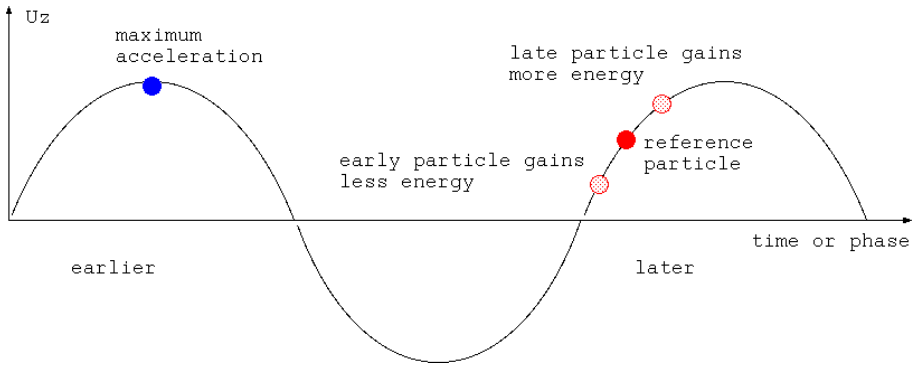

Figure 43: Cavity gain

of the RF-system and is given by $h = f_{RF}/f_0$ where $f_0 1/T_0$ is the revolution frequency and T_0 is the revolution time. In a linear accelerator the relative phase of the cavities has to be adjusted to the distance between cavities and the speed of the particles.

The design phase ϕ_d of the RF-cavity that the reference particle should experience is determined by external requirements, such as to achieve maximum acceleration which is indicated by the particle located at the crest of the oscillation in Fig. 43. In other accelerators the design phase can be determined by the requirement to replenish the losses experienced for example by the emission of synchrotron radiation or the interaction with a target. For definiteness sake we consider for the moment a storage ring where the beam looses energy U_l per turn, which must be replenished in the RF system. The design phase ϕ_d is then given by

$$U_l = e\hat{V}\sin\phi_d \tag{139}$$

where $\hat{V}$ is the peak acceleration voltage that a particle can experience in the cavity. Another particle arriving at a slightly different time τ from the reference particle that arrives at phase ϕ_d will experience the following energy gain in the cavity

$$\Delta E = e\hat{V}\left[\sin(\phi_d + \omega_{RF}\tau) - \sin\phi_d\right] \ . \tag{140}$$

This situation is illustrated in Fig. 43 where the reference particle is located at a given phase and an early and a late particle are also shown that experience a energy kick with respect to the reference particle given by the previous equation. We note that it is convenient to use the phase variable ϕ instead of arrival time τ. These variables are related by $\phi = \omega_{RF}\tau$.

In order to find equations of motion for the longitudinal dynamics we have to relate the energy change ΔE to the relative momentum deviation $\delta = \Delta p/p$ which is given by some manipulations involving relativistic kinematics

$$\frac{\Delta E}{E} = \beta^2\frac{\Delta p}{p} = \beta^2\delta \ . \tag{141}$$

Finally we make the assumption that the energy chnage that the particle experiences in a cavity is small compared to its energy which means that we can replace a derivative with respect to time d/dt by averaging the change over one revolution period T_0. For the temporal variation of δ we then find

$$\frac{d\delta}{dt} \approx \frac{1}{T_0\beta^2}\frac{\Delta E}{E} = \frac{e\hat{V}}{T_0\beta^2 E}\left[\sin(\phi_d+\phi) - \sin\phi_d\right] \tag{142}$$

which relates the change in relative momentum per revolution period $d\delta/dt$ to the arrival time τ, here expressed in terms of the phase variable $\phi = \omega_{RF}\tau$.

The time difference with respect to the reference particle that it takes for a particle with momentum deviation $\delta = \Delta p/p$ to travel around the accelerator is given by

$$\frac{\tau}{T_0} = \frac{1}{T_0}\left(\frac{C(\delta)}{v(\delta)} - T_0\right) = \left(\alpha - \frac{1}{\gamma^2}\right)\delta = \eta\delta \tag{143}$$

where we used that the circumverence is dependent on the momentum of the accelerator as given my the moemntum compaction factor α and that the speed v also depends on the momentum deviation δ. Eq. 143 describes the variation of the arrival time τ. If we multiply this equation with ω_{RF} we observe that it describes the change of the phase over one revolution period. If we again use the assumption that the variables τ or ϕ and δ change slowly we can again replace the temporal derivative d/dt by an abverage over the revolution period T_0 and can write

$$\frac{d\phi}{dt} \approx \frac{\omega_{RF}\tau}{T_0} = \omega_{RF}\eta\delta \tag{144}$$

which relates the changes in the phase to the relative momentum deviation δ. Together with Eq. 142 this describes the longitudinal dynamics of a particle under the influence of a cavity and the phase slip it encounters when traversing the accelerator. Average losses are accounted for by the design phase ϕ_d.

As an aside we observe that Eq. 143 which defines η can be also be rewritten in the following way

$$\frac{\Delta T}{T} = -\frac{\Delta f}{f} = \left(\alpha - \frac{1}{\gamma^2}\right)\delta = \eta\delta \tag{145}$$

which implies that a small change in momentum (or energy) of the particle δ causes the revolution time to change a little bit, which corresponds to the negative change in revolution frequency of that particle.

Returning to Eq. 142 and 144 we observe that we can differentiate the latter equation once more with respect to time and insert $\dot{\delta}$ from Eq. 142 with the result

$$\ddot{\phi} - \frac{\omega_{RF}\eta}{T_0\beta^2}\frac{e\hat{V}}{E}\left[\sin(\phi_d+\phi) - \sin\phi_d\right] = 0\ . \tag{146}$$

We observe that for small phase deviations ϕ the previous equation can be rewritten by using $\sin(\phi_d + \phi) - \sin\phi_d \approx \phi\cos\phi_d$ which allows us to express Eq. 146 in the following form

$$\ddot{\phi} + \Omega_s^2\phi = 0 \tag{147}$$

where we introduce the synchrotron frequency Ω_s by

$$\Omega_s^2 = -\frac{\omega_{RF}\eta\cos\phi_d}{T_0\beta^2}\frac{e\hat{V}}{E} \tag{148}$$

and which is valid for small oscillations. Equation 147 is the differential equation for a harmonic oscillator and describes oscillations with angular frequency Ω_s. These oscillations in the phase-momentum phase space are called *synchrotron oscillations.* The motion of the phase and momentum deviation δ will be according to $\phi = \hat{\phi}\sin\Omega_s t)$ and $\delta = \hat{\delta}\cos\Omega_s t)$ where $\hat{\phi}$ and $\hat{\delta}$ are the maximum amplitudes of the phase and the momentum deviation. Note that $\hat{\phi}$ and $\hat{\delta}$ are related by $\hat{\phi} = (\omega_{RF}\eta/\Omega_s)\hat{\delta}$.

Note that there are only stable oscillations if Ω_s^2 is positive, which implies that $-\eta\cos\phi_d$ must be positive. Considering the definition of the phase slip factor $\eta = \alpha - 1/\gamma^2$ in terms of the momentum compaction factor α and the kinematic factor γ we see that for low energies the $\gamma-$factor dominates and makes η negative which implies that the $\cos\phi_d$ must also be positive. The energy for which η is zero is called *transition energy*, and at low energies we are, colloquilally speaking, below transition, and at energies with $1/\gamma^2 \leq \alpha$ above transition. In the latter case we find that $\cos\phi_d$ must be negative and consequently the phase must be around 180 degrees, near the point where the RF voltage crosses zero from positive to negative voltages. In a real ring the particles will assemble at the design-phase where the motion is stable: below transition around zero phase and above transition around 180 degrees.

In the previous paragraphs we discussed small amplitude oscillations and here we will drop this approximation, but will assume for the sake of simlifying the algebra that we have no losses U_l and therefore the design phase ϕ_d is either zero or 180 degree. Equation 146 in this case simplifies to

$$\ddot{\phi} + \Omega_s^2\sin\phi = 0 \tag{149}$$

which is formally identical to Eq. 147 if the oscillation amplitudes are small. In the general case Eq. 149 is the equation of a mathematical pendulum and we can exploit the theory that was developed.

We start by noting that the equation of motion has an integral of motion because multiplying Eq. 149 with $\dot{\phi}$ yields

$$0 = \dot{\phi}\ddot{\phi} + \Omega_s^2\dot{\phi}\sin\phi = \frac{d}{dt}\left[\frac{1}{2}\dot{\phi}^2 - \Omega_s^2\cos\phi\right] \tag{150}$$

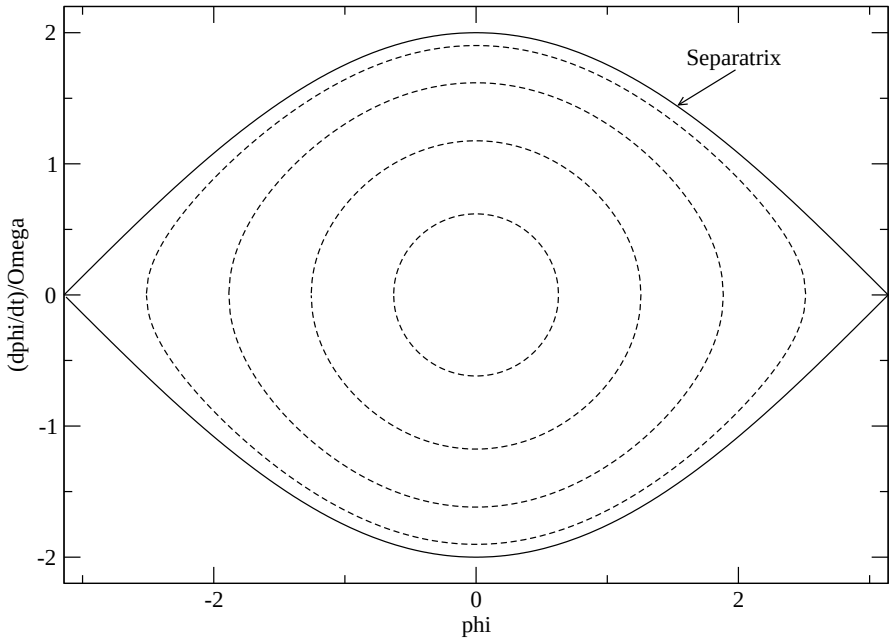

Figure 44: Phase portrait of the longitudinal phase space.

which means that the quantity in the square bracket is constant and does not change with time t. This leads to

$$\frac{1}{2}\dot{\phi}^2 - \Omega_s^2 \cos\phi = A \ . \tag{151}$$

The constant A we can express in terms of the maximum phase excursion $\hat{\phi}$ because at $\phi = \hat{\phi}$ we have $\dot{\phi} = 0$ and get $A = -\Omega_s^2 \cos\hat{\phi}$. Inserting into the previous equation yields

$$\frac{1}{2}\dot{\phi}^2 + \Omega_s^2(\cos\hat{\phi} - \cos\phi) = 0 \ . \tag{152}$$

What we have doe here is that we characterized the possible trajectories that the equation of motion describes by the maximum amplitude in much the same way as we can distinguish different oscillations of a children's swing by their maximum amplitude. A alternative way would be to describe the trajectories by their total energy.

We can now solve Eq. 152 for $\dot{\phi}$ and plot it as a function of ϕ and obtain

$$\dot{\phi} = \Omega_s \sqrt{2(\cos\phi - \cos\hat{\phi})} \tag{153}$$

and plot $\dot{\phi}$ with $\hat{\phi}$ as parameter that assumes values of $\hat{\phi} = \pi/5, 2\pi/5, \ldots, \pi$. We show the plots in the $\phi, \dot{\phi}$ phase plane in Fig. 44 and observe that for small maximum phase amplitude $\hat{\phi}$ the phase space is elliptical, but as the amplitude increases, the phase space is increasingly distorted up to a limiting curve for $\hat{\phi} = \pi$ which is called the *separatrix* and is the limit of stable oscillations.

Note that $\dot{\phi}$ is related to the energy deviation $\delta = \Delta p/p$ by Eq. 144, which implies that the vertical axis in Fig. 44 is actually just a rescaled δ where the rescaling is given by $\dot{\phi} = \eta\omega_{RF}\delta$. This implies that there is a *maximum momentum acceptance* δ_{max}, namely the height of the separatrix which is often called the *bucket half-height.* It is given by

$$\delta_{max} = \frac{2\Omega_s}{\eta\omega_{RF}} = 2\sqrt{\frac{1}{\eta\omega_{RF}\beta^2 T_0}\frac{e\hat{V}}{E}} \tag{154}$$

which clearly shows that the momentum acceptance is proportional to the root of the cavity voltage $\hat{V}$. If a particle should for some reason such as badly injected or having received an anomalously large energy loss in the interaction with a target should lie outside the separatrix, it will not be able to perform stable synchrotron oscillations.

We found that the particles with small phase angles ϕ perform harmonic synchrotron oscillations. This will, however, change as the amplitude increases and the motion becomes increasingly anharmonic. To illustrate this we calculate the oscillation period T_p as a function of the amplitude by rearranging Eq. 153 in the following way

$$\begin{aligned} \Omega_s \int_0^{T_p/4} dt &= \frac{\Omega_s T_p}{4} = \int_0^{\hat{\phi}} \frac{d\phi}{\sqrt{2(\cos\phi - \cos\hat{\phi})}} \\ &= \frac{1}{2}\int_0^{\hat{\phi}} \frac{d\phi}{\sqrt{(\sin^2(\hat{\phi}/2) - \sin^2(\phi/2)}} \end{aligned} \tag{155}$$

where we calculate the time to get from one zero to the extreme phase $\hat{\phi}$ which constitutes a quarter an oscillation period $T_p/4$. The integral on the right hand side can be brought into the standard form of a complete elliptic integral $K(x)$ by the substitution $\sin\psi = \sin(\phi/2)/\sin(\hat{\phi}/2)$ whence we arrive at

$$\frac{\Omega_s T_p}{2} = 2\int_0^{\pi/2} \frac{d\psi}{\sqrt{1 - \sin^2(\hat{\phi}/2)\sin^2\psi}} = 2K(\sin(\hat{\phi}/2)) \ . \tag{156}$$

For the oscillation period we the obtain

$$T = \frac{4}{\Omega_s}K(\sin\hat{\phi}/2) = T_{s0}\left(\frac{2}{\pi}\right)K(\sin\hat{\phi}/2) \tag{157}$$

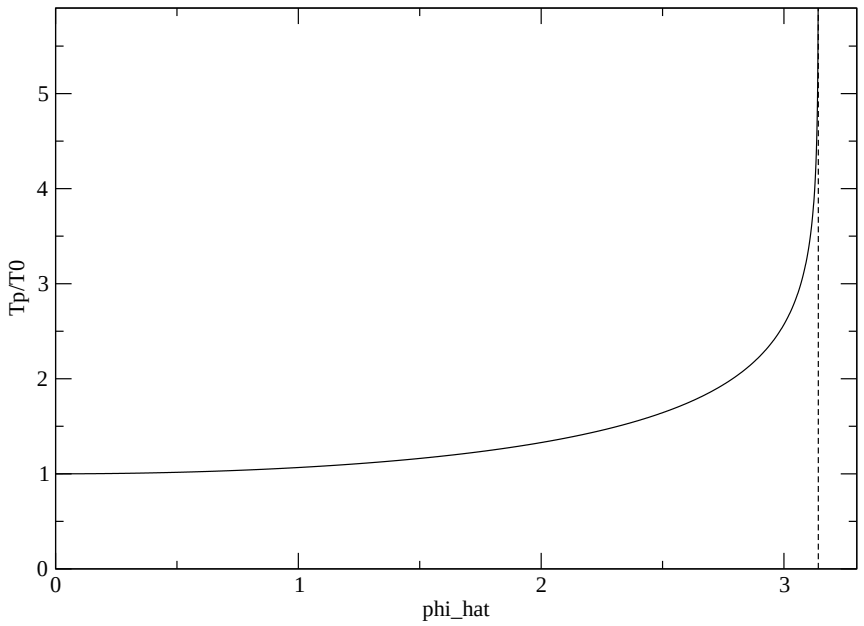

Figure 45: The synchrotron period.

where we define the small oscillation amplitude revolution period T_{s0}. We can use the series expansion for the elliptic function [19]

$$K(x) = \frac{\pi}{2}\left[1 + \frac{x^2}{4} + \frac{3x^4}{8}\right] \tag{158}$$

and for small oscillation amplitudes we find for the amplitude dependence of the revolution period

$$T \approx T_{s0}\left(1 + \frac{\hat{\phi}^2}{16} + \frac{3\hat{\phi}^4}{128} + \ldots\right) \ . \tag{159}$$

For the entire range of values $0 \leq \hat{\phi} \leq \pi$ we plot the revolution period T_p normalized to the small amplitude period T_{s0} as a function of the amplitude $\hat{\phi}$ in Fig. 45. Note that the osillation period diverges as the amplitude $\hat{\phi}$ approaches π which corresponds to a starting phase close to one of the nodes of the separatrix.

This concludes the discussion of the longitudinal phase space and we now turn to the discussion of the diagnostic equipment that is used in accelerators to find out how much beam is where and how big.

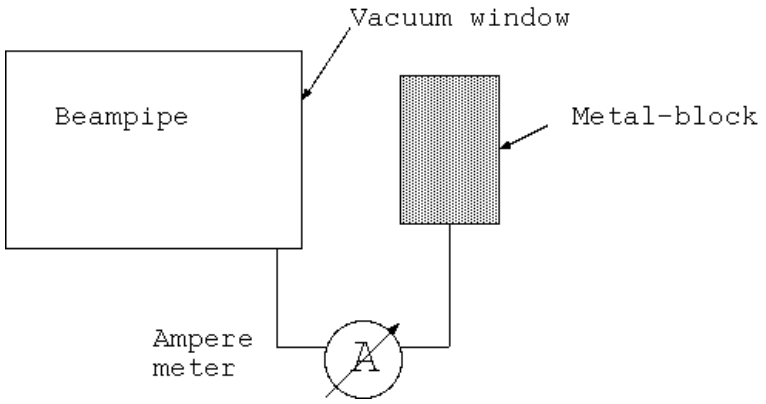

Figure 46: Conceptual schematic of a faraday cup.

7 Diagnostics

We now discuss various methods that are usually emplyed in accelerators to measure properties of the beam, which is required to optimize the performance and sometimes find out what is wrong with the accelerator.

7.1 Zeroth Moment: Current

The simplest method to measure the beam current, provided the beam power is suffciently low, is to just dump it in a metal block that is connected via an Ampere-meter to ground. If positively charged ions are diagnosed electrons from the absorber material will rush towards the ions and try to compensate their charge, thus constituting an electronic current that can be measured. Just using a metallic block is not very accurate and the more refined version that shields secondary electrons by magnetic fields and separates the absorber from the accelerator vacuum is called a *Faraday-cup.* Careful attention must be paid to stray capacitances and inductances, if the device must be fast, meaning measure rapid changes in the current. At higher beam currents and beam powers the absorber must be cooled and one must make sure that the beam is really stopped inside the absorber. This type of beam current measurement device is, since it stops the beam, by its nature, invasive.

Another device to measure the current is a *wall-gap monitor* shown in Fig. 47 which is based on the measurement of the image currents in the beam pipe. The vacuum chamber shields the magnetic fields that the beam generates such that no magnetic field is visible outside. On the other hand, inside the vacuum chamber the magnetic field is given by Ampere's law

$$\oint \vec{B}\vec{dl} = \mu_0 I_{beam} \tag{160}$$

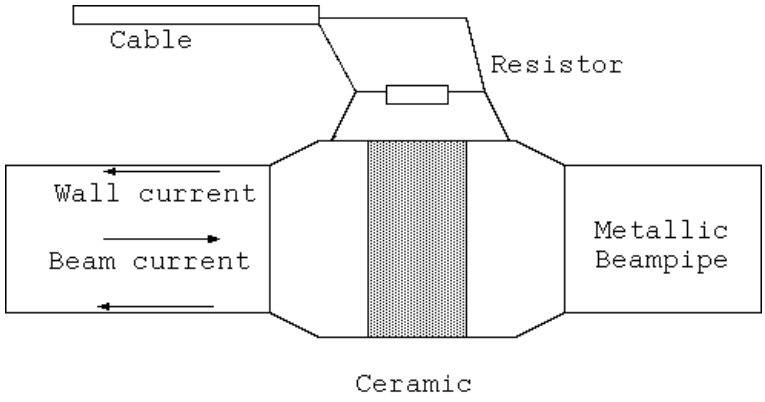

Figure 47: Schematic of a wall-gap monitor.

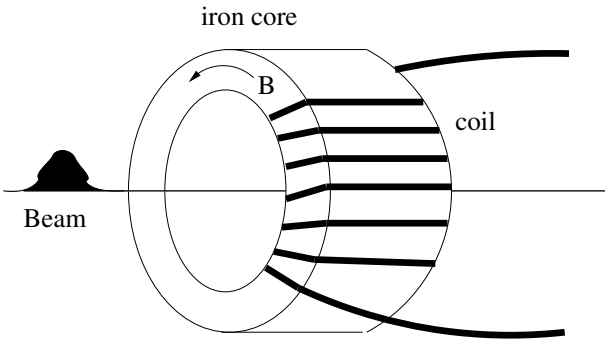

Figure 48: Schematic of a beam current transformer.

which implies that in the vacuum chamber a current propagates, which is equal in magnitude but has opposite sign. This wall current can be measured by inserting a ceramic in the beam pipe and bypassing the ceramic with resistors such that the wall currents are forced to travel through the resistors. Across the resistors a voltage drop develops that can be measured with a sensitive voltmeter. Provided that careful attention is paid to shielding and electrical design very fast signals into the GHz range can be resolved which corresponds to a time resolution on the pico-second scale. Note that a wall-gap monitor does not intercept the beam and is therefore a non-invasive current measurement system.

In the wall-gap monitor we picked up the image currents with a resistor, but we can also directly detect the magnetic field of the beam by placing a circular iron core around the beam inside the beam pipe. A passing bunched beam induces a temporally varying flux inside the iron that can be detected by winding several windings around the iron which act as secondary windings of a transformer, hence the name *current transformer*. The device is illustrated in Fig. 48.

Note that the current transformer detects the magnetic field due to a tempo-

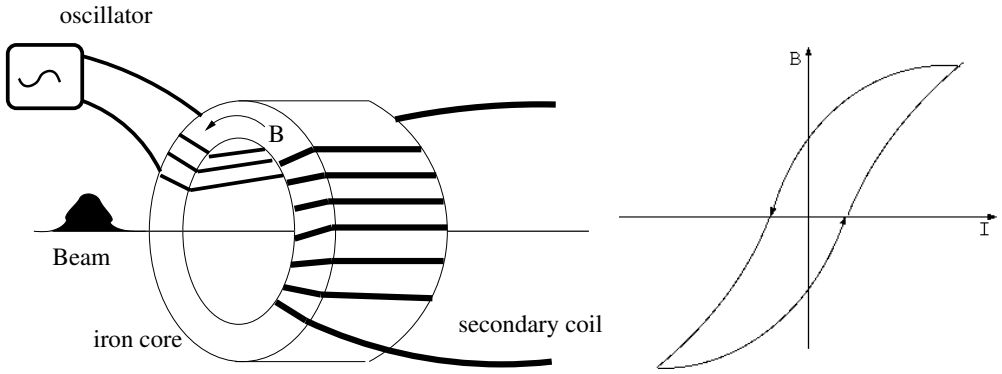

Figure 49: Schematic of a DCCT and a hysteresis curve.

rally changing beam current, which works well for bunched beams, but does not work for circulating beams that are un-bunched, so-called coasting beams. This type of situation is common in nuclear physics rings that are equipped with an electron cooler. A current measurement device that is able to measure the average or DC component of the current is called a *direct current charge transformer* or *DCCT*. The operation method is illustrated in Fig. 49. Similar to the normal beam transformer discussed in the previous paragraph an iron ring is encirceling the beam. There are two coils wound onto the ring. The first one is coupled to an oscillator which excites the ring such that the magnetic field inside the ring is driven into saturation. The voltage that is induced in a second winding around the ring will contain not only the driving signal from the oscillator, but also harmonics at higher frequencies. Since the hysteresis curve, also shown in Fig. 49, is antisymmetric normally only odd harmonics are generated. If, on the other hand, an extra beam current is passing through the ring, the hysteresis curve will be shiften up or downwards and will no be asymmetric any more. Therefore even harmonics are generated in the secondary coil. If a second wire, carrying a DC current, is passed through the ring, and adjusted as to cancel the second harmonic, we can very acurately measure the compensating current. The mesurement is thus based on a null-measurement which is typically very accurate. Note, however, that the bandwidth of the DCCT is limited by the exciting frequency of the driving oscillator.

We point out that a more thorough discussion of the transformers can be found in Ref. [3].

7.2 First Moment: Position

We now turn to the discussion of devices that allow to measure the transverse position of the beam which is of paramount importance for the operation of accelerators because a 'wrong' beam position is normally one of the first and

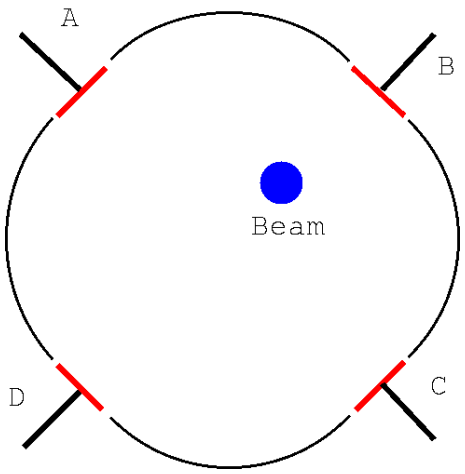

Figure 50: Button beam position monitor.

most easily recognized indicator if something is wrong. Furthermore, often only the desired performance is only achievable if the beam is well-centered in the magnets. Under some circumstances highly accurate measurements are desirable that determine the beam position to an accuracy within μm or better, and in other circumstances the position must be measured on a turn by turn base, which is rather demanding, especially for small rings with a high repetition rate. Another extreme requirement, that may occur, is the desire to measure the positions of two counter-propagating beams near the interaction point of a collider. This requires direction sensitive devices. Since we normally want to use the beam in an experiment, rather than use it in the measurement procedure itself, we need non-invasive diagnostics. This implies that either the wall-currents that we discussed earlier in the context of a wall-gap monitor or the electro-magnetic fields that accompany the beam will be analyzed and the beam position deduced.

The first BPM we discuss is a a so-called *Button BPM* of which a schematic is shown in Fig. 50. There are usually four small isolated buttons mouted in the vacuum chamber that pick up part of the wall currents. The isolated button appears like a capacitor and a displacement current flows into it. The BPM therefore only measures the AC component of the beam current which makes it especially suitable for short bunches that carry a high number of harmonics. The positions can be deduced from the signals in the four buttons in the following way

$$x = k_x \frac{(A+D)-(B+C)}{A+B+C+D} \quad , \quad y = k_y \frac{(A+B)-(C+D)}{A+B+C+D} \tag{161}$$

where we have $k_x = k_y = R/\sqrt{2}$ if the beam pipe is circular and the size of the buttons is small compared to the radius R of the pipe. The latter requirement implies that the signal induced in the buttons is rather small, because only a

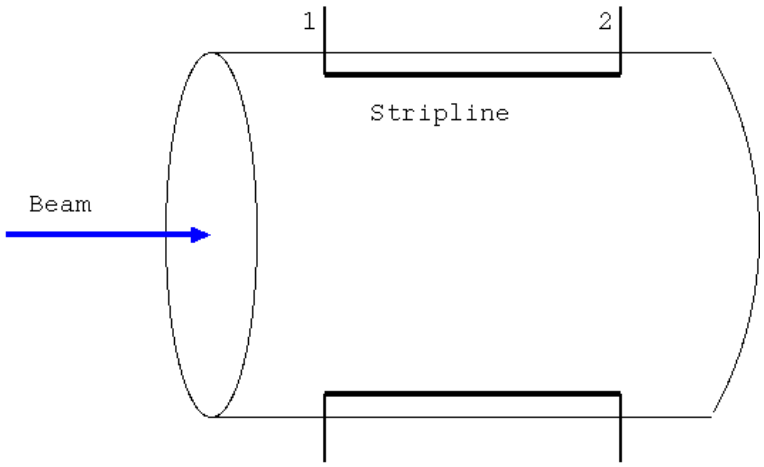

Figure 51: Schematic of a stripline beam position monitor.

small portion of the wall currents are intercepted and this makes these BPM more suitable for high intensity beams. Button BPMs normally are the preferred choice for synchrotron radiation sources, due to the requirement for short bunches and high intensity.

The proportionality constant is easily calculated by noting that the wall current density in the wall is proportional to the electric field of the bunch on the beam pipe wall. One may envision this by the longitudinally compressed electric field pushes the charges in the wall back and forth, while the beam passes. The electric field on the pipe wall can be easily calculated by the method of image charges and by conformal mapping.

Exercise: calculate the coefficients k_x and k_y in Eq. 161 for a round and an infinitely wide horizontal beam pipe.

The second BPM that we consider is a *stripline BPM* of which a schematic is shown in Fig. 51. A conducting strip is mounted isolated on the inside of the vacuum chamber and the magnetic field of the passing beam changes the flux in the area between the beam pipe and the stripline, which induces a voltage in the stripline. If the bunch length is shorter than the length of the stripline two pulses, one positive and one negative, can be picked up at eiher terminal of the stripline. This allows to distinguish signals from couter-propagating beams. The signals from four striplines arranged in much the same way as in the button BPM allows to extract the position information from these BPM.

Another technique is based on placing a diagonally sliced box inside the beam pipe which has the advantage that the sense-electrodes are particularly large, making these BPM suitable for ion-storage rings, that often operate at very low intensities. The position is then deduced from the difference of the signals from one and the neighboring electrode and normalized to the sum. A picture of such a

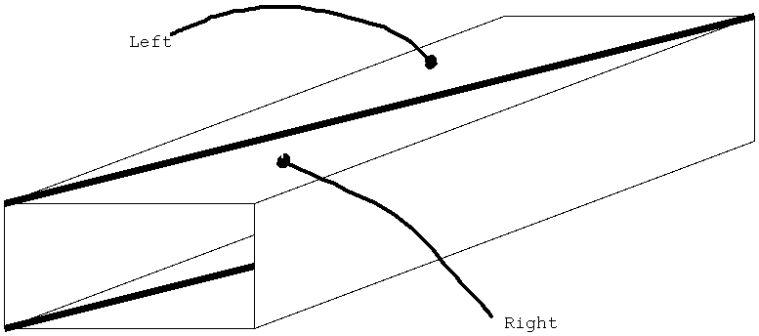

Figure 52: Shoebox beam position monitor.

shoebox bpm is shown in Fig. 52. Normally one box, that can have a rectangular cross-section, as shown, or a round cross-section, is used for the horizontal and one for the vertical direction.

A fourth position sensotive device is a *cavity BPM* which is based on a resonating structure that supports a mode which is spatially anti-symmetric and that has a zero on the beam axis, for example a TM_{210}-mode, as is shown in Fig. 53. If the beam is centered it can not excite the shown mode, but if it is off center the charges in the beam can work agains the electric field and thereby excite the mode. The shown mode can then excite the electrons in the antenna shown on the bottom, providing a means to couple out the signal. The information whether the beam is too high or too low can be obtained from phase comparing the signal from the antenna with a reference signal.

7.3 Second Moment: Beam Size

We now turn to the discussion of devices to measure the transverse and longitudinal size of the beam. Of course the most straightforward way of measuring the transverse beam size is by placing a luminescent screen into the beam and observing the emitted light with a video camera and digitizing the picture on a computer as is shown in Fig. 54. This is obviously an invasive measurement, especially at low energies where the beam is absorbed in the screen. There are often problems with these type of screens due to blind spots where the screen material is not as effective as on the surrounding area. Limited dynamic range and saturation of the intensity are other problems as well as cameras failing due to high radiation doses.

A variant of the luminescent screen are screens based on *optical transition radiation* (OTR) where the screen is made of a very thin foil of, for example AlO_2, which disturbs some beams, such as high energy electron beams, very little. The screen has an index of refraction larger than unity which causes light to travel slower as the speed of light in vacuum such that it can match the

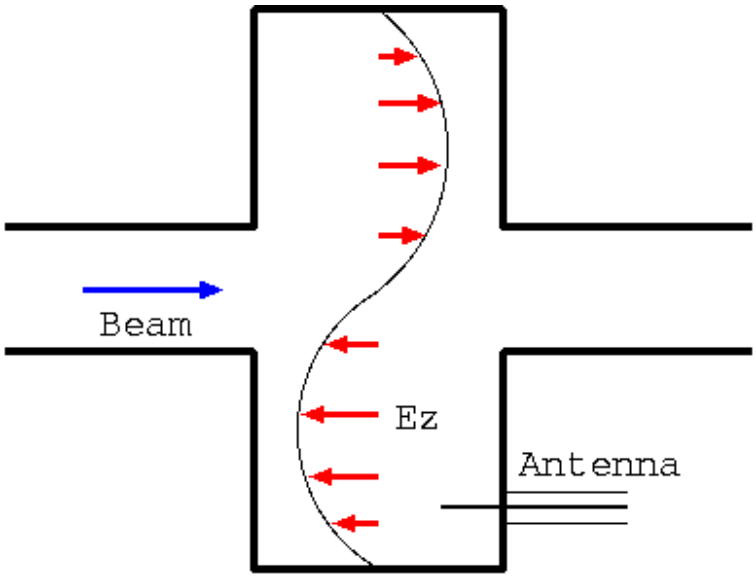

Figure 53: Cavity based beam position monitor.

also slowed-sown beam particles. Therefore electro-magnetic waves can couple to the beam particles. This is an effect very similar to the emission of Cerenkov radiation. The emitted radiation is then picked up by a camera, possibly after being transported in an optical beam line made of lenses and mirrors.

Another monitor is based on the ionization of very thin, typically a few μm, thick wires, as illustrated in Fig. 55, which is called a *secondary emission monitor* (SEM) grid. The electrons that are knocked out of the wire by the intercepted beam cause a small current to flow in the wires that is first amplified and then measured in an ampere-meter. Of course all wires must be read out in parallel which requires one amplifier per wire and all of them must have be well-balanced. Moreover, a large number of wires must be passed from the inside of the beam pipe, where the SEM-grid is located to the outside, where the amplifiers are located, which requires advanced vacuum feedthroughs. All of this make SEM grids rather expensive apart from the mechanical problems due to heat deposition in the wires that can lead to melting. The result of SEM grid measurements are normally displayed as a histogram of the current in the wires as a function of the wire position, showing a transverse beam profile. The center and the width of the distribution in the histogram then represent the beam position and width.

Instead of having a large number of wires permanently in the beam we can use a *wire-scanner* which consists of a fork that moves a thin wire through the beam. In this case only a single amplifier for the wire is used, but some mechanical devices such a stepper-motors or pneumatic pistons are required to move the wire. Of course, the position of the wire must be known while it traverses the beam which is usually done by resistive or optical position encoders. Often it is

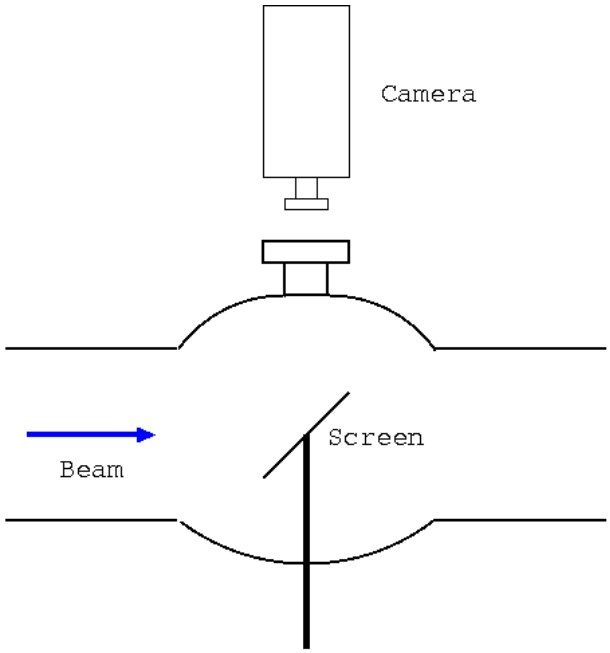

Figure 54: Schematics of the operating principle of a luminescent screen.

possible to avoid electronic readout of the secondary emission current from the wire but observe nearby fast ion chambers for radiation protection which will generate a signal related to the number of beam particles hitting the wire. It should also be noted that the wires are normallyb only several micrometer thich but nevertheless represent a thick target for the beam and perturb it which can be critical especially for circular accelerators.

A variant of the wire scanner is the *Magnesium Jet profile monitor* developed in Novosibirsk for VEPP-3 but a copy was installed in CELSIUS. In the Magnesium Jet profile monitor the wire is replaced by a thin stream of Magnesium vapour that is transversely swept across the beam and the mechanical position of the nozzle is recorded. The Magnesium is ionized by the beam and the electrons are collected by a transverse electric field into a photo-multiplier which will produce a signal proportional to the number of Magnesium atoms hit by the beam.

A further variant of the same theme is the *residual gas monitor* where the Magnesium Jet is replaced by the residual gas that is normally present in the beam pipe to some extent. The gas is ionized and the electrons are collected by transverse and magnetic fields onto a multichannel plate that is position sensitive and yields information about the beam size.

Electron and positrons at high energies emit synchrotron light in bending dipole magnets which can be imaged onto a camera with e.g. a pinhole and thereby a direct image of the transverse elecron beam distribution is obtained.

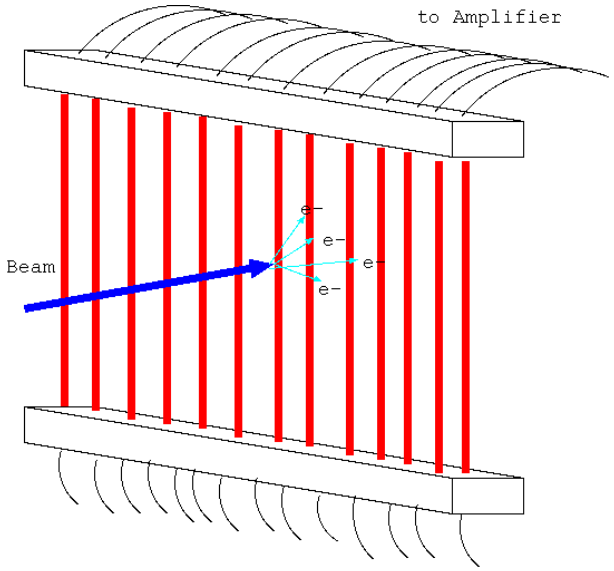

Figure 55: Schematics of a secondary emission monitor grid.

The second moment of the longitudinal distribution or the arrival times of bunches longer than several nano-seconds can be obtained by directly observing the sum signal of a BPM on a fast oscilloscope. This works well with long bunches with a length of several meters, but once the bunches get shorter other means are required.

Very short electron bunches down to the pico-second range (0.3 mm) can be measured with a streak camera which is a mechanical fast rotating mirror that smears the short light-pulse spatially onto a light detector. The width of the 'streak' is related to the pulse length of the original light pulse. More modern designs focus the light pulse onto a photocathode and the emitted and accelerated electrons are 'streaked' transversely by a transversely varying electric field onto a screen.

The same trick can also be applied to the primary electron beam if it is transversely deflected by a transversely deflecting mode in a cavity. This works like a backwards version of a cavity BPM, shown in Fig. 53. Here the antenna excites a deflecting mode and depending on the arrival time longitudinal slices of the beam receive different transverse kicks that can be streaked over a luminescent screen. Such a transversely deflecting cavity called LOLA is installed at the FLASH facility at DESY. Alternatively, a longitudinal accelerating cavity can be used to give different longitudinal slices different energies that will show up as a streak on a screen at a downstream location with dispersion.

Measuring the length of bunches in the femto-second range requires techniques borrowed from ultra-short laser science. See Ref. [31] for a recent review. In one case the electric field of an electron bunch that is passed close to an electro-

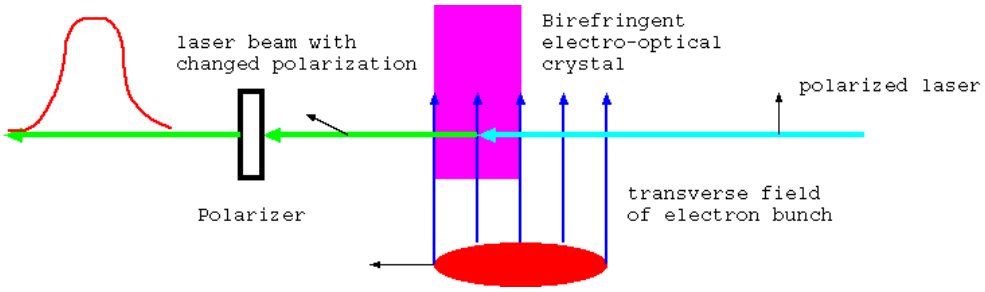

Figure 56: Schematics of an electro-optical sampling experiment.

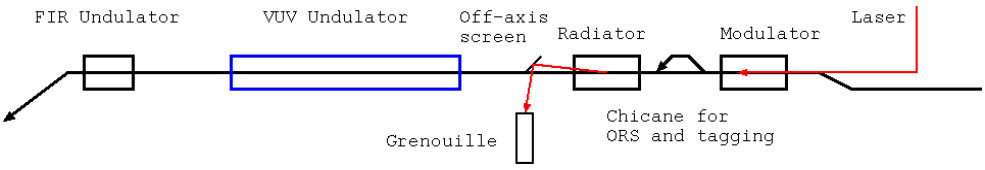

Figure 57: Schematics of the optical replica synthesizer.

optical crystal changes the polarization properties of the crystal which in turn affects a laser pulse that is passed through the crystal simultaneously, as is shown in Fig. 56. Normally the relative polarization of crystal and laser is chosen such that all laser power is absorbed in the crystal, and only if the polarization of the crystal is changed due to the electric field of the passing electron bunch a laser pulse that resembles the longitudinal profile of the electron bunch appears on the other side of the crystal. Several variations of this theme are in operation at DESY, SLAC and other laboratories.

Another novel idea that exploits laser diagnostic capabilities to completely diagnose femto-second light pulses is the Optical Replica Synthesizer (ORS), which creates an optical copy of the longitudinal electron distribution and utilizes frequency resolved gating (FROG) techniques [32] to diagnose the profile of the light pulse. In the ORS a seed laser modulates the energy of the electron bunch in an undulator where the transverse electric field of the laser couples to the transverse velocity component of the electron motion. The energy modulation is turned into a longitudinal density modultion with a period length of the laser wavelength in a small chicane. The micro-bunched beam is then made to radiate in a second undulator tuned to the same wavelength. The emitted light pulse has the wavelength of the seed laser and the envelope of the longitudinal electon distribution. A prototype of the ORS will be installed in FLASH at DESY in 2007 with active participation of Uppsala University.

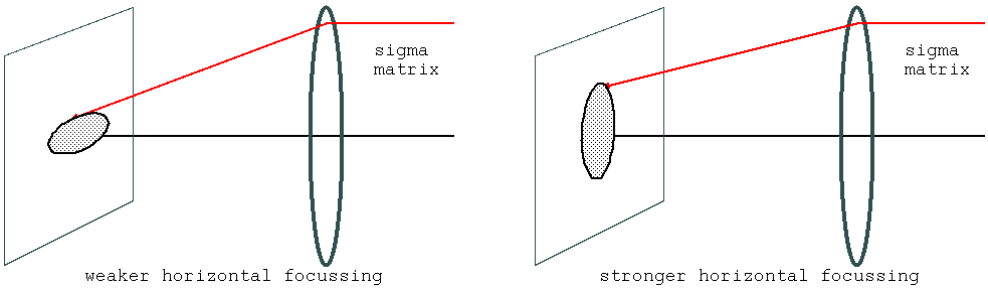

Figure 58: Quadrupole scan.

7.4 Emittance and Beta Function

Turning back to the transverse beam size measurements we now want to discuss methods of how to experimentally determine the emittance and beta function from size measurements. We will restrict ourselves to the one-dimensional case where the beam sigma-matix is only a 2×2 matrix which has three independent parameters, namely

$$\sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \langle x^2 \rangle & \langle xx' \rangle \\ \langle xx' \rangle & \langle x'^2 \rangle \end{pmatrix} = \varepsilon \begin{pmatrix} \beta & -\alpha \\ -\alpha & \gamma \end{pmatrix} \tag{162}$$

with $\gamma = (1 + \alpha^2)/\beta$. All we *can* directly measure is the beam size but we still want to beable to determine the other parameters such as emittance or $\langle xx' \rangle$ as well and we obviously need at least three measurements to determine the three independent parameters, either the three second moments such as $\langle x'^2 \rangle$ or $\varepsilon, \beta, \alpha$. We start by determining the moments first and later calculate the emittance and beta from them.

Consider Fig. 58 which shows the spot on a luminescent screen for two different settings of an upstream quadrupole. In the first case on the left, the quadrupole is focussing weakly and the beam is horizontally wide, but vertically small. If we increase the quadrupole excitation, the horizontal beam size will shrink, but the vertical beam size will increase, because quadrupoles focus in one plane but defocus in the other. We will now see how the horizontal beam size will change as a function of the quadrupole excitation for an input beam specified by its moments. For this we need the transfer matrix from the entrance of the quadrupole to the screen

$$R = \begin{pmatrix} 1 & l \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix} = \begin{pmatrix} 1 - l/f & l \\ -1/f & 1 \end{pmatrix} \tag{163}$$

where l is the distance from the quadrupole to the screen and f is the focal length of the quadrupole. We now transport the input beam σ from Eq. 162 with the transfer-matrix R from Eq. 163 to the screen by $\bar{\sigma} = R\sigma R^t$ and obtain for the 11

component of $\bar{\sigma}$ which is the beam size squared on the screen

$$\begin{aligned}\bar{\sigma}_{11} &= R_{11}^2\sigma_{11} + 2R_{11}R_{12}\sigma_{12} + R_{12}^2\sigma_{22} \\ &= (1-l/f)^2\sigma_{11} + 2l(1-l/f)\sigma_{12} + l^2\sigma_{22} \\ &= \left(\frac{l}{f}\right)^2\sigma_{11} - \left(\frac{l}{f}\right)(2\sigma_{11}+2l\sigma_{12}) + (\sigma_{11}+2l\sigma_{12}+l^2\sigma_{22})\end{aligned} \tag{164}$$

which shows that the squared beam size on the screen plotted as a function of the relative quadrupole excitation l/f is a parabola and the coefficients of the parabola are related to the beam parameters of the incoming beam. By setting the quadrupole to at least three different values an recording the beam sizes on the screen we can therefore determine the three moments σ_{ij}.

From the three moments we can in turn determine the emittance from

$$\varepsilon = \sqrt{\det\sigma} = \sqrt{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \tag{165}$$

and the beta and alpha functions from

$$\beta = \frac{\sigma_{11}}{\varepsilon} \quad \text{and} \quad \beta = -\frac{\sigma_{12}}{\varepsilon} \tag{166}$$

An alternative to measuring the emittance from a quadrupole scan is to measure the beam size at at least three consecutive location. The transfer matrices from the start location to the screens or wire scanners labelled n are denoted by a superscript n such as R^n. The measured beam size squared at screen n we will denote by σ_n^2. The dependence of the measurements on the incoming beam at an upstream location, labelled 0 is then determined by the following set of equations

$$\begin{aligned}\sigma_1^2 &= (R^1)_{11}^2\sigma_{11} + 2R_{11}^1R_{12}^1\sigma_{12} + (R^1)_{12}^2\sigma_{22} \\ \sigma_2^2 &= (R^2)_{11}^2\sigma_{11} + 2R_{11}^2R_{12}^2\sigma_{12} + (R^2)_{12}^2\sigma_{22} \\ \sigma_3^2 &= (R^3)_{11}^2\sigma_{11} + 2R_{11}^3R_{12}^3\sigma_{12} + (R^3)_{12}^2\sigma_{22}\end{aligned} \tag{167}$$

which can be written in matrix form as

$$\begin{pmatrix}\sigma_1^2\\ \sigma_2^2\\ \sigma_3^2\end{pmatrix} = \begin{pmatrix}(R^1)_{11}^2 & 2R_{11}^1R_{12}^1 & (R^1)_{12}^2\\ (R^2)_{11}^2 & 2R_{11}^2R_{12}^2 & (R^2)_{12}^2\\ (R^3)_{11}^2 & 2R_{11}^3R_{12}^3 & (R^3)_{12}^2\end{pmatrix}\begin{pmatrix}\sigma_{11}\\ \sigma_{12}\\ \sigma_{22}\end{pmatrix} \tag{168}$$

and we can solve for the unknown σ_{ij} by simple matrix inversion. If there are more measurements in for example one or more addiational screens, we can just one ore more equations to Eq. 167 which leads to more rows in the corresponding matrix equation 168 which then is over-determined and we can solve it in the least-square sense using the usual $(A^tA)^{-1}$ tricks from *Numerical Recipes*..

These measurements can be extended to determine the coupled 4×4 beam matrix if there are skew quadrupoles available to change the coupling.

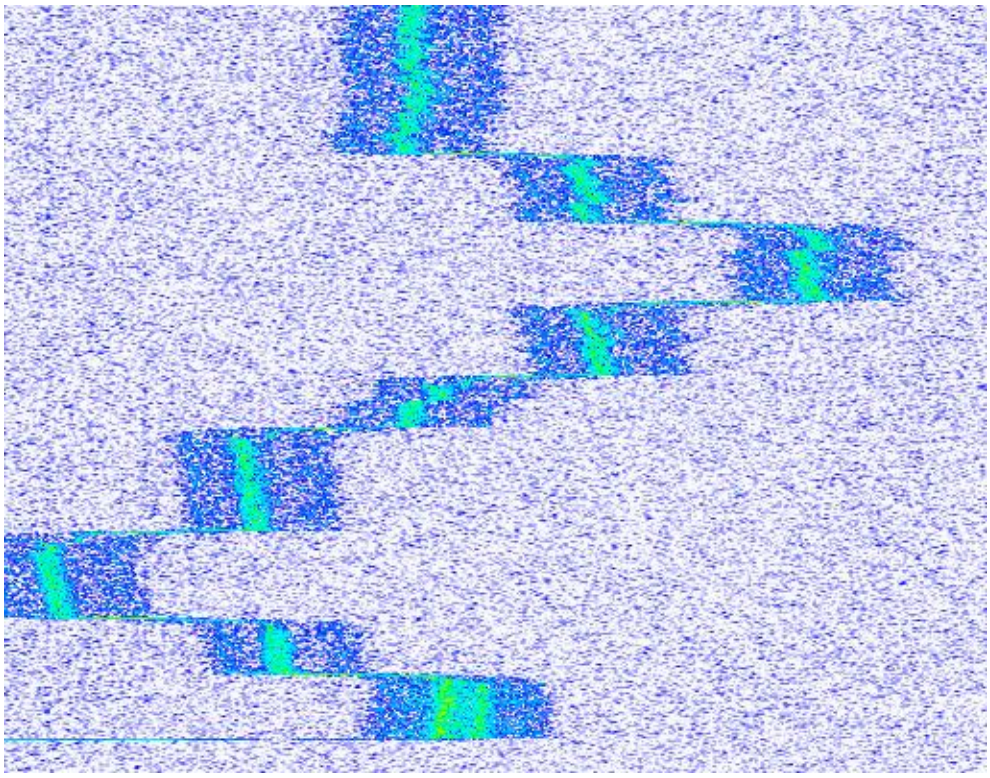

Figure 59: Schottky spectrum of 248.25 MeV/u deuterons in CELSIUS at the second harmonic (4.500480 MHz, 1020 Hz full bandwidth) when stepping the HVPS in steps of 20 V around 136 kV. Time runs from top to bottom.

7.5 Schottky Signals

In circular proton or ion accelerators the accelerating radiofrequency can be turned off because the particles do not loose energy due to synchrotron radiation. With absent RF the longitudinal focussing is absent and the beam will smear out around the entire ring without a bunch structure. This is commonly called a coasting beam. Since the momentum spread of the beam σ_p is non-zero, the particles will have a spread of revolution frequencies, proportional to $\eta\sigma_p$. In practice every particle has its particular energy and revolution frequency. This means that each particle will induce a signal in a sensitive detector, a 'ping' depending on it revolution frequency. Since there are very many particles in the ring, we can expect to observe a distribution of frequencies, representing the momentum distribution of the beam. This is called the Schottky signal. In Fig. 59 we show a spectrogram recorded on a fast spectrum analyzer in CELSIUS. The horizontal axis represents the revolution frequency and vertically time increases from top to bottom. The colors encode the signal intensity. In the experiment documented in Fig. 59 we have deliberately changed the revolution frequency by

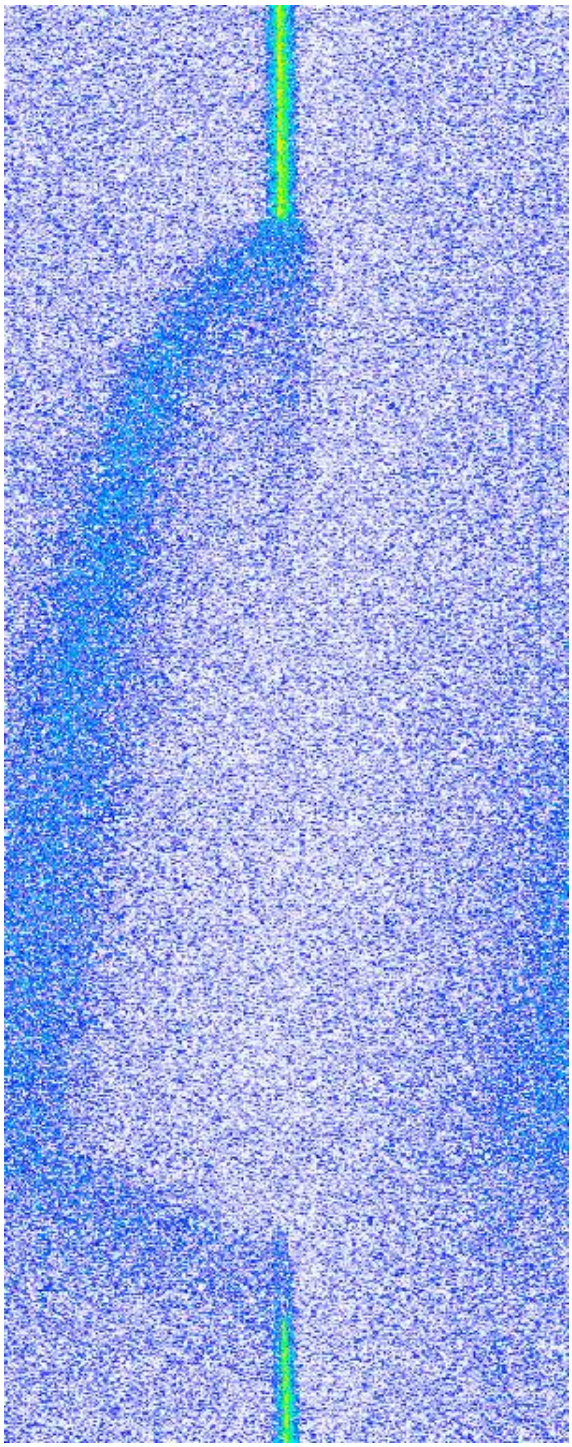

Figure 60: Schottky signal of deuterons in CELSIUS at the 2nd harmonic at 4.50048 MHz with full span of 4.07 kHz. Time increases from top to bottom.

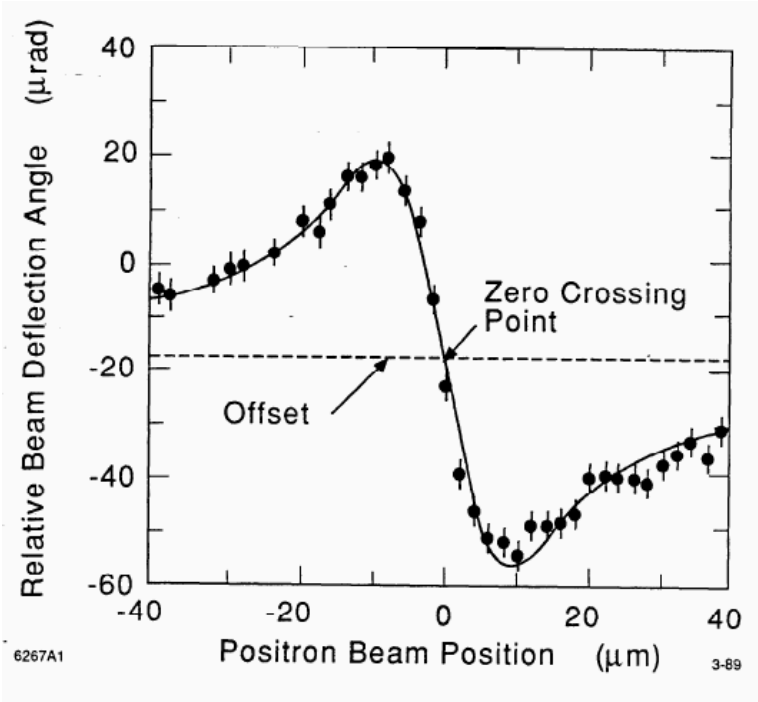

Figure 61: Beam beam deflection curve

changing the high voltage of the electron cooler in CELSIUS which changes the speed of the, in this case, deuterons, which shows up as a changed revolution frequency. The width of the distribution corresponds to the momentum spread.

In another experiment shown in Fig. 60 we had a target used in nuclear physics experiments present and the beam is initially narrow due to the presence of the electron cooler, which is, however, turned off after a short while, which causes the beam to widen and also the average frequency changes to lower values which is due to the energy lost by the beam in the target. Normally it is replenished by the electron cooler, but that was turned off in this case. Since the target thickness is proportional to the rate of energy loss which shows up as a frequency change of the Schottky signal. We routinely used this method to determine the target thickness in CELSIUS.

7.6 Beam-Beam Diagnostics

at SLAC and for future linear colliders

7.7 Timing

Technically, timing diagnostic is also a is also a measurement of the first moment, but ...

8 Imperfections and their Correction

In this chapter we will discuss imperfections of the magnets used in an accelerator, how to diagnose them and finally how to correct them. In this context it is useful to recall how magnets are chracterized in terms of their multipolarity. The field in magnet has to fulfill Maxwell's equations and it can be shown that the magnetic field therefore can be derived from a magnetic potential.

The effect of a thin multipole of order n, where $n = 1$ denotes quadrupole, $n = 2$ denotes the sextupole and so forth, on the beam is that of a kick that depends on the initial position x and y of the particle when entering the magnet and is given by

$$\Delta x' + i\Delta y' = \frac{k_n L}{n!}(x + iy)^n \tag{169}$$

where $\Delta x'$ and $\Delta y'$ denote the horizontal and vertical kick that the particle receives and $k_n L$ is the integrated strength of the magnet, normalized to momentum, or $B\rho$, of the beam

$$k_n L = \frac{d^n B/dx^n}{B\rho} L \ . \tag{170}$$

where L is the length of the magnet. The kick that a particle receives in the corresponding skew-magnets can be calculated by multiplying the right hand side by the imaginary unit i.

The most obvious magnetic imperfection is an *excitation error* which can come from a power supply error or a mis-calibration of the magnet itself. In this case the type of the error is equal to the magnet type, the field pattern is correct, but the the magnitude of the field is wrong. This can be represented by a somewhat erroneous $k_n L$ of the magnet.

Another type of magnetic error is *misalignment* which can be visualized by a magnet not standing of the design orbit, either longitudinally, or transversely. The effect of transverse misalignment by for example a horizontal distance Δ_x can be most easily undersood by a change of coordinates and the kick that a particle receives is no longer given by Eq. 169, but by

$$\Delta x' + i\Delta y' = \frac{k_n L}{n!}(x + \Delta x + iy)^n \ . \tag{171}$$

If we restrict ourselves to the horizontal plane, consider a sextupole with $n = 2$ and expand the term in the bracket we see that the particle receives a horizontal kick $\Delta x'$ given by

$$\Delta x' = \frac{1}{2} k_2 L(x^2 + 2\Delta_x x + \Delta_x^2) \ . \tag{172}$$

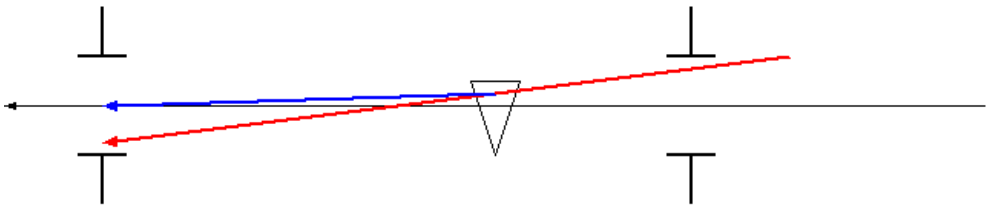

Figure 62: One-to-one tseering.

Note that the particle now receives a kick that is proportional x^2 as before but also a kick proportional to x and a kick independent of x the latter which represents a dipole kick of magnitude $k_2 L\Delta_x^2/2$ that affects a particle, even though it is on the design orbit. No surprise here, the particle is on the design orbit, but the sextupole is not. The term linear in x affects the particles in the same way as an additional quadrupole with strength (or inverse focal length) $k_2 L\Delta_x$. Observe that a misaligned sextupole gives rise to a quadrupole and dipole errors. This is a general behavior of misaligned multipoles that give rise to a multitude of lower multipoles. This effect is called *feed-down* of the multipoles.

A further source of magnetic errors are the end-fields of magnets, which always be described by their multipolar contents in terms of a sequence of $k_n L$. The magnitude of these coefficients can usually only be determined experimentally by magnetic bench measurements or by numerical codes that model the magnetic field from the the geometry of the magnet and the excitation of the coils.

In the following sections we now turn to some of the more common imperfections that appear in the daily operation of accelerators and see how they are diagnosed and corrected.

8.1 Orbit

The most common correction in accelerators is arguably orbit correction. The orbit of the stored beam is usually observed with beam position monitors (BPM) and often the the orbit that is displayed on a computer screen is nowhere near the design orbit which should read zero on all BPM, if these were properly aligned initially. The source of a non-zero orbit are dipole errors in the accelerator which can come from all the error sources we discussed in the previous section such as misaligned magnets. Normally there are also dipole steering magnets installed in an accelerator that can be used to deflect the beam in such a way as to zero the position on the BPM. If for example there is on steering magnet upstream of every BPM with a large R_{12} transfer matrix element between the corrector and the BPM a position error Δ_x can be corrected by exciting the corrector as to give a deflection angle θ with $R_{12}\theta$ such that $\Delta_x + R_{12}\theta = 0$. If there is one corrector for every BPM this method is called one-to-one steering of a beam line. Observe

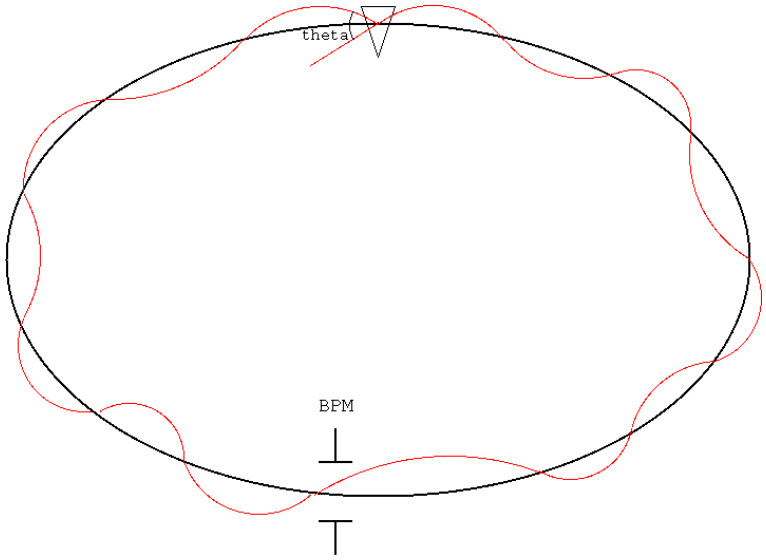

Figure 63: Closed orbit.

that in a straight beam line a corrector will only affect the orbit downstream of its position.

Note that the situation in a circular accelerator is slightly different because exciting a single steering magnet will cause a perturbation of the orbit everywhere in the accelerator, because the orbit perturbation must be periodic with the circumference. The equilibrium orbit in a ring is therefore often referred to as *closed orbit.* The concept of the closed orbit is shown in Fig. 63 where the dipole corrector at the top of the ring gives the beam a kick of magnitude θ which causes the orbit to oscillate around the ring. The orbit immediately after the corrector we denote by $(x, x')_{co}$ and follow the orbit vector around the ring by first propagating around the ring with a transfermatrix and then applying the kick $(0, \theta)$. To simplify the algebra we assume that we have $\alpha = 0$ and the transfermatrix that describes the propagarton of the beam around the ring is then only characterized by the beta function β and the tune Q. After one turn the requirement for a closed solution yields the following equation

$$\begin{pmatrix} x \\ x' \end{pmatrix}_{co} = \begin{pmatrix} \cos(2\pi Q) & \beta \sin(2\pi Q) \\ -\sin(2\pi Q)/\beta & \cos(2\pi Q) \end{pmatrix} \begin{pmatrix} x \\ x' \end{pmatrix}_{co} + \begin{pmatrix} 0 \\ \theta \end{pmatrix} \tag{173}$$

Solving for the closed orbit results in

$$x_{co} = \frac{\beta}{2} \cot(\pi Q) \qquad , \qquad x'_{co} = \frac{\theta}{2} \tag{174}$$

which implies that the kick-angle θ is affects the closed orbit in the way shown in Fig. 63. Furthermore note that the position at the corrector x_{co} changes as well and the change depends on the beta function β and the tune Q.

An important point to note is the dependence on the tune through the cotangent which has a singularity at multiples of π which implies that the orbit is exceedingly sensitive to dipole perturbations if the tune Q is an integer.

In order to find out the position change of the closed orbit change x_i at the BPM i that the steerer caused we need to propagate the closed orbit change at the corrector labelled j to the BPM using the transfer matrix between the corrector and the BPM. Such a transfer matrix between two positions with beta functions β_i and β_j and phase advance ψ_{ij} is given by

$$x_i = \frac{\sqrt{\beta_i \beta_j}}{2\sin(\pi Q)} \cos(\psi_{ij} - \pi Q)\theta_j \tag{175}$$

Note that the factor in front of the kick angle θ_j takes the role of he R_{12} transfer matrix element ih a beam line in the sense that it describes the change in orbitb poition at a point i due to a change in angle at position j, but taking the constraint that the orbit must be closed into account. The factor is commonly called *response coefficient* decause it describes the response of the BPM reading to a change of the steering magnet. If we tolerate a little sloppy notation, we can write this response coefficient by $C^{ij} = \partial BPM_i/\partial COR_j$. In the case of a straight beam line the response coefficients are equal to the earlier encountered transfer matrix elements such as R_{12} for the horizontal plane or R_{34} in the vertical plane.

In the previous paragraph we expressed the response coefficients between positions in terms of beta functions and phase advances. Which is the commonly used way to operate if the beamline is un-coupled, which means that changing horizontal corrector will only affect the horizontal orbit and not the vertical. In the more general case with arbitrary 4×4 possibly coupled transfer matrices we can still calculate the response coefficients. We will denote the 4×4 transfer matrix that propagates the beam from a point labelled j to itself by R^{jj} and call it the full-turn matrix at point j. The orbit perturbation that we have at point j we denote by a vector with 4 components $\vec{\theta}_j = (0, \theta_x, 0, \theta_y)^t$ where $\theta_{x/y}$ are the kick angles in the horizontal and vertical plane, respectively. The closed orbit perturbation $\vec{x}_j$ immediately after point j is then given by

$$\vec{x}_j = R^{jj}\vec{x}_j + \vec{\theta}_j \tag{176}$$

where we denote orbit vectors with four components by an arrow. Similarly to what we did in the previous paragraph we solve for $\vec{x}_j$

$$\vec{x}_j = (1 - R^{jj})^{-1}\vec{\theta}_j \tag{177}$$

where the 1 in the brackes denotes a a 4×4 unit matrix. Now we have the effect of a general perturbation at the source location labelled j. If we want to know it at another location labelled i we simply have to propagate the orbit vector $\vec{x}_j$ by the 4×4 transfer matrix R^{ij} from point j to point i and get

$$\vec{x}_i = R^{ij}\vec{x}_j = R^{ij}(1 - R^{jj})^{-1}\vec{\theta}_j = C^{ij}\vec{\theta}_j \tag{178}$$

where we introduced the 4×4 response matrix C^{ij} between location j and i defined by

$$C^{ij} = R^{ij}\vec{x}_j = R^{ij}(1 - R^{jj})^{-1} \ . \tag{179}$$

It is useful to norte that the response matrix element in a straight beam line were R^{ij} and the closed-orbit constraint is accounted for by the term $(1 - R^{jj})^{-1}$.

The response matrix C^{ij} is a 4×4 matrix and the response of a horizontal position observation to a horizontal angle kick is the matrix element C^{ij}_{12} in the horiontal plane. In a coupled beam line the matrix element C^{ij}_{32} denotes the variation of the vertical orbit (first lower index '3') to a horizontal kick (second lower index '2'). If the accelerator is uncoupled all matrices R and C are block-diagonal and the 32 element is zero.

By now it should be obvious that the calculations of the response coefficients is largely an exercise in linear algebra once the transfer matrices R^{ij} are known and the latter can often easily obtained by numerical accelerator optics codes, such as MAD [8] where the`sectormap` command provides a file with all transfer matrices of beam line that can be used to subsequently calculate the response coefficients.

We now address the question how the knowledge of the response coefficients can be used to correct the orbit of a circular accelerator globally. We assume that we have a number N of BPM in the ring and M corrector magnets. To simplify the notation we assume that we only correct the orbit in the horizontal plane and that due to the reasons discussed in the first secton of this chapter the BPMs read a non-zero orbit, that we want to correct to zero or some other 'golden orbit' that was determined by some other means, such as favourable background situation. The readings of the N BPM we denote by x_i with $i = 1, \ldots, N$ and the desired orbit by $\hat{x}_i$. The kick angles of the steering or correction magnets, here assumed to be horizontal, are denoted by θ_j with $\theta = 1, \ldots, M$. If we knew (note the conditional!) all the M corrector values we could calculate the the change of position at the $i-$th BPM Δx_i by summing over the correctors, weighted by the response coefficients C^{ij}_{12} between corrector j and BPM i

$$\Delta x_i = \sum_{j=1}^{M} C^{ij}_{12}\theta_j \tag{180}$$

where the sum extends over all corrector magnets. The desired change in orbit is of course he difference between the measured orbit x_i and the desired orbit $\hat{x}_i$, or $\Delta x_i = x_i - \hat{x}_i$. We just make the correctors add the negative of the measured orbit to the beam, thus zero'ing the difference between the desired and initially measured orbit. We can now explitely write Eq. 180 in matrix form

$$\begin{pmatrix} x_1 - \hat{x}_1 \\ x_2 - \hat{x}_2 \\ \vdots \\ x_N - \hat{x}_N \end{pmatrix} = \begin{pmatrix} C^{11}_{12} & C^{12}_{12} & \ldots & C^{1M}_{12} \\ C^{21}_{12} & C^{22}_{12} & \ldots & C^{2M}_{12} \\ \vdots & \vdots & \ddots & \vdots \\ C^{N1}_{12} & C^{N2}_{12} & \ldots & C^{NM}_{12} \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_M \end{pmatrix} \ . \tag{181}$$

Since the BPM usually have a finite measurement resolution σ_i which may be different from BPM to BPM we might want to trust the better BPM with smaller σ_i (say $10\,\mu$m) more than those with worse resolution of $100\,\mu$m. This we can incorporate into the matrix inversion by dividing the corresponding rows by the respective BPM resolution, thereby giving the proper weight to the corresponding equations. Explicitely rewriting Eq. 181 in this case reads

$$\begin{pmatrix} (x_1-\hat{x}_1)/\sigma_1 \\ (x_2-\hat{x}_2)/\sigma_2 \\ \vdots \\ (x_N-\hat{x}_N)/\sigma_N \end{pmatrix} = \begin{pmatrix} C_{12}^{11}/\sigma_1 & C_{12}^{12}/\sigma_1 & \dots & C_{12}^{1M}/\sigma_1 \\ C_{12}^{21}/\sigma_2 & C_{12}^{22}/\sigma_2 & \dots & C_{12}^{2M}/\sigma_2 \\ \vdots & \vdots & \ddots & \vdots \\ C_{12}^{N1}/\sigma_N & C_{12}^{N2}/\sigma_N & \dots & C_{12}^{NM}/\sigma_N \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_M \end{pmatrix} . \quad (182)$$

or somewhat loosely $\vec{x}-\vec{\hat{x}} = A\vec{\theta}$ where the quantities with an arrow denote the respective column vectors and A denotes the matrix with the response coefficients in Eq. 181. The way we formulated the problem was backwards: we assumed knowledge of the correctors and deduced the change in the BPM readings. In practice we normally want to know the converse: we want to determine the corrector excitations θ_j that minimizes the BPM readings $x_i - \hat{x}_i$. To this end we observe that we just have to invert the matrix A. If the number of BPM N and correctors M is equal we indeed can determine the required corrector excitations θ_j from

$$\vec{\theta} = -A^{-1}(\vec{x}-\vec{\hat{x}}) \quad (183)$$

where the minus sign makes sure that adding this zeroes the orbit. Here we made the simplifying assumption that the matrix is square ($N = M$) and non-degenerate, but this is not always the case.

First we treat the case where there are more BPM to measure the orbit than steering magnets to correct it which is often the case, because saving steerers saves magnets and power supplies and therefore money. We thus have $N > M$ and there are more rows than columns. Imagine an accelerator with only one steering magnet. In that case the matrix a would be a $N \times 1$ matrix or a column vector. Having more constraints, which is what the BPM readings are, than unknowns, which is what the corrector excitations are, we have an over-determined linear system, which we can solve in the least-square sense, which means that we try to determine $\vec{\theta}$ such that the rms orbit $\sum_i (x-\hat{x})^2/N$ is minimized. In ref. [26] we find that solving the linear least squares problem $\Delta\vec{x} = A\vec{\theta}$ is done by

$$\vec{\theta} = (A^T A)^{-1} A^T \Delta\vec{x} \; . \quad (184)$$

where A^T is the transpose of matrix A. This so-called pseudo-inverse can be calculated by normal matrix inversion provided that the matrix $A^T A$ is non-degenerate. In the case of non-degeneracy the inverse can still be calculated using Singular Value Decomposition (SVD) which is also used if there are less BPM than correctors. This case we treat in the next paragraph.

If we have less BPM than correctors the matrix A in Eq. 181 has more columns than rows. Imagine a single BPM and many correctors, in that case the matrix a is one long row vector, which cannot be 'inverted' in the normal sense, because it is highly degenerate. In such cases SVD comes to the rescue, because it constructs a set of basis vectors for the subspace of the matrix A that is degenerate and also that for the non-degenerate subspace. The trick is now to project out the non-degenerate part of the matrix and invert only that part and simultaneously set the projection of the solution vector $\vec{\theta}$ onto the degenerate subspace to zero. The effect of this somewhat bizarre operation is that the rms of the orbit is minimized using the corrector excitation $\vec{\theta}$ with the minimum rms, i.e. $\sum_j \theta_j^2$ is minimum while simultaneously minimizing the orbit on the BPMs. The interested reader is referred to Ref. [26] for the very readable discussion of SVD. I will make this paragraph more explcit in a later version.

Another often used algorithm to 'invert' the matrix A is the MICADO algorithm, which is especially suitable for rings with many (up to several hundreds, like LEP) BPM and correctors. MICADO operates according to the following principle. First pick the corrector that improves the orbit the most and correct. Then repeat this procedure with the remaining orbit and correctors. This appears like a lot of searching, but the algorithm can be very efficiently coded using Householder-transformations of the matrix A [27].

In synchrotron radiation sources the beam also has to be stabilized on the monochromator used to select the photon energy of the emitted synchrotron radiation. Often the position of this signal on the can be monitored and be used as a virtual BPM. It can be included in beam optics codes that are used for modelling and correcting accelerators by using a trick due to Martin Lee from SLAC and introduce a module with the transfer-matrices

$$(L)(BPM)(-L) \tag{185}$$

at the source point of the synchrotron radiation in the bending or undulator magnet. Here (L) denotes the transfer matrix for a drift space and (BPM) resulting in a measurement value x_i of the position and $(-L)$ is the inverse transfer-matrix transporting the 'virtual' beam back onto the orbit.

There are many articles covering orbit correction and ref. [28] is a good starting point for further studies.

8.2 Bumps and Knobs and Rock & Roll

Occasionally particular displacement or angle of the orbit at a special point is required without affecting other parameters. An examples is changing the position of the beam at the interaction point without changing the angle. This is used to bring colliding beams that initially miss each other into collision as isillustrated in Fig. 64. Another example is the steering of the synchrotron radiation light onto a monochromator as is shown in Fig. 65.

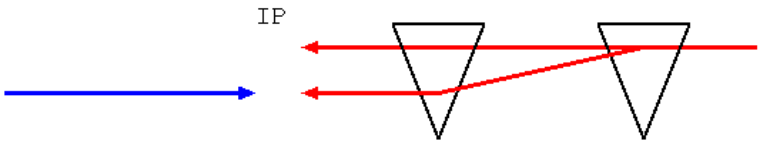

Figure 64: Two steering dipole correctors are used to change the position of the beam at the IP without affecting the angle.

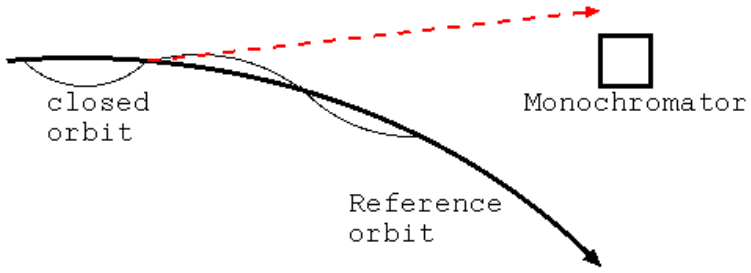

Figure 65: The angle of the source point of the emission of synchrotron radiation must be adjusted such that the radiation hits the monochromator at the end of an often long synchrotron radiation beam line.

Often it is desirbale to only affect the orbit locally, meaning that a certain position of the orbit is steered to, but the total effect of the perturbation should be kept as local as possible. This brings us to the concept of a local bump which is illustrated in Fig. 66. The first two steerers, labelled 1 and 2, at the right will adjust the orbit to achieve displacement and angle x_0 and x_0', at a center point. The following two correctors 3 and 4 will undo the changes of the orbit. This type of local orbit bump is often encountered close to an injection septum or a collimator. Even the entire orbit of a ring or a beam line can be steered using a sequence of such bumps.

To calculate the excitation of the steerers for such a bump is straight-forward. We denote the kick angles of the steerers by θ_j and the transfer matrix between corrector j and the center point R^{0j} and to the final point R^{fj}. Then we write down how every corrector affects the position and angle at the respective points

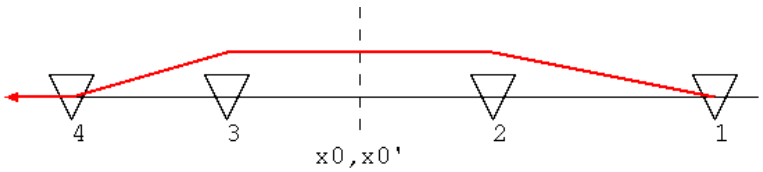

Figure 66: Illustration of a bump with four dipoles.

where we thus formulate as constraints in Eq. 187

$$\begin{pmatrix} x_0 \\ x_0' \\ x_f = 0 \\ x_f' = 0 \end{pmatrix} = \begin{pmatrix} R_{12}^{01} & R_{12}^{02} & 0 & 0 \\ R_{22}^{01} & R_{22}^{02} & 0 & 0 \\ R_{12}^{f1} & R_{12}^{f2} & R_{12}^{f3} & R_{12}^{f4} \\ R_{22}^{f1} & R_{22}^{f2} & R_{22}^{f3} & R_{22}^{f4} \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_3 \end{pmatrix} . \tag{186}$$

The interpretation of this equation is very intuitive. Consider the first equation, that defines the upper row of the matrix appearing in Eq. 187. We want to have x_0 at the center point and only the upstream correctors 1 and 2 can affect it. The next line does the same thing for the angle x_0' at the center point. The third and fourth row come from the requirement that the position and angle at the end of the bump should be zero, colloquially called closing the bump. Here all four correctors are involved.

This set of equations triagonal and the upper two equations can be solved by themselves by inverting a 2×2 matrix. The thus found angles θ_1 and θ_2 can then be inserted in the two lower equations leading to another linear system for the unknowns θ_3 and θ_4 that can be solved by inverting another 2×2 matrix.

To illustrate the procedure we assume that involved transfer matrices are driftspaces and that all elements are l=1 m apart from another. Eq. 187 for this simplified example then reads

$$\begin{pmatrix} x_0 \\ x_0' \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2l & l & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 4l & 3l & l & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_3 \end{pmatrix} . \tag{187}$$

Solving the first two equations we find

$$\theta_1 = x_0/l - x_0' \quad \text{and} \quad \theta_2 = -x_0/l + 2x_0' \tag{188}$$

and inserting in the last two equations yields

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} x_0 + 2lx_0' + l\theta_3 \\ x_0' + \theta_3 + \theta_4 \end{pmatrix} \tag{189}$$

which is trivially solved for θ_3 and θ_4. We can summarize all results in the following matrix form

$$\begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_3 \end{pmatrix} = \begin{pmatrix} 1/l & -1 \\ -1/l & 2 \\ -1/l & -2 \\ 1/l & 1 \end{pmatrix} \begin{pmatrix} x_0 \\ x_0' \end{pmatrix} \tag{190}$$

Note that this is a prescription of how to change the steerer excitations θ_j in order to achieve the desired bump amplitude x_0 and x_0'. This is an example of a so-called multi-knob, which is a prescription for moving a group of power summplies

```
! ================================================
! SCNFF IP:  ELECTRON X POSITION KNOB
! USING A3X AND A6X NORTH
! UNITS: BDES[KG-M]/MICROMETER
!
SET/LABEL=EXIP36
SET/SENS=64
SET/NOWARN
DEF/DEV=(XCOR,FF11,5170,BDES)/COEF= -29.563E-6   ! A6X NORTH
DEF/DEV=(XCOR,FB69, 520,BDES)/COEF= 309.008E-6   ! A3X NORTH
```

Figure 67: A multiknob used in the SLC final focus to change the horizontal position of the electron beam at the IP.

or other devices in a coordinated fashion in order to achieve some objective. If we wanted for example a knob that moves the position x_0 without affecting the angle x_0', we just take the first column of the matrix in Eq. 190.

These type multi-knobs are often found in accelerators and we show an example that was used in the SLC final focus in Fig. 67.

8.3 Tune

The tune of a circular accelerator is one of the most important parameters for the stability of operating the ring and therefore needs to be adjusted to values found to be advantageous. Apart from the orbit the tune is probably the most adjusted parameter. To correct it, we must, however, first measure it. This is usually done by feeding the analog signal from the difference channel of a beam position monitor to a spectrum analyzer. Alternatively, if available, one can use a fast turn-by-turn position measurement system that records the position of the beam for every turn and perform a Fourier analysis by e.g. FFT, of the thus obtained time series. Especially in rings with small circumference and consequently high revolution frequency turn-by-turn systems are difficult an/or expensive to build. In some accelerators systems based on PLLs are in use. The horizontal tune can be obtained from the horizontal positions of the BPMs and the vertical tune from the vertical positions.

We need to digress on the measurement procedure for a moment and consider a single BPM that is used to measure the tune by recording the turn-by-turn position signals and calculating the FFT. Since the BPM is at a fixed position in the ring, it cannot obeserve the integer number of full oscillations between successive passes of the beam and consequently only the fractional part of the tune is measured. This is also obvious from the FFT which results in a spectrum between 0 and 0.5 if only the modulus is displayed. Consequently only the fractional part

of Q, denoted by $[Q]$ or that of $1-Q$ is determined in this way. Normally Q and $1-Q$ can be distinguished by changing a horizontally focussing quadrupole and observing the change of the tune, which should increase for increasing quadrupole excitation.

Typical sources of tune errors are the fringe fields of magnets that were not properly taken into account in the modelling of the accelerator or power supply calibration errors. Even swapped cables resulting in quadrupoles with wrong polarity are not unheard of.

We now assume that the have a system that returns the measured fractional tunes Q_x and Q_y and two quadrupoles with a power supply that allows independent adjustment of their excitation. For simplicity we assume that they are thin quadrupoles with focal strength f_1 and f_2. The effect of quadrupole 1 on the tunes is given by

$$\Delta Q_x = \frac{\beta_{1x}}{4\pi f_1} \quad \text{and} \quad \Delta Q_y = -\frac{\beta_{1y}}{4\pi f_1} \tag{191}$$

where $beta_{1x}$ and β_{1y} are the horizontal and vertical beta function at the location of quadrupole 1 and corresponding equations for quadrupole 2. Note the minus sign in the second equation, because the quadrupole is focussing in one plane and defocussing in the other. If the changes are small the tune changes from the two quadrupoles can be simply added with the result

$$\begin{aligned} \Delta Q_x &= \frac{\beta_{1x}}{4\pi f_1} + \frac{\beta_{2x}}{4\pi f_2} \\ \Delta Q_y &= -\frac{\beta_{1y}}{4\pi f_1} - \frac{\beta_{2y}}{4\pi f_2} \end{aligned} \tag{192}$$

which can be written in matrix form

$$\begin{pmatrix} \Delta Q_x \\ \Delta Q_y \end{pmatrix} = \frac{1}{4\pi} \begin{pmatrix} \beta_{1x} & \beta_{2x} \\ \beta_{1y} & \beta_{2y} \end{pmatrix} \begin{pmatrix} 1/f_1 \\ 1/f_2 \end{pmatrix} \tag{193}$$

and the desired correction strengths $1/f_i$ can be obtained by inverting the 2×2 matrix in the previous equation which yields the result

$$\begin{pmatrix} 1/f_1 \\ 1/f_2 \end{pmatrix} = \frac{4\pi}{\beta_{1x}\beta 2y - \beta_{2x}\beta_{1y}} \begin{pmatrix} \beta_{2y} & -\beta_{2x} \\ -\beta_{1y} & \beta_{1x} \end{pmatrix} \begin{pmatrix} \Delta Q_x \\ \Delta Q_y \end{pmatrix} . \tag{194}$$

Selecting the negative of the calculated $1/f_i$ will allow to correct the measured ΔQ, provided the changes in the tune are small. If the measured tune errors are large, the correction can be iterated a few times and normally converges, unless the beta-functions are completely changed in the process.

8.4 Coupling

In most parts of this note we discussed the un-coupled motion in an accelerator, where the horizontal and vertical betatron oscillations are independent of one

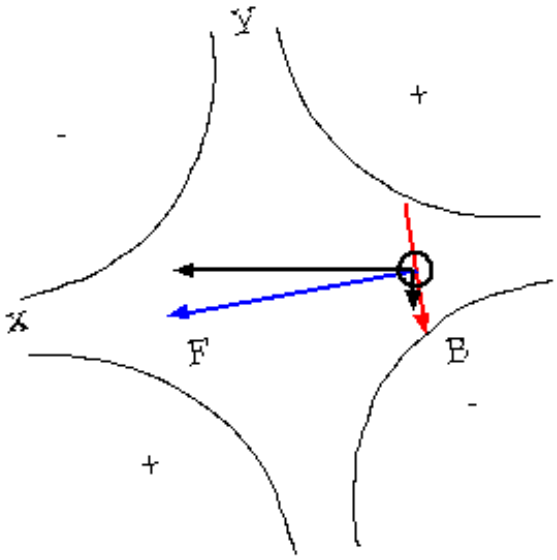

Figure 68: Rotated Quadrupole.

another. This independence is normally perturbed by quadrupoles with a roll error caused by a quadrupole tilted in the $x-y-$plane. An example is shown in Fig. 68 The force component of a horizontally offset particle has both a horizontal and a vertical component. A horizontal offset thus causes a vertical kick and couples the horizontal and vertical betatron motion. Other sources of coupling are solenoidal fields which are often part of the detectors in high energy accelerators or electron coolers in nuclear physics storage rings.

A further source of betatron coupling is a misaligned vertical orbit in strong sextupole magnets which often appear in synchrotron light sources for the compensation of the chromaticity. The kick that a particle receives in a sextupole is quadratic in the transverse coordinates and is given by Eq. 169 with $n = 2$. A vertically misaligned orbit is described by $y \rightarrow y + \Delta y$ and inserting in Eq. 169 we obtain

$$\begin{aligned} \Delta x' + i\Delta y' &= \frac{k_2 L}{2}(x + i(y + \Delta y))^2 \\ &= \frac{k_2 L}{2}\left[(x + iy)^2 + 2\Delta y(ix - y) - \Delta y^2\right] . \end{aligned} \tag{195}$$

The imaginary part of the right hand side describes the vertical kick and we observe a contribution proportional to $k_2 L \Delta y x$ of the vertical (imaginary) part. This part describes a vertical kick from a horizontally displaced particle, which is just what coupling is.

After having discussed the most common causes of coupling we need to consider why coupling is unwanted or even dangerous. One reason is that in an un-

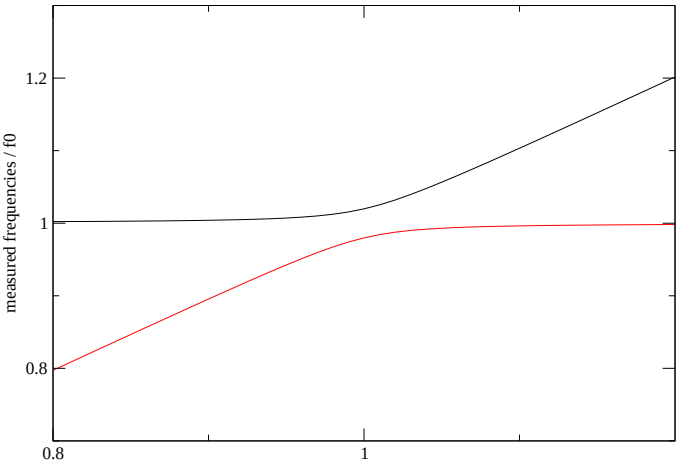

Figure 69: Closest tune.

coupled lattice only horizontal and vertical lines in the tune diagram can appear. In a coupled lattice resonances represented by diagonals in the tune diagram can be excited and careful counting shows that the number of lines thereby doubles and the regions between low-order resonances drastically shrinks. A second reason why coupling is often unwated is that correction of orbit and tune becomes more complicated. Consider orbit correction in an uncoupled accelerator where correction involves the inversion of the horizontal and the vertical resonse matrix independently. In a coupled accelerator only a single matrix with the combined response of all correctors and all BPM, both horizontal and vertical must be inverted. But this matrix is twice the size and inversion becomes numerically more unstable, especially if the response matrix is only known to finite precision. Moreover in a coupled lattice the notion of betafunction is not uniquely defined and the straightforward tune calculation method discussed in the previous section only works approximately.

In synchrotron light sources coupling will define the vertical emittance which should ideally be as small as possible to minimize the vertical source size of the emitted synchrotron radiation. In an earlier section we found that the horizontal emittance is produced by the emission of synchrotron radiation in dipole magnets if the dispersion function is non-zero in the dipoles. Thus only horizontal emittance is generated naturally in a perfect ring, but coupling will transfer some of this emittance, which can be visualized as transverse oscillation energy, into the vertical plane, thereby increasing the vertical emittance. Normally all magnets are carefully aligned initially and special attention is given to the vertical orbit

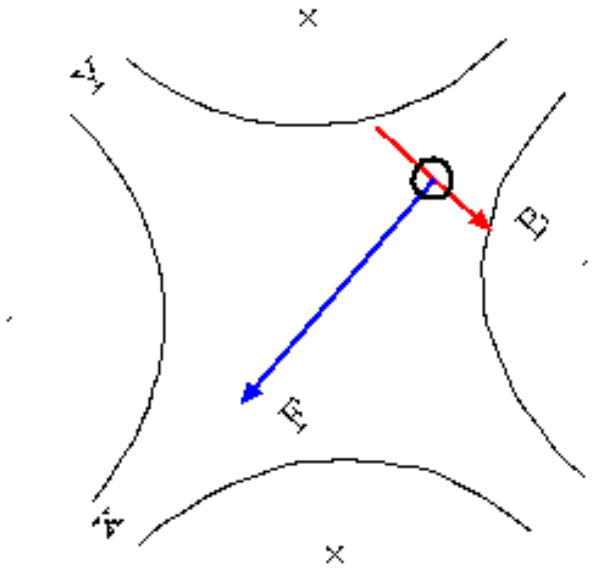

Figure 70: Skew quadrupole.

in light sources to minimize this effect.

Coupling is diagnosed most often in the so-called closest-tune method, when the (fractional) horizonal and vertical tune are brought close to each other by tuning the normal quadrupoles suitably. If plotting the tunes as a function of the excitation of the normal quadrupoles we find that for an uncoupled ring the two tune lines cross each other, but for a coupled ring the tunes 'repel' each other, as is illustrated in Fig. 69. The horizontal axis is proportional to the normal quadrupole excitation and as it is changed the betatron frequencies that are plotted on the vertical axis change. Obviously they cannot cross in this case, but in the case of an uncoupled lattice they were able to cross. Therefore, one can experimentally minimize the coupling by exciting correction skew-quadrupoles or solenoids in such a way as to allow the tunes to get closer in a closest-tune scan.

Another way to minimize the tune is to change a horizontal steering magnet and observe a vertical orbit change on beam position monitors, though this is a rather crude measurement.

Coupling can be corrected by introducing suitable magnets that cause coupling by themselves and exciting them in a way as to cancel the coupling generated in unknown or un-wanted ways. A cooler-solenoid is often compensated by a nearby so-called anti-solenoid that has the same integrated excitation $\int Bds$ as the primary solenoid. Strong Detector solenoids are often corrected by skew quadrupoles which are quadrupoles rotated by 45 degrees around their longitudinal axis as is shown in Fig. 70. Placing four skew quadrupoles in the ring will

allow full compensation of all coupling degrees of freedom.

8.5 Chromaticity

The chromaticity was already discussed earlier and is caused by the energy dependence of the focussing strength of the quadrupoles. Higher energy particles have larger foacl lengths and therefore will make the tune energy dependent and will spear out the working point of the accelerator in teh tune diagram. Moreover it will affect the growth rate of instabilities. In a linear collider th chromaticity will cause a dilution of the spot size and counteracts the demand for high luminosities which requires small spot sizes at the interaction point. The correction of the chromaticity is typically done by sextupoles placed at dispersive locations which will create an effective momentum dependent focussing force, that can be tuned to counteract the naturally occuring chromaticity in the quadrupoles.

8.6 Beta function measurements

The beata functions play such a central role in the design of an accelerator that one often wants to directly measure them in order to verify that the magnet lattice really behaves as expected. A simple way to measure the beta functions is to utilize

$$\Delta Q = \frac{\beta}{4\pi\Delta f} = \frac{\Delta k_1 l \beta}{4\pi} \tag{196}$$

where β is the average beta function inside a quadrupole of length l whose excitation k_1 is changed by a small amount Δk_1. Observing the tune change as a function of the excitation change of the quadrupole allows the calculation of the beta function inside the manipulated quadrupole. Often several quadrupoles are powered in series by a single power supply which prevents the independent change of a quadrupole and only the sum of the beta functions of all quadrupoles can be determined. In order to overcome this limitation the quadrupoles in some rings have so-called shunt resistor coupled in parallel that diverts some of the current through the resistor, thereby lowering the current flowing through the quadrupole by a known amount allowing the determination of the quadrupole at the location of a single quadrupole only.

Shunting the quadrupoles has a second very attractive feature: It allows the determination of the misalignmenty of the quadrupole, because a misaligned quadrupole, whose excitation is changed will give the particles a kick proportional to $\Delta k_1 l \Delta x$ if the misalignment Δx is horizontal. The kick can be measured by the beam position system and knowing the quadrupole excitation change Δk_1 will permit the calculation of the misalignment Δx. Of course the vertical misalignment of the quadrupole can be determined in the same way by looking at the vertical orbit.

8.7 Response Matrix

In order to obtain detailed information about quadrupole gradients, BPM, orbit correctors, and other parameters we carefully compare the BPM-corrector response matrix, measured by varying a corrector and observing the changing BPM signal, with model predictions from MAD[8] or other beam optics modeling programs. The response coefficients are calculated by propagating the closed orbit change due to a corrector to a BPM. This leads to $\hat{C}^{ij}_{12/34} = [R^{ij}(1-R^{jj})^{-1}]_{12/34}$, where index 12 refers to the horizontal and 34 to the vertical plane. R^{ij} is the transfer matrix between the corrector and the BPM and R^{jj} is the full-turn matrix starting after the corrector location. In addition, a horizontal corrector kick Θ_j changes the length of the closed orbit by $D_j\Theta_j$, where D_j is the dispersion at the corrector. In the presence of RF the beam's energy changes by $\Delta p/p = -D_j\Theta_j/(\alpha - 1/\gamma^2)C$ in order to keep the revolution frequency constant, where α is the momentum compaction factor, γ is the normalized energy, and C is the circumference. This energy shift will be visible at BPM i as an additional orbit shift, proportional to the dispersion D_i at the BPM. Thus, the complete response is given by

$$C^{ij}_{12} = \left[R^{ij}(1-R^{jj})^{-1}\right]_{12} - \frac{D_i D_j}{(\alpha - 1/\gamma^2)C} \; . \tag{197}$$

Clearly the second term, which is often ignored, will be important in some accelerators, particularly near transition, where $1\gamma^2$ approaches the momentum compaction factor α. In the response matrix approach we can hope to reach very high accuracies due to the large number measurements, which is on the order of $2 \times N_{bpm} \times N_{cor}$ and can be very large, but only fitting for a number of parameters on the order of $2 \times (N_{bpm} + N_{cor}) + N_{quad}$. In such a high precision measurement systematic errors in the response matrix would appear if neglecting the term with the dispersion in Eq. 197. These systematic errors are easily avoided by including the dispersive term.

The analysis is done by two codes, CALIF[33] and LOCO[34]. CALIF sets up equations that relate the measured ($\bar{C}^{ij}$) and the model (C^{ij}) response coefficients in a Taylor-series

$$x^i \bar{C}^{ij} y^j = C^{ij} + \sum_k \frac{\partial C^{ij}}{\partial g_k} \delta g_k \tag{198}$$

where x^i is the BPM scale, y^j is the corrector scale, and g_k is the quadrupole gradient error. The derivative can be calculated in differencing two MAD runs. CALIF then does a linear fit to either $\{x^i, \delta g_k\}$ while keeping y^j fixed or to $\{y^j, \delta g_k\}$ while keeping x^i fixed. This process is iterated while adjusting the BPM resolution errors until the χ^2/DOF is close to unity.

LOCO[34] fits for the deviations of the scales from unity $\Delta x^i, \Delta y^j$, and the

gradient errors δg_k with measured BPM resolution error bars using

$$\bar{C}^{ij} = C^{ij} + \sum_k \frac{\partial C^{ij}}{\partial g_k} \delta g_k + C^{ij} \Delta x^i - C^{ij} y^j \ . \tag{199}$$

As a further refinement LOCO also fits for the energy change from the correctors due to the presence of RF. Equation 199 shows a global degeneracy between BPM and corrector scale errors, which leads to a degenerate linear system, which can, however, be solved using Singular Value Decomposition solvers[26]. Moreover, note that g_k can be any parameter, it need not necessarily be a gradient, but could equally be the length of a drift space, offset in a sextupole or any other parameter.

Equations 198 or 199 can be cast into a linear set of equations of the standard form $\vec{y} = A\vec{x}$, where $\vec{y}$ is based on measured data and $\vec{x}$ is the vector of parameters to be found. The solution is simply $\vec{x} = (A^T A)^{-1} A^T y$, where $(A^T A)^{-1}$ is the covariance matrix[26], containing the errors on the fitted parameters.

9 Instabilities

In the previous sections we treated the particles in an accelerator as independent of each other, and ignored current dependent effects where particles within the beam affect and possibly perturb their neighboring particles. One obvious candidate of such forces are space-charge forces coming from the fact that all particles within a beam have the same charge and therefore repel each other. For low-intensity beams this will play only a minor role but as the beam current or charge per beam increases the repulsion within the beam increases. In this section we will discuss this and other similar effects where parts of the beam affect other parts and often this type of interaction between parts of the beam leads to instabilities. In some cases even to beam loss. The investigation of these intensity dependent effects is often of very practical necessity, because these effects limit the achievable storable beam current and therefore the performance of an accelerator. This is particularly true for modern machines, so-called factories, that are in use or planned to do experiments on nuclear or sub-nuclear reactions with minute cross-sections and require large numbers of accelerated particles in order to achieve useable count-rates in the experiments.

We start by looking at the simplest intensity dependent effect, the repulsion between particles of equal charge within the same beam.

9.1 Space charge

Consider a round beam with radius a of constant charge density ρ. The electric field $\vec{E}$ on a surface of radius r inside the beam can be calculated from Gauss'

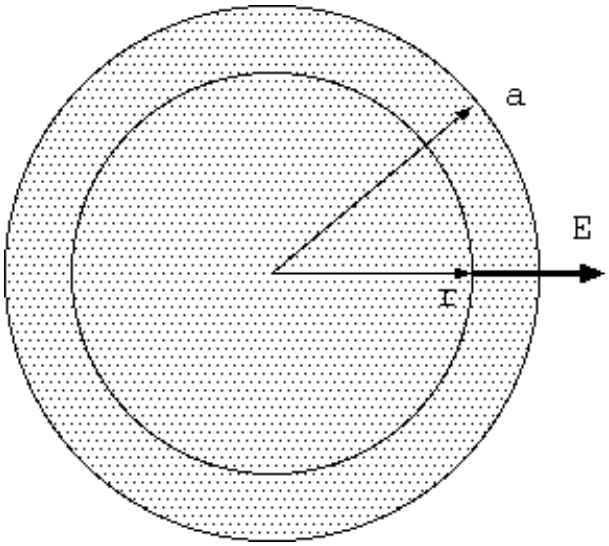

Figure 71: Electric field on the surface of a homogenously charged cyclider.

law by integrating Maxwell's equations

$$\frac{1}{\varepsilon_0}\int \rho d^3V = \int \vec{\nabla}\cdot\vec{E}d^3V = \int \vec{E}\cdot d^2A\ . \tag{200}$$

The first integral can be evaluated to yield $\pi r^2 l\rho/\varepsilon_0$ and the surface integral yields $2\pi rlE$ where E is the radial component of the electric field which evaluates to

$$E = \frac{\rho}{2\varepsilon_0}r = \frac{I}{2\pi a^2\varepsilon_0\beta c}r \tag{201}$$

where we expressed the chrage density ρ by the current I and the speed of the particles βc which are related by $I = \rho\pi a^2\beta c$. We find that the force eE is linear in the distance from the beam center r and repulsive due to the equal charge of all involved particles.

All charged particles in the beam move with speed βc and therefore constitute an electric current I that will create an azimuthal magnetic field B as given by Ampere's law. The current I_r inside the circle of radius r is $I_r = Ir^2/a^2$ and we find for the magnetic field

$$B = \frac{\mu_0 I}{2\pi a^2}r = \frac{I}{2\pi\varepsilon_0 c^2 a^2}r\ . \tag{202}$$

A particle moving at radius r will therefore experience the Lorenz force $F = evB$. The direction of the the magnetic field can be determined from the right-hand rule where the thumb denotes the source particle motion and the other fingers indicate the direction of the magnetic field. The force that a probe-particle experiences

can equally be determined by the other right-hand rule, namely that the thumb denotes the direction of the probe particle and the index finger the direction of the magnetic field. The diretion of the force is then given by the middle finger which in this case points at the center of the source current. In summary we find that the magnetic forces are attractive.

Adding the forces from the electric repulsion and the magnetic attraction we find the total space charge force F to be

$$F = \frac{eI}{2\pi a^2 \varepsilon_0 \beta c} r - \frac{eI\beta c}{2\pi \varepsilon_0 c^2 a^2} r = \frac{eIr}{2\pi a^2 \varepsilon_0 \beta c}(1-\beta^2) = \frac{eIr}{2\pi a^2 \varepsilon_0 \gamma^2 \beta c} \ . \qquad (203)$$

Here we see very clearly that the electric and magnetic forces nearly cancel and especially at high energy with large kinematic factor γ the direct space charge force, as the force described in eq. 203 is called, is strongly suppressed by a factor $1/\gamma^2$. It therefore only plays a role in low-energy accelerators that operate at high intensities.

Note that the direct space-charge force is repulsive and linear in the transverse coordinate r. This will affect the motion of the particles within the beam but not the center of gravity of the entire beam, because the center of the force coincides with the center of the beam. This additional force behaves like a defocussing quadrupole in both planes and will cause the tune of the particles to be reduced by the amount

$$\Delta Q_{x,y} = -\frac{r_0 N}{2\pi \varepsilon_{x,y} \beta^2 \gamma^3} \qquad (204)$$

where r_0 is the classical particle radius and $\varepsilon_{x,y}$ is the emittance in the repsective planes. It should be noted that the tune shift is not visible in the centroid motion of the beam visible on BPM and is called an incoherent tune-shift as opposed to a coherent tune-shift that were visible on the BPM.

The tune shift we discussed in the previous paragraphs was the direct space-charge, because only the beam was involved. In earlier sections we, however, found that the beam environment, and especially the vacuum chamber plays an important role for the electro-magnetic fields generated by the beam which are, amongst other things, also responsible for space-charge forces. In short, the beam pipe will affect the space-charge forces and that is what we will investigate now. We start by considering a beam with line density λ centered between two parallel horizontal plates that represent the vacuum chamber on which the electric field lines have to arrive purely parallel in order to satisfy the boundary conditions on a perfectly conducting metallic surface. This is achieved by introducing a hirarchy of image charges as shown in Fig. 72. Note that the bounadry condition of the lower side of the upper wall is satisfied by the image charge at $2h$. But now we habe two charges that we have to compensate on the lower wall. This is accomplished by introducing an additional image charge at $-4h$ that assures that the field lines emanating from the image charge at $2h$ is perpendicular on

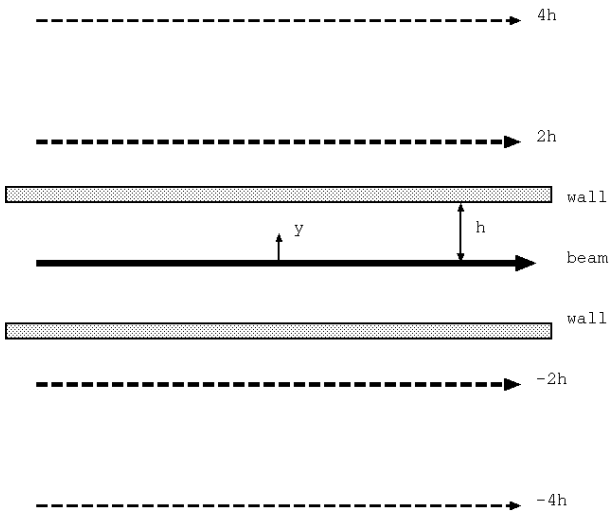

Figure 72: Image currents.

the lower wall. The same argument holds for the images charges at $-2h$ and $4h$. Of course the charges at $\pm 4h$ require compensating charges at $\pm 6h$ and so forth. We therefore arrive at an infinite sum of image charges with alternating polarity. The electric field at a vertical distance y from the center of the beam can then be calculated from

$$E_y = \sum_{n=1}^{\infty} (-1)^n \frac{\lambda}{2\pi\varepsilon_0} \left(\frac{1}{2nh + y} - \frac{1}{2nh - y} \right) = \frac{\lambda}{4\pi\varepsilon_0 h^2} \frac{\pi^2}{12} y \qquad (205)$$

where we made use of $\sum_{n=1}^{\infty} (-1)^n / n^2 = \pi^2/12$. Again we find a vertical force linear in the distance from the center of the beam y that is, in fact, focussing in te sense that is points back towards the center of the beam. The horizontal component of the force can be found from $\operatorname{div} \vec{E} = 0$ and can be found to have the opposite sign but equal magnitude. It is therefore defocussing.

The focussing and defocussing forces due to the image charges that the beam induces in the wall lead to tune shifts, similar to the direct space charge forces. Note, however, that a displaced beam will alter the image charges and thereby the fields. In this way the beam as a whole can affect its own motion that will be visible on BPMs. The physics behind space charge tune shifts was first discussed by L. Laslett [36] and the relative magnitude of the contributions for various beam pipe geometries is characterized by so-called Laslett-coefficients and the respective tune shifts are called Laslett tune-shift.

9.2 Negative Mass Instability

The focussing and defocussing forces discussed affected the tune of the particles and could lead particle losses if the tune was equal to a strong resonance, but no

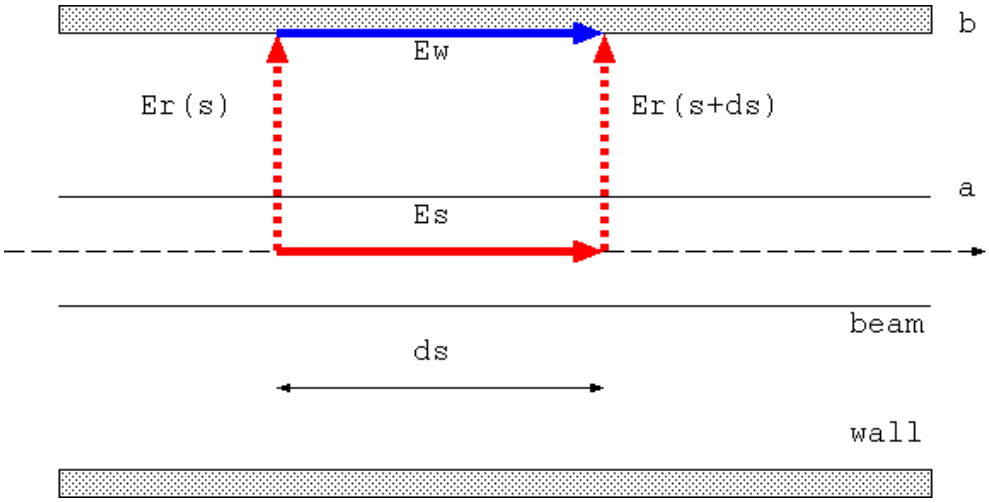

Figure 73: Calculation of the longitudinal field.

feed-back like instability occured. In this section we will now consider a simple system with an unbunched (or coasting) beam in a circular accelerator, whose line charge-density λ is perturbed by a small amount. We will see that this small perturbation in the charge density will induce longitudinal electric fields that in some cases reduces the initial perturbation and in other cases increases the initial perturbation. The latter case then leads to an instability.

We consider a round beam of radius a in a circular beam pipe of radius b and calculate the longitudinal electric field E_s in the center of the beam by

$$\begin{aligned} E_r &= \frac{\lambda}{2\pi\varepsilon_0 r} \quad \text{and} \quad B_\phi = \tfrac{\mu_0\lambda\beta c}{2\pi r} \quad \text{for} \quad r \geq a \\ E_r &= \frac{\lambda r}{2\pi\varepsilon_0 a^2} \quad \text{and} \quad B_\phi = \tfrac{\mu_0\lambda\beta c r}{2\pi a^2} \quad \text{for} \quad r < a \end{aligned} \tag{206}$$

$$\oint \vec{E}\cdot d\vec{l} = \frac{\partial}{\partial t}\int \vec{B}\cdot d\vec{A} \tag{207}$$

$$\begin{aligned} \oint \vec{E}\cdot d\vec{l} &= (E_s - E_w)ds \\ &\quad + \frac{\lambda(s+ds)}{2\pi\varepsilon_0}\left[\int_0^a \frac{r}{a^2}dr + \int_a^b \frac{dr}{r}\right] - \frac{\lambda(s)}{2\pi\varepsilon_0}\left[\int_0^a \frac{r}{a^2}dr + \int_a^b \frac{dr}{r}\right] \\ &= \left[E_s - E_w + \frac{\partial\lambda/\partial s}{4\pi\varepsilon_0}(1+2\ln(b/a))\right]ds \end{aligned} \tag{208}$$

9.3 Wake Fields and Impedances

bla

9.4 Robinson Instability

bla

9.5 Microwave Instability

bla

10 Making Luminosity

bla

10.1 Beam-Target Interaction

bla

10.2 Beam-Beam Interaction

bla

11 Technical Components

bla

11.1 Particle Sources

bla

11.2 Vacuum System

bla

11.3 RF-systems

bla

11.4 Injection

bla

11.5 Electron Cooler

bla

11.6 Stochastic Cooling

bla

12 Magnets

bla

12.1 Normal-conducting Magnets

bla

12.2 Super-conducting Magnets

bla

12.3 Permanent Magnets

bla

13 Synchrotron Radiation

bla

14 Radiation Protection

bla

15 Spin Dynamics

bla

References

[1] D. Edwards, M. Syphers, *An Introduction to the Physics of High Energy Accelerators,* John Wiley and Sons, New York, 1993.

[2] S. Y. Lee, *Accelerator Physics, 2nd ed.,* World Scientific, New Jersey, 2004.

[3] K. Wille, *Physik der Teilchenbeschleuniger und Synchrotronstrahlungsquellen,* Teubner, Stuttgart, 1996.

[4] H. Wiedemann, *Particle Accelerator Physics I and II (2nd ed.)* Springer Verlag, Berlin, 2003.

[5] A. Chao, M. Tigner, *Handbook of accelerator physics and Engineering,* World Scientific, New Jersey, 1999.

[6] `inventors.about.com/library/inventors/blcathoderaytube.htm`

[7] `en.wikipedia.org/wiki/Van_de_Graaff_generator`

[8] `mad.web.cern.ch/mad/`

[9] `www.ansoft.com/maxwellsv/`

[10] `www.als.lbl.gov/`

[11] `www.maxlab.lu.se/`

[12] Karl L. Brown, *A First-and Second-Order Matrix Theory for the Design of Beam Transport Systems and Charged Particle Spectrometers,* SLAC 75, Revision 3, SLAC, 1972.

[13] `pc532.psi.ch/optics.htm`

[14] R. Servranckx, et.al. *Users Guide to the Program DIMAD* unpublished 2001. `www.slac.stanford.edu/accel/nlc/local/AccelPhysics/codes/dimad/`

[15] Trace3D is available at `laacg1.lanl.gov/`

[16] Parmila is available at `laacg1.lanl.gov/`

[17] Winagile can be downloaded from `bryant.home.cern.ch/bryant/`

[18] `www.cs.wisc.edu/~ghost/gsview/index.htm`

[19] M. Abramowitz, I. Stegun, *Handbook of Mathematical Functions,* Dover, New York, 1972.

[20] J. D. Jackson, *Classical Electrodynamics (2nd ed),* John Wiley and Sons, New York, 1975.

[21] H. Klein, *Basic Concepts I,* in Cern Accelerator School on RF Engineering for Particle Accelerators at Exeter College, Oxford, UK, April 1991, CERN Yellow Report 92-03, Vol. I.

[22] `user.web.cern.ch/User/Physics/LibraryArchives/LibraryArchives.html` hosts the CERN Yellow Reports

[23] CERN Accelerator School web-site: `http://cas.web.cern.ch/cas/`

[24] G. Fischer, *Iron Dominted Magnets,* SLAC-PUB-3726. Available at `www.slac.stanford.edu/pubs/slacpubs/3000/slac-pub-3726.html`

[25] E. Beadle, V. Castillo (eds.), *Accelerator Instrumentation, Upton NY, 1989,* AIP Conference proceedings 212, New York, 1990.

[26] W. Press, et. al., *Numerical Recipes,* Cambridge University Press, 1986.

[27] Y. Marti, B. Autin, *Closed orbit correction of A.G. machines using a small number of magnets.* CERN-ISR-MA-73-17, 1973.

[28] C. Borchetta, *Review of Orbit Control,* Proceedings of EPAC98 in Stockholm, abvailable at `www.jacow.org`

[29] R. Brinkmann, *Insertions,* in the 1986 CERN accelerator school in Århus, published as CERN yellow report 87-10, 1987.

[30] K. Brown, *A conceptual Design of Final Focus Systems for Linear colliders,* in Frontiers of Particle Beams, M. Month, S. Turner (eds.), Springer Lecture Notes in Physics 296, Springer Verlag, Berlin, 1986. (slac-pub-4159) http://www.slac.stanford.edu/cgi-wrap/getdoc/slac-pub-4159.pdf

[31] S. Jamison, et.al., *Femtosecond Bunch Length Measurements,* presented at EPAC06 in Edinburgh, June 2006.

[32] R. Trebino, "Frequency Resolved Optical Gating", Kluiwer Academic, Boston, 2000.

[33] W. J. Corbett, M. Lee, V. Ziemann, *A Fast Model Calibration Procedure for Storage Rings,* Proceedings of the Particle Accelerator Conf., Washington, D.C., 1993.

[34] J. Safranek, *Experimental Determination of Storage Ring Optics using Orbit Response Measurements,* BNL note 63382, July 1996.

[35] K. Schindl, *Space Charge,* in Proceedings of the Joint US-CERN-Japan-Russia School on Particle Accelerators in Montreux, 1998.

[36] L. J. Laslett, *On intensity limitations imposed by transverse space-charge effects in circular particle accelerators,* BNL-Report 7534, p. 325.