

Introduction to Particle Accelerators and their Limitations

B.J. Holzer

CERN, Geneva, Switzerland

Abstract

The paper gives an overview of the principles of particle accelerators and their historical development. After introducing the basic concepts, the main emphasis is on sketching the layout of modern storage rings and discussing their limitations in terms of energy and machine performance. Examples of existing machines, among them the Large Hadron Collider (LHC) at CERN, demonstrate the basic principles of and the technical and physical limits that we face in the design and operation of particle colliders. The push for ever higher beam energies motivates the design of future colliders as well as the development of more efficient acceleration techniques.

Keywords

Particle accelerators; history; basic concepts; technologies; technical limits; physical limits.

1 Introduction

The study of matter, from initial theories about the structure of the atom to the discovery of the nucleus and, subsequently, of a variety of particles and their interactions, has been summarized in a scientific picture often called the ‘standard model’; along its way it has driven the development of powerful tools to create the particle beams that are needed to analyse the detailed structure of matter. The largest accelerator to date, the proton–proton Large Hadron Collider (LHC) at CERN in Geneva, operates with an energy per beam of 7 TeV, which corresponds to an available centre-of-mass energy of $E_{\text{cm}} = 14$ TeV. The LHC is part of a long tradition of technical and physical progress in creating particle beams, accelerating them, and achieving successful collision with micrometre beam sizes. This article gives a basic introduction to the physics of particle accelerators and discusses some of their limitations. The author has arbitrarily selected ten limitations to focus on, although in fact there are many more that the reader may find in other publications and which could be studied further (and some of them overcome, hopefully).

1.1 Limit I: The geography

For the fun of it, let us start at the end of the line: the largest accelerators and the fact that beyond physical and technical limits there is a serious boundary condition—the landscape. For a given technology, pushing the particle energy of a storage ring to higher and higher values will necessitate larger and larger machines, and we may suddenly encounter the problem that our device no longer fits in the garage at our institute or, as shown in Fig. 1, not even in the entire region surrounding our facility. As regards LHC [1], the largest storage ring at present, and the Geneva region, the space between Lake Geneva and the Jura mountains defined the size of the tunnel for the present LHC. As a consequence, the maximum feasible beam energy available for high-energy physics experiments is determined by the geographical boundary conditions of the Geneva countryside. Certainly, there are extremely high particle energies in cosmic rays, but you will agree that the accelerators driving these are also much much larger!

Before we discuss the high-energy frontier machines, let us take a brief look at the path paved by the ingenious scientific developments dating back to the discovery of the nucleus by Ernest Rutherford. Figure 2 (taken from [2]) shows a comparison of the scattering events, plotted as a function of the scattering angle, predicted by the Thomson model of the atom and the experimental results obtained by



Fig. 1: The proton–proton collider LHC at CERN, Geneva

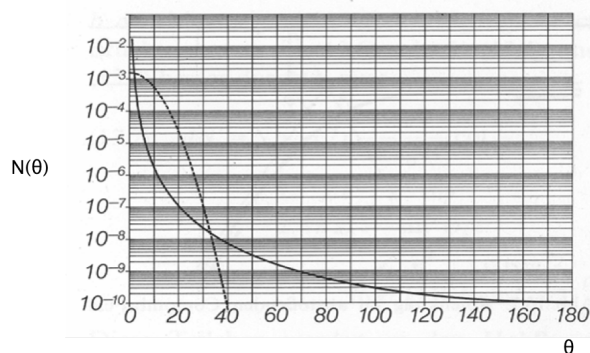


Fig. 2: Comparison of the Thomson model of the atom and the results of Rutherford's experiment; the number of scattering events of alpha particles scattered at the gold foil is plotted as a function of the measured (or predicted) scattering angle.

Rutherford: the discrepancy between the model, which assumes equally distributed charges in matter, and the observed data is evident. As a consequence, the concept of the nucleus was born.

Using alpha particles on the level of MeV is not ideal for precise, triggerable and healthy experiments. So Rutherford discussed with two colleagues, Cockcroft and Walton, the possibility of using artificially accelerated particles. Based on this idea, within only four years Cockcroft and Walton invented the first particle accelerator ever built, and in 1932 they gave the first demonstration of the splitting of a nucleus (lithium) by using a 400 keV proton beam.

Their acceleration mechanism was based on a rectifier or Greinacher circuit, consisting of a number of diodes and capacitors that transformed a relatively small AC voltage to a DC potential which corresponds, depending on the number of diode/capacitor units used, to a multiple of the applied basic potential. The particle source was a standard hydrogen discharge source connected to the high-voltage part of the system, and the particle beam was accelerated to ground potential, hitting the lithium target [3]. A schematic view of the mechanism is shown in Fig. 3 (for details see, for instance, Ref. [4]), and a photograph of such a device which has been used at CERN for many years is presented in Fig. 4. Cockcroft and Walton were awarded the Nobel prize for their invention.

1.2 Limit II: Voltage breakdown in DC accelerators

In parallel to Cockcroft and Walton, but based on a completely different technique, another type of DC accelerator had been invented: Van de Graaff designed a DC accelerator [5] that used a mechanical

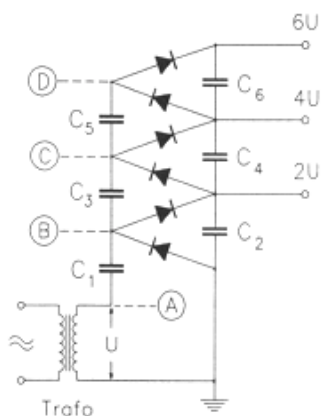


Fig. 3: Schematic layout of the rectifier circuit used by Cockcroft and Walton to generate the high DC voltage needed for their accelerator.

transport system to carry charges, sprayed on a belt or chain, to a high-voltage terminal; see Fig. 5, taken from Ref. [6].

In general these machines can reach higher voltages than the Cockcroft–Walton devices, but they are more limited in terms of particle intensity. Common to all DC accelerators is the limitation on the achievable beam energy due to high-voltage breakdown effects (discharges). Without using an insulating gas (SF_6 in most cases), electric fields will be limited to about 1 MV m^{-1} , and even with the most sophisticated devices, like the one in Fig. 6, acceleration voltages on the order of MV cannot be overcome. In fact, the example in Fig. 6 shows an approach that has been applied in a number of situations: injecting a negative ion beam (even H^- is used) and stripping the ions in the middle of the high-voltage terminal allows one to profit from the potential difference twice and thus to make another step of gain in beam energy.

Given the obvious limitations of the DC machines described above, the next step forward is natural. In 1928, Widerøe developed the concept of a AC accelerator. Instead of rectifying the AC voltage, he connected a series of acceleration electrodes in an alternating manner to the output of an AC supply. The schematic layout is shown in Fig. 7 where, for a instant in time, the direction of the electric field is indicated. In principle, this device can produce step by step a multiple of the acceleration voltage, as long as for the negative half-wave of the AC voltage the particles are shielded from the decelerating field.

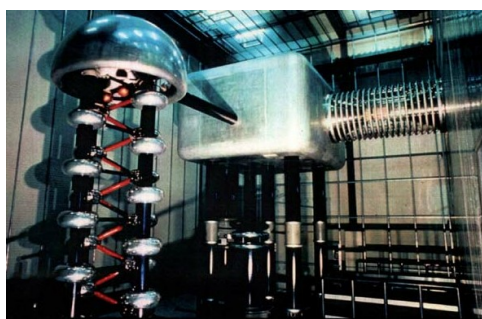


Fig. 4: A Cockcroft–Walton generator that was used at CERN as a pre-accelerator for the proton beams; the device has since been replaced by the more compact and efficient RFQ technique.

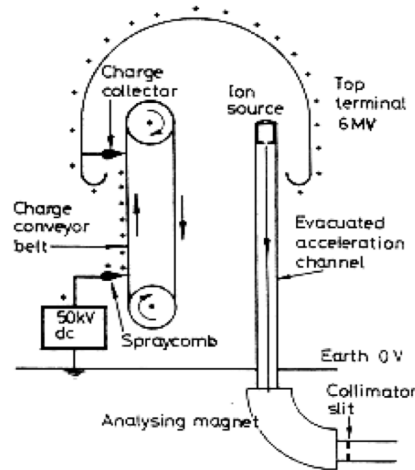


Fig. 5: Schematic design of a Van de Graaff accelerator



Fig. 6: A typical example of a tandem Van de Graaff accelerator; these are very reliable machines for precise measurements in atomic and nuclear physics. (Photo courtesy of the Max Planck Institute for Nuclear Physics, Heidelberg.)

The energy gain after the n th step is therefore

$$E_n = n \cdot q \cdot U_0 \cdot \sin(\psi_s), \quad (1)$$

where n denotes the acceleration step, q the charge of the particle, U_0 the applied voltage per gap, and ψ_s the phase between the particle and the changing AC voltage.

1.3 Limit III: The size of the accelerating structure

A key quantity in such a Widerøe structure is the length of the drift tubes that will protect the particles from the negative half-wave of the sinusoidal AC voltage. For a given frequency of the applied radio-frequency (RF) voltage, the length of the drift tube is defined by the speed of the particle and the duration of the negative half-wave of the sinusoidal voltage, as shown in Fig. 8.

The time-span of the negative half-wave is defined by the applied frequency, $\Delta t = \tau_{\text{rf}}/2$, so for the length of the n th drift tube we get

$$l_n = v_n \cdot \frac{\tau_{\text{rf}}}{2}. \quad (2)$$

Given the kinetic energy of the particle,

$$E_{\text{kin}} = \frac{1}{2}mv^2, \quad (3)$$

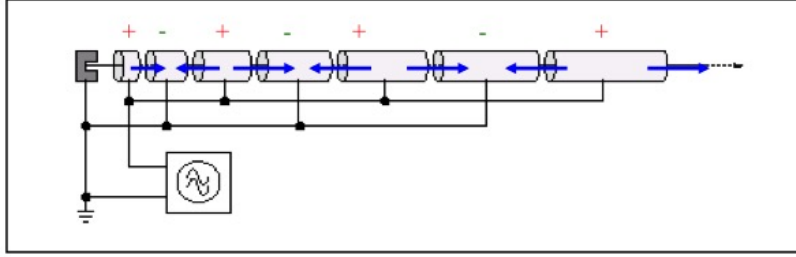


Fig. 7: Schematic view of the Widerøe principle as a fundamental concept for AC (or RF) acceleration

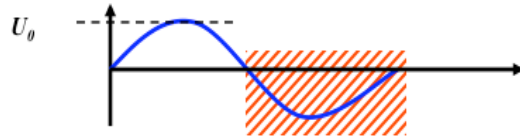


Fig. 8: The frequency (and hence the period) of the RF system and the particle speed determine the length of the drift tubes in the Widerøe structure.



Fig. 9: Unilac at GSI, Darmstadt; clearly visible are the structure of the drift tubes and their increasing length as a function of the particle energy.

we obtain directly that

$$l_n = \frac{1}{\nu_{rf}} \cdot \sqrt{\frac{nqU_0 \sin \psi_s}{2m}}, \quad (4)$$

which defines the design concept of the machine. Figure 9 shows a photograph of such a device, the Unilac at the Institute for Heavy Ion Research (GSI) in Darmstadt, Germany.

Two remarks should be made in this context.

- The short derivation here is based on the classical approach, and in fact these accelerators are usually optimum for ‘low-energetic’ proton or heavy-ion beams. Typical beam energies (referring to protons) are on the order of 10 MeV; for example, the present Linac 2 at CERN delivers the protons for LHC operation with an energy of 50 MeV, corresponding to a relativistic β of 0.31.
- For higher energies, even in the case of protons or ions, the speed will at some point approach the speed of light, and the length of the drift tubes and hence the dimension of the whole accelerator will reach a size that may no longer be feasible. More advanced ideas are needed in order to keep the machine within reasonable dimensions, and the next natural step in the historical development was to introduce magnetic fields and bend the particle beam into a circle.

2 Pushing for the highest energies: synchrotrons and storage rings

A significant step forward in achieving high beam energies involves the use of circular structures. In order to apply over and over again the accelerating fields, we try to bend the particles onto a circular path and so bring them back to the RF structure where they will receive the next step-up in energy. To do this, we introduce magnetic (or electric) fields that will deflect the particles and keep them on a well-defined orbit during the complete acceleration process. The Lorentz force that acts on a particle will therefore have to compensate exactly the centrifugal force due to the bent orbit. In general, we can write

$$\mathbf{F} = q \cdot (\mathbf{E} + \mathbf{v} \times \mathbf{B}). \quad (5)$$

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

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

$$F_{\text{Lorentz}} = e \cdot v \cdot B, \quad (6)$$

$$F_{\text{centrifugal}} = \frac{\gamma m_0 v^2}{\rho}. \quad (7)$$

Assuming an idealized homogeneous dipole oriented along the particle orbit, we define the condition for a perfect circular orbit as equality between these two forces; this yields the following condition for the idealized ring:

$$\frac{p}{e} = B \cdot \rho, \quad (8)$$

where we refer to protons and have accordingly set $q = e$. This condition relates the so-called beam rigidity $B\rho$ to the particle momentum that can be carried in the storage ring, and it will ultimately define, for a given magnetic field of the dipoles, the size of the storage ring.

In reality, instead of a continuous dipole field the storage ring will be built out of several dipoles, powered in series to define the geometry of the ring. For a single magnet, the particle trajectory is shown schematically in Fig. 10. In the free space outside the dipole magnet, the particle trajectory follows a straight line. As soon as the particle enters the magnet, it is bent onto a circular path until it leaves the magnet at the other side.

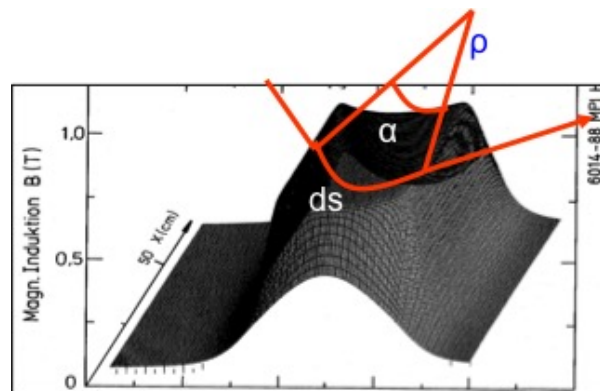


Fig. 10: Dipole field of a storage ring and schematic path of the particles

2.1 Limit IV: The magnetic guide field

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

$$\alpha = \frac{ds}{\rho} = \frac{B ds}{B \cdot \rho}. \quad (9)$$

We therefore require that

$$\frac{\int B dl}{B \cdot \rho} = 2\pi. \quad (10)$$

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

In the case of the LHC, the dipole field has been pushed to the highest achievable values; 1232 superconducting dipole magnets, each of length 15 m, define the geometry of the ring and, via Eq. (10), the maximum momentum for the stored proton beam. Using the equation given above, for a maximum momentum of $p = 7 \text{ TeV}/c$ we obtain a required magnetic field of

$$B = \frac{2\pi \cdot 7000 \cdot 10^9 \text{ eV}}{1232 \cdot 15 \text{ m} \cdot 2.99792 \cdot 10^8 \text{ m s}^{-1}}, \quad (11)$$

or

$$B = 8.33T, \quad (12)$$

to bend the beams. For convenience we have expressed the particle momentum in units of GeV/c . Figure 11 shows a photograph of the LHC dipole magnets, built out of superconducting NbTi filaments, which are operated at a temperature of $T = 1.9 \text{ K}$.



Fig. 11: Superconducting quadrupole of the LHC storage ring

2.2 Focusing properties

In addition to the main bending magnets that guide the beam onto a closed orbit, focusing fields are needed to keep the particles close together. In modern storage rings and light sources, the particles are kept in the machine for many hours, and a carefully designed focusing structure is needed to maintain the necessary beam size at different locations in the ring.

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

$$B_x = g \cdot y, \quad B_y = g \cdot x. \quad (13)$$

The constant g is called the gradient of the magnetic field and characterizes the focusing strength of the magnetic lens in both transverse planes. For convenience it is (like the dipole field) normalized to the particle momentum. The normalized gradient is denoted by k and defined as

$$k = \frac{g}{p/e} = \frac{g}{B\rho}. \quad (14)$$

The technical layout of such a quadrupole is depicted in Fig. 11. As in the case of the LHC dipoles, the quadrupole magnet is built in superconducting technology.

Now that we have defined the basic building blocks of a storage ring, we need to arrange them in a so-called magnet lattice and optimize the field strengths in such a way as to obtain the required beam parameters. An example of what such a magnet lattice looks like is given in Fig. 12. This photograph shows the dipole (orange) and quadrupole (red) magnets in the TSR storage ring in Heidelberg. Eight dipoles are used to bend the beam in a ‘circle’, and the quadrupole lenses between them provide the focusing to keep the particles within the aperture limits of the vacuum chamber.



Fig. 12: TSR storage ring, Heidelberg, as a typical example of a separate-function strong focusing storage ring

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

$$\frac{B(x)}{p/e} = \frac{1}{\rho} + k \cdot x + \frac{1}{2!}mx^2 + \frac{1}{3!}nx^3 + \dots. \quad (15)$$

Following the arguments above, for the moment we take into account only constant (dipole) or linear (quadrupole) terms. The higher-order field contributions will be treated later as (hopefully) small perturbations.

The particles will now follow the ‘circular’ path defined by the dipole fields, and in addition will undergo harmonic oscillations in both transverse planes. The situation is shown schematically in Fig. 13. An ideal particle will follow the design orbit that is represented by the circle in the diagram. Any other particle will perform transverse oscillations under the influence of the external focusing fields, and the amplitude of these oscillations will ultimately define the beam size.

Unlike a classical harmonic oscillator, however, the equations of motion in the horizontal and vertical planes differ somewhat. Assuming a horizontal focusing magnet, the equation of motion is

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

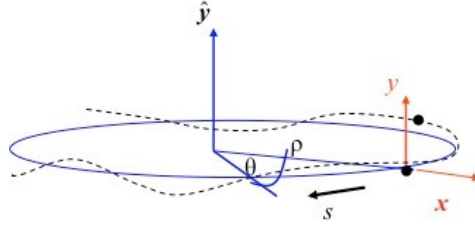


Fig. 13: Coordinate system used in particle beam dynamics; the longitudinal coordinate s moves around the ring with the particle considered.

where k is the normalized gradient introduced above and the $1/\rho^2$ term represents the so-called weak focusing, which is a property of the bending magnets. In the vertical plane, on the other hand, due to the orientation of the field lines and by Maxwell's equations, the forces instead have a defocusing effect; also, the weak focusing term disappears:

$$y'' - y \cdot k = 0. \quad (17)$$

The principal problem arising from the different directions of the Lorentz force in the two transverse planes of a quadrupole field is sketched in Fig. 14. It is the task of the machine designer to find an adequate solution to this problem and to define a magnet pattern that will provide an overall focusing effect in both transverse planes.

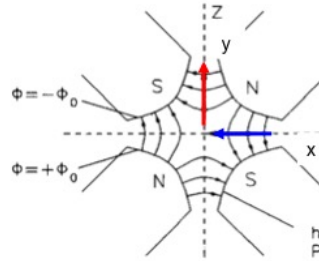


Fig. 14: Field configuration in a quadrupole magnet and the direction of the focusing and defocusing forces in the horizontal and vertical planes.

Following closely the example of the classical harmonic oscillator, we can write down the solutions of the above equations of motion. For simplicity, we focus on the horizontal plane; a 'focusing' magnet is therefore focusing in this horizontal plane and at the same time defocusing in the vertical plane. Starting with initial conditions for the particle amplitude x_0 and angle x'_0 in front of the magnet element, we obtain the following relations for the trajectory inside the magnet:

$$x(s) = x_0 \cdot \cos(\sqrt{|K|} s) + x'_0 \cdot \frac{1}{\sqrt{|K|}} \sin(\sqrt{|K|} s), \quad (18)$$

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

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

$$\begin{pmatrix} x \\ x' \end{pmatrix}_s = \mathbf{M}_{\text{foc}} \begin{pmatrix} x \\ x' \end{pmatrix}_0, \quad (20)$$

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

$$\mathbf{M}_{\text{foc}} = \begin{pmatrix} \cos(\sqrt{|K|} s) & \frac{1}{\sqrt{|K|}} \sin(\sqrt{|K|} s) \\ -\sqrt{|K|} \sin(\sqrt{|K|} s) & \cos(\sqrt{|K|} s) \end{pmatrix}.$$

Schematically, the situation is visualized in Fig. 15.

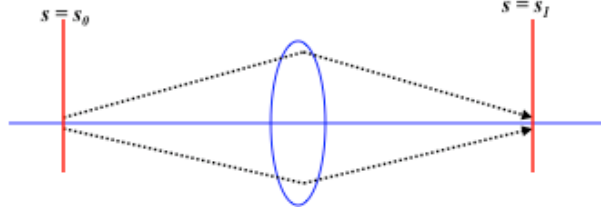


Fig. 15: Schematic principle of the effect of a focusing quadrupole magnet

In the case of a defocusing magnet, we obtain analogously that

$$\begin{pmatrix} x \\ x' \end{pmatrix}_s = \mathbf{M}_{\text{defoc}} \begin{pmatrix} x \\ x' \end{pmatrix}_0, \quad (21)$$

with

$$\mathbf{M}_{\text{defoc}} = \begin{pmatrix} \cosh(\sqrt{|K|} s) & \frac{1}{\sqrt{|K|}} \sinh(\sqrt{|K|} s) \\ \sqrt{|K|} \sinh(\sqrt{|K|} s) & \cosh(\sqrt{|K|} s) \end{pmatrix};$$

see Fig. 16.

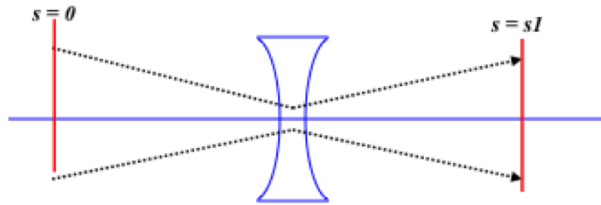


Fig. 16: Schematic principle of the effect of a defocusing quadrupole magnet

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

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

This matrix formalism allows us to combine the elements of a storage ring in an elegant way and so it is straightforward to calculate the particle trajectories. As an example, we consider the simple case of an alternating focusing and defocusing lattice, a so-called FODO lattice [4]; see Fig. 17.

As we know the properties of each and every element in the accelerator, we can construct the corresponding matrices and calculate step by step the amplitude and angle of a single particle's trajectory around the ring. Even more conveniently, we can multiply out the different matrices and, given initial conditions x_0 and x'_0 , obtain directly the trajectory at any location in the ring:

$$\mathbf{M}_{\text{total}} = \mathbf{M}_{\text{foc}} \cdot \mathbf{M}_{\text{drift}} \cdot \mathbf{M}_{\text{dipole}} \cdot \mathbf{M}_{\text{drift}} \cdot \mathbf{M}_{\text{defoc}} \cdots \quad (22)$$

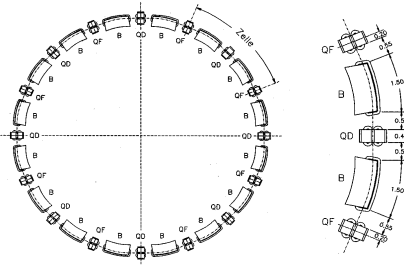


Fig. 17: A simple periodic chain of bending magnets and focusing/defocusing quadrupoles forming the basic structure of a storage ring.

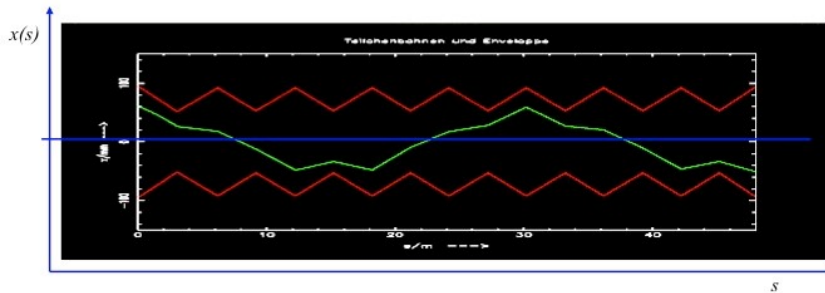


Fig. 18: Calculated particle trajectory in a simple storage ring

The trajectory thus obtained is shown schematically in Fig. 18.

We emphasize the following facts in this context.

- At each moment, or in each lattice element, the trajectory is a part of a harmonic oscillation.
- However, due to the different restoring or defocusing forces, the solution will look different at each location.
- In the linear approximation that we make use of in this context, all particles experience the same external fields, and their trajectories will differ only because of their different initial conditions.
- There seems to be an overall oscillation in both transverse planes while the particle is travelling around the ring. Its amplitude stays well within the boundaries set by the vacuum chamber, and its frequency in the example of Fig. 18 is roughly 1.4 transverse oscillations per revolution, which corresponds to the eigenfrequency of the particle under the influence of the external fields.

Coming closer to a real existing machine, we show in Fig. 19 the orbit measured during one of the first injections into the LHC storage ring. The horizontal oscillations are plotted in the upper half of the figure and the vertical oscillations in the lower half, on a scale of ± 10 mm. Each histogram bar indicates the value recorded by a beam position monitor at a certain location in the ring, and the orbit oscillations are clearly visible. By counting (or, better, fitting) the number of oscillations in both transverse planes, we obtain values of

$$Q_x = 64.31, \quad Q_y = 59.32. \quad (23)$$

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

As the tune characterizes the particle oscillations under the influence of all external fields, it is one of the most important parameters of the storage ring. Therefore it is usually displayed and controlled at

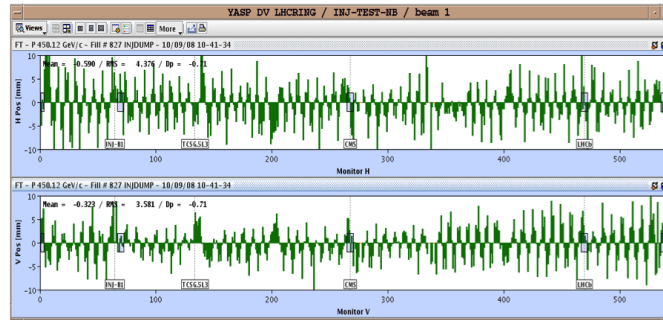


Fig. 19: Measured orbit in LHC during the commissioning of the machine

all times by the control system of such a machine. As an example, Fig. 20 shows the tune diagram of the HERA proton ring [7]; it was obtained via a Fourier analysis of the spectrum measured from the signal of the complete particle ensemble. The peaks indicate the two tunes in the horizontal and vertical planes of the machine, and in a sufficiently linear machine a fairly narrow spectrum is obtained.



Fig. 20: Tune signal of a proton storage ring (HERA-p)

Briefly referring back to Fig. 18, the question is what the trajectory of the particle will look like for the second turn, or the third, or after an arbitrary number of turns. Now, as we are dealing with a circular machine, the amplitude and angle, x and x' , at the end of the first turn will be the initial conditions for the second turn, and so on. After many turns the overlapping trajectories begin to form a pattern, such as that in Fig. 21, which indeed looks like a beam having here and there a larger and smaller beam size but still remaining well-defined in its amplitude by the external focusing forces.

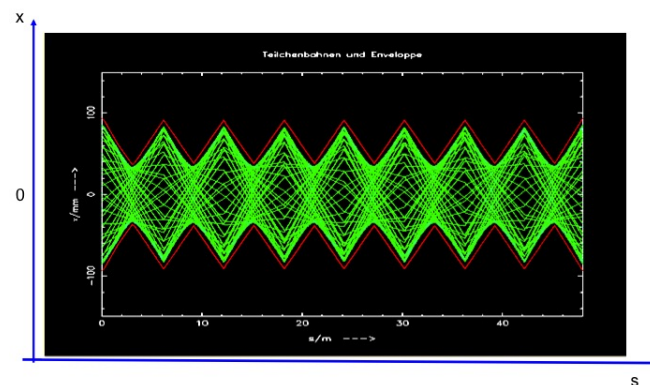


Fig. 21: Many single-particle trajectories together form a pattern that corresponds to the beam size in the ring

To make a long story short [8], a mathematical function, which we call β or amplitude function, can be defined that describes the envelope of the single-particle trajectories. With this new variable, we can rewrite the equation for the amplitude of a particle's transverse oscillations as

$$x(s) = \sqrt{\epsilon} \sqrt{\beta(s)} \cos(\psi(s) + \phi), \quad (24)$$

where ψ is the phase of the oscillation, ϕ is its initial condition, and ϵ is a characteristic parameter of the single particle or, if we are considering a complete beam, of the ensemble of particles. Indeed, ϵ describes the space occupied by the particle in the transverse (here simplified two-dimensional) (x, x') phase space. More specifically, the area in (x, x') space that is covered by the particle is given by

$$A = \pi \cdot \epsilon, \quad (25)$$

and, as long as we consider conservative forces acting on the particle, this area is constant according to Liouville's theorem. Here we take these facts as given, but we point out that, as a direct consequence, the so-called emittance ϵ cannot be influenced by whatever external fields are applied; it is a property of the beam, and we have to take it as given and handle it with care.

To be more precise, and following the usual textbook treatment of accelerators, we can draw in phase space the ellipse of the particle's transverse motion; see, for example, Fig. 22. While the shape and orientation are determined by the optics function β and its derivative, $\alpha = -\frac{1}{2}\beta'$, the area covered is constant.

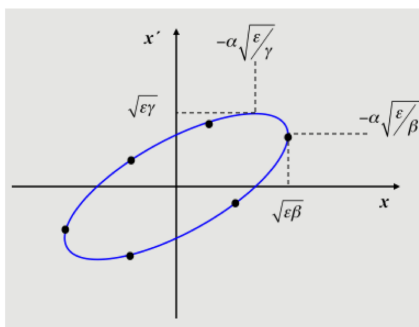


Fig. 22: Ellipse in (x, x') phase space

Let us talk a bit more about the beam as an ensemble of many (typically 10^{11}) particles. Referring to Eq. (24), at a given position in the ring the beam size is defined by the emittance ϵ and the function β . Thus, at a certain moment in time the cosine term in (24) will be 1 and the trajectory amplitude will reach its maximum value. Now, if we consider a particle at one standard deviation (sigma) of the transverse density distribution, then using the emittance of this reference particle we can calculate the size of the complete beam, in the sense that the complete area (within one sigma) of all particles in (x, x') phase space is surrounded (and thus defined) by our one-sigma candidate. Thus the value $\sqrt{\epsilon \cdot \beta(s)}$ will define the one-sigma beam size in the transverse plane. As an example, we use the values for the LHC proton beam: in the periodic pattern of the arc, the beta function is $\beta = 180$ m and the emittance at flat-top energy is roughly $\epsilon = 5 \times 10^{-10}$ rad m. The resulting typical beam size is therefore 0.3 mm. Now, clearly we would not design a vacuum aperture of the machine based on a one-sigma beam size; typically, an aperture requirement corresponding to 12σ is a good rule to guarantee a sufficient aperture, allowing for tolerances from magnet misalignments, optics errors and operational flexibility. In Fig. 23 the LHC vacuum chamber is shown, including the beam screen used to protect the cold bore from synchrotron radiation; it corresponds to a minimum beam size of 18σ .

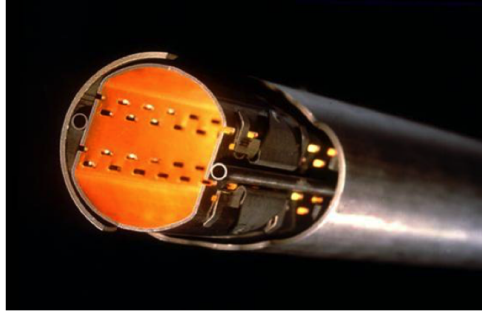


Fig. 23: The LHC vacuum chamber with the beam screen to shield the superconducting magnet bore from synchrotron radiation.

3 Particle colliders

3.1 Limit V: Fixed-target collider

The easiest way to perform physics experiments with particle accelerators is to bang the accelerated beam onto a target and analyse the resulting events. While nowadays in high-energy physics we do not apply this technique any more, it still plays an essential role in the regime of atomic and nuclear physics experiments. The advantage is that it is quite simple once the accelerator has been designed and built, and the particles produced are easily separated due to the kinematics of the reaction. The situation is illustrated in Fig. 24. The particle ‘a’ that is produced and accelerated in the machine is directed onto the particle ‘b’, which is at rest in the laboratory frame. The particles produced from this collision are labelled ‘c’ and ‘d’ in this example.

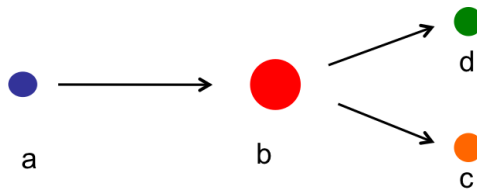


Fig. 24: Schematic diagram of fixed-target collider

While the set-up of such a scheme is quite simple, it is worth taking a closer look at the available energy in the centre-of-mass system. The relativistic overall energy is given by

$$E^2 = p^2 c^2 + m^2 c^4, \quad (26)$$

which holds for a single particle but is equally valid for an ensemble of particles. Most important, the rest energy of the particle ensemble is constant (and is sometimes called the ‘invariant mass of the system’).

Considering the system of two particles colliding, we can write

$$(E_a^{\text{cm}} + E_b^{\text{cm}})^2 - (p_a^{\text{cm}} + p_b^{\text{cm}})^2 c^2 = (E_a^{\text{lab}} + E_b^{\text{lab}})^2 - (p_a^{\text{lab}} + p_b^{\text{lab}})^2 c^2. \quad (27)$$

In the frame of the centre-of-mass system we get, by definition,

$$p_a^{\text{cm}} + p_b^{\text{cm}} = 0, \quad (28)$$

while in the laboratory frame where particle ‘b’ is at rest we have simply

$$p_b^{\text{lab}} = 0. \quad (29)$$

The equation for the invariant mass therefore simplifies to

$$W^2 = (E_a^{\text{cm}} + E_b^{\text{cm}})^2 = (E_a^{\text{lab}} + m_b \cdot c^2)^2 - (p_a^{\text{lab}} \cdot c^2). \quad (30)$$

In other words, the energy that is available in the centre-of-mass system depends on the square root of the energy of particle ‘a’, which is the energy provided by the particle accelerator:

$$W \approx \sqrt{2E_a^{\text{lab}} \cdot m_b \cdot c^2} \quad (31)$$

—a quite unsatisfactory situation!

To meet the demand for higher and higher energies in particle collisions, the design of modern high-energy accelerators has naturally concentrated on the development of particle colliders, where two counter-rotating beams are brought into collision at one or several interaction points (Fig. 25).



Fig. 25: Schematic diagram of the collision of two particles with equal energy

If we calculate the available energy in the centre-of-mass system for the case of two colliding beams of identical particles, we get

$$(p_a^{\text{cm}} + p_b^{\text{cm}})^2 = 0 \quad (32)$$

and, by symmetry, also

$$(p_a^{\text{lab}} + p_b^{\text{lab}})^2 = 0. \quad (33)$$

So the full energy delivered to the particles in the accelerator is available during the collision process:

$$W = E_a^{\text{lab}} + E_b^{\text{lab}} = 2 E_a^{\text{lab}}. \quad (34)$$

A ‘typical’ example of a high-energy physics event in such a collider is shown in Fig. 26.

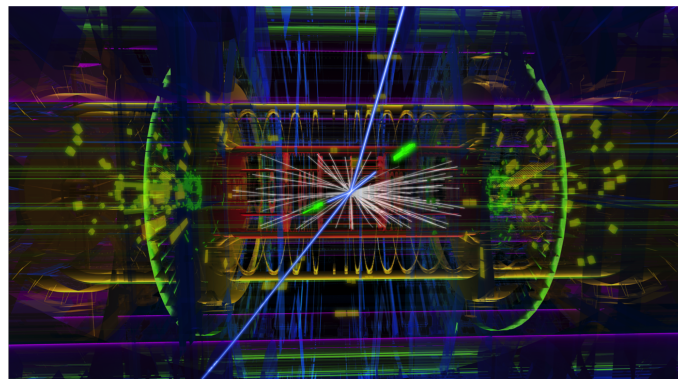


Fig. 26: ‘Typical’ event observed in a collider ring—a Higgs particle measured in the ATLAS detector

3.2 Limit VI: The unavoidable particle detectors

While it is quite clear that a particle collider ring is a magnificent machine in the quest for higher energies, there is a small problem involved, namely the ‘particle detector’. In the arc of the storage ring we can usually find a nice pattern of magnets providing us with a well-defined beam size, expressed as the beta

function. However, special care has to be taken when our colleagues from high-energy physics wish to install a particle detector. Especially when working at the energy frontier, just like for the accelerators, these devices tend to expand considerably in size with the energies required. In Fig. 27, the largest particle detector installed in a storage ring is shown as an impressive example: the ATLAS detector at the LHC.

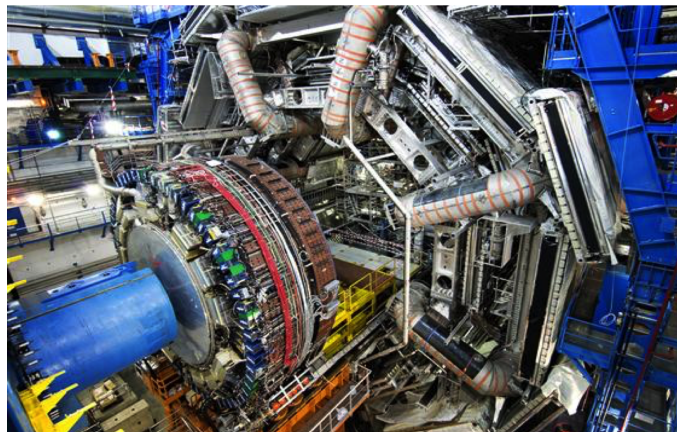


Fig. 27: ATLAS detector at the LHC, which is 46 m in length and has an overall weight of 7000 t

The storage ring has to be designed to provide the space needed for the detector hardware and at the same time create the smallest achievable beam spots at the collision point, which is usually right in the centre of the detector. Unfortunately these requirements are a bit contradictory. The equation for the luminosity of a particle collider depends on the stored beam currents and the transverse spot size of the colliding beams at the interaction point (IP):

$$L = \frac{1}{4\pi e^2 f_0 b} \cdot \frac{I_1 I_2}{\sigma_x^* \sigma_y^*}. \quad (35)$$

At the same time, however, the beta function in a symmetric drift grows quadratically as a function of the distance between the beam waist and the first focusing element, i.e.

$$\beta(s) = \beta^* + \frac{s^2}{\beta^*}. \quad (36)$$

The smaller the beam at the IP, the faster it will grow until we can apply—outside of the detector region—the first quadrupole lenses. As a consequence, this trend sets critical limits on the achievable quadrupole aperture or, for a given aperture, the achievable quadrupole gradient. The focusing lenses right before and after the IP, being placed as close as possible to the detector, are generally the most critical and most expensive magnets in the machine, and their aperture requirement ultimately determines the luminosity that can be delivered by the storage ring.

For the experts we would like to add that even if the bare aperture requirement can be fulfilled, the resulting chromaticity that is created in the mini-beta insertion and the sextupole strengths that are needed to correct for it usually pose the next limit that we will face.

3.3 Limit VII: The relative rareness of Nobel prize-winning reactions

The rate of events produced in a particle collision process depends not only on the performance of the colliding beams but first and foremost on the probability of creating such an event, the so-called cross-section of the process. In the case of the Higgs particle, which is without doubt the highlight of LHC Run 1, the overall cross-section is displayed in Fig. 28.

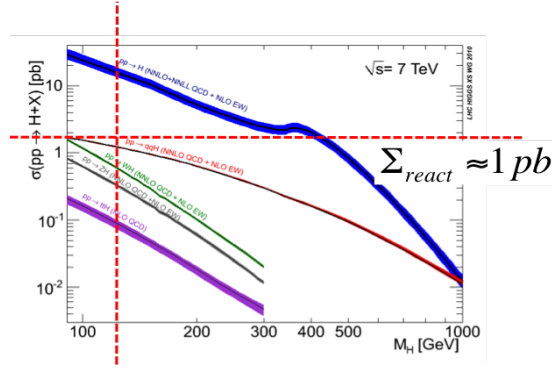


Fig. 28: Cross-section of the Higgs particle for different production processes, courtesy of the CMS collaboration

Without going into details, we can state that the cross-section for Higgs production is on the order of

$$\Sigma_{\text{react}} \simeq 1 \text{ pb}. \quad (37)$$

During the three years of LHC Run 1, i.e. the period 2011–2013, an overall luminosity of

$$\int L dt = 25 \text{ fb}^{-1} \quad (38)$$

was accumulated.

Combining these two numbers using the fact that the event rate of a reaction is $R = L \cdot \Sigma_{\text{react}}$, we get a total number of ‘some thousand’ Higgs particles produced—for a Nobel prize-winning event just at the edge of reliable statistics. Therefore, the particle colliders have to be optimized not only for the highest achievable energies but also for maximum stored beam currents and small spot sizes at the interaction points so as to optimize the luminosity of the machine.

3.4 Limit VIII: The luminosity of a collider ring

Following the arguments above, the design goal here is to prepare, accelerate and store two counter-rotating particle beams in order to profit best from the energy of the two beams during the collision process. Still, there is a price to pay: unlike in fixed-target experiments, where the ‘particle’ density of the target material is extremely high, in the case of two colliding beams the event rate is basically determined by the transverse particle density that can be achieved at the IP. Assuming Gaussian density distributions in both transverse planes, the performance of such a collider is described by the luminosity

$$L = \frac{1}{4\pi e^2 f_0 b} \cdot \frac{I_1 I_2}{\sigma_x^* \sigma_y^*}. \quad (39)$$

While the revolution frequency f_0 and the bunch number b are ultimately determined by the size of the machine, the stored beam currents I_1 and I_2 and the beam sizes σ_x^* and σ_y^* at the IP have their own limitations.

The most serious limitation comes from the beam–beam interaction itself. During the collision process, individual particles of the counter-rotating bunches feel the space charge of the opposing bunch. In the case of a proton–proton collider, this strong field acts like a defocusing lens, and has a strong impact on the tune of the bunches [9].

In Fig. 29 the situation is shown schematically. Two bunch trains collide at the IP, and during the collision process a direct beam–beam effect is observed. In addition to that, before and after the actual collision, long-range forces exist between the bunches that have a nonlinear component; see Fig. 30.

As a consequence [9], the tune of the beams is not only shifted with respect to the natural tune of the machine but also spread out, as different particles inside the bunches see different contributions from the beam–beam interaction.

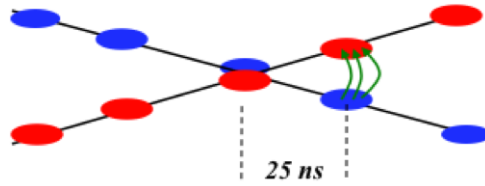


Fig. 29: Schematic view of the beam–beam interaction during the crossing of bunch trains

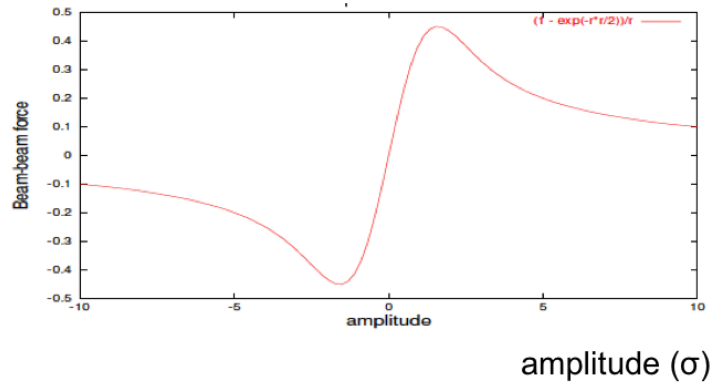


Fig. 30: Beam–beam force as a function of the transverse distance of the particle to the centre of the opposing bunch.

Therefore, in the tune diagram, we no longer obtain a single spot representing the ensemble of particles, but rather a large array that depends in shape, size and orientation on the particle densities, the distance of the bunches at the long-range encounters, and the single-bunch intensities. The effect has been calculated for the LHC and is displayed in Fig. 31.

In a number of cases a useful approximation can be applied, as for distances of about $1-2\sigma$, the beam–beam force in Fig. 30 can be linearized and acts like a quadrupole lens. Accordingly, a tune shift can be calculated to characterize the strength of the beam–beam effect in a collider. Given the parameters described above, and introducing the classical particle radius r_p , the amplitude function β^* at the IP and the Lorentz factor γ , we can express the tune shift due to the linearized beam–beam effect as

$$\Delta Q_y = \frac{\beta_y^* \cdot r_p \cdot N_p}{2\pi \gamma (\sigma_x + \sigma_y) \sigma_y}. \tag{40}$$

In the case of the LHC, the design value of the beam–beam tune shift is $\Delta Q = 0.0033$, and in daily operation the machine is optimized to run close to this value, which places the ultimate limit on achievable bunch intensities in the collider.

4 Lepton colliders

4.1 Limit IX: Synchrotron light—the drawback of electron storage rings

In proton or heavy-ion storage rings, the design can more or less follow the rules discussed above. But the situation changes drastically as the particles become more and more relativistic. Bent on a circular

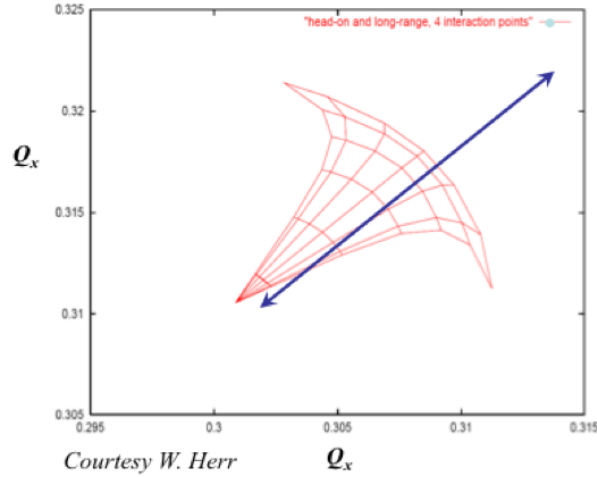


Fig. 31: Calculated tune shift due to beam–beam interaction in the LHC

path, electrons in particular will radiate an intense light, the so-called synchrotron radiation, which will have a strong influence on the beam parameters as well as on the design of the machine.

Summarizing the situation briefly here, the power loss due to synchrotron radiation depends on the bending radius and the energy of the particle beam:

$$P_s = \frac{2}{3} \alpha \hbar c^2 \frac{\gamma^4}{\rho^2}, \tag{41}$$

where α represents the fine structure constant and ρ the bending radius in the dipole magnets of the ring. As a consequence, the particles will lose energy turn by turn. To compensate for these losses, RF power has to be supplied to the beam at any moment. An example that illustrates the problem nicely is shown in Fig. 32. It plots the horizontal orbit of the former Large Electron–Positron Collider (LEP) storage ring. The electrons, travelling from right to left in the plot, lose a considerable amount of energy in each arc and hence deviate from the ideal orbit towards the inner side of the ring. The effect on the orbit is large: up to 5 mm orbit deviation was observed in the example of Fig. 32. In order to compensate for these losses, four RF stations were installed in the straight sections of the ring to supply the necessary power.

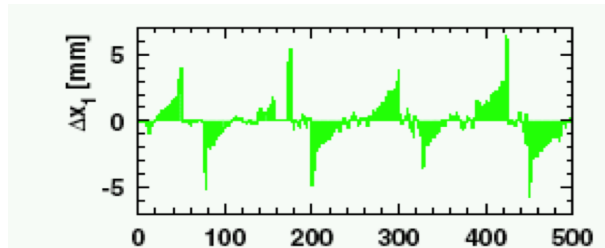


Fig. 32: Measured horizontal orbit of the LEP electron beam; due to synchrotron radiation losses, the particle orbit is shifted towards the inner side of the ring in each arc.

The strong dependence of the synchrotron radiation losses on the relativistic γ factor sets severe limits on the beam energy that can be carried in a storage ring of a given size. The push for ever higher energies means either that storage rings even larger than LEP need to be designed or, to avoid synchrotron radiation, linear accelerating structures should be developed.

Currently, the next generation of particle colliders is being studied [10]. The ring design of the future circular collider (FCC) foresees a 100 km ring to carry electrons (and positrons) of up to 175 GeV energy. The size of this storage ring is far beyond the dimensions of anything that has been designed up to now. A sketch of the machine layout is given in Fig. 33, where the yellow dashed circle delineates the 100 km ring and the white circle represents the little LHC machine.

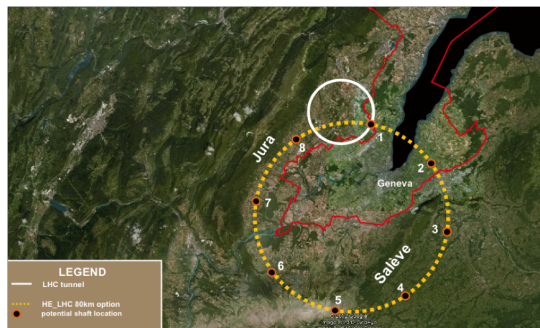


Fig. 33: Schematic view of a possible 100 km FCC design in the Geneva region

For the maximum projected electron energy of $E = 175$ GeV, synchrotron radiation would cause an energy loss of 8.6 GeV, or an overall power of 47 MW of the radiated light at full beam intensity.

4.2 Limit X: Acceleration gradients in linear structures

As far as lepton beams are concerned, ring colliders suffer from the severe limitation caused by synchrotron radiation losses, and at a certain point the construction of such large facilities would not seem reasonable any more. To avoid the problem of synchrotron radiation, linear structures that were discussed earlier and used in the infancy of particle accelerators have become in vogue again. Still, the advantage of circular colliders cannot be completely ignored: even with a modest acceleration gradient in the RF structures, the particles will get turn by turn a certain boost in energy and will at some point reach the desired flat-top energy in the ring.

In a linear accelerator, this kind of repetitive acceleration is by design not possible; within a single pass through the machine, the particles will have to be accelerated to full energy. In order to keep the structure compact, the highest acceleration gradients will therefore be needed. One of the most prominent designs proposed for a possible future collider is the CLIC design [11]. Within one passage through the 25 km long accelerator, the electron beam will get up to 3 TeV, and the same is true for the opposing positron beam. An artist's rendering of this machine is shown in Fig. 34.



Fig. 34: Proposed location of the CLIC linear collider along the Jura mountains in the Geneva region

The main parameters of the CLIC design are listed in Table 1. The accelerating gradient, i.e. the energy gain per meter, is especially to be emphasized; it has been pushed to the maximum value

that is technically feasible, and the limit is ultimately due to the breakdown of the electric field in the accelerating structure.

Table 1: Main parameters of the CLIC study

	500 GeV	3 TeV
Site length	13 km	48 km
Loaded acceleration gradient (MV m^{-1})	12	
Beam power per beam (MW)	4.9	14
Bunch charge (10^9 e+/e)	6.8	3.7
Horizontal/vertical normalized emittance ($10^{-6}/10^{-9}$ m)	2.4/25	0.66/20
Beta function (mm)	10/0.07	
Beam size at IP: horizontal/vertical (nm)	45/1	
Luminosity ($\text{cm}^{-2} \text{s}^{-1}$)	2.3×10^{34}	5.9×10^{34}

A picture of such a CLIC-type structure is shown in Fig. 35. On the right-hand side is an electron microscope photo of the surface after a voltage breakdown. At the spot of the sparking, a little crater can be seen, indicating possible damage to the surface and, as a consequence, deterioration of the achievable gradient which has to be avoided under all circumstances [12]. Although considerably higher than the typical values in circular machines, the gradient $E_{\text{acc}} = 100 \text{ MV m}^{-1}$ in a linear machine still leads to a design of overall length approximately 50 km for a maximum achievable energy of $E_{\text{max}} = 3 \text{ TeV}$.

5 Conclusion

To summarize, for future lepton ring colliders (or, to be more precise, electron–positron colliders), synchrotron radiation losses set a severe limit on the achievable beam energy; and very soon the size of the machines will become uneconomical. For a given limit in synchrotron radiation power, the dimensions of the machine would have to grow quadratically with the beam energy. Linear colliders are therefore proposed as the preferred way to go. In this case, the maximum achievable acceleration gradient is the key issue. New acceleration methods, namely plasma-based set-ups in which gradients have been observed that are much higher than those seen with conventional techniques, are a most promising concept for the design of future colliders. An impressive example is shown in Fig. 36: within a plasma cell of only a few centimetres in length, electrons are accelerated to several GeV. The gradients achievable are orders of magnitude higher than in any conventional machine (see, e.g., Ref. [13]). Still, there are problems to overcome, such as issues with overall efficiency, beam quality (mainly the energy spread of the beam), and the achievable repetition rate. Nevertheless, we are convinced that this is a promising field worthy of much further study—and this is what the present school is about.

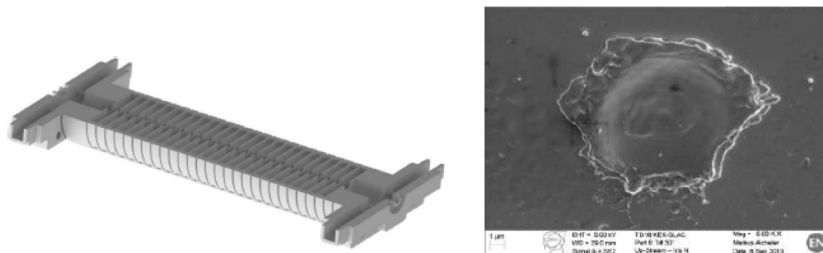


Fig. 35: Accelerating structure of the CLIC test facility CTF3; the electron microscope photo on the right shows the damage to the surface due to discharges in the module.

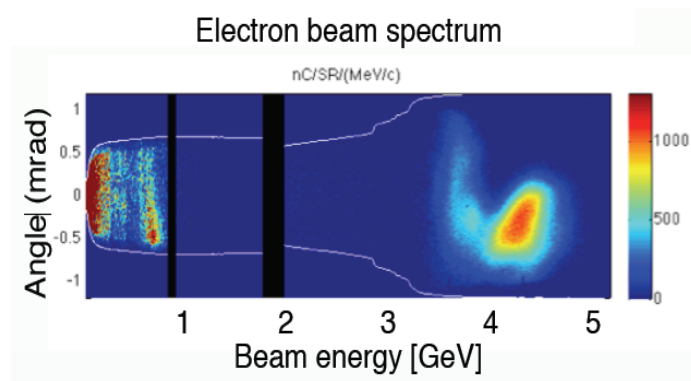


Fig. 36: Electron beam accelerated in the wake potential of a plasma cell; up to 4 GeV is obtained within only a few centimetres of length [13].

References

- [1] LHC design report, edited by O.S. Brüning et al., CERN-2004-003 (CERN, Geneva, 2004). <http://dx.doi.org/10.5170/CERN-2004-003-V-1>, <http://dx.doi.org/10.5170/CERN-2004-003-V-2>, <http://dx.doi.org/10.5170/CERN-2004-003-V-3>
- [2] C. Gerthsen, O. Kneser and H. Vogel, *Physik* (Springer, Berlin, 1982). <http://dx.doi.org/10.1007/978-3-662-09308-5>
- [3] J.D. Cockcroft and E.T.S. Walton, *Proc. R. Soc. (Lond.)* **A136** (1932) 619, **A137** (1932) 229, **A144** (1934) 704.
- [4] K. Wille, *The Physics of Particle Accelerators* (Oxford University Press, Oxford, 2000).
- [5] R.J. Van de Graaff, *Phys. Rev.* **38** (1931) 1919.
- [6] P.J. Bryant, in Proceedings of the CAS-CERN Accelerator School: 5th General Accelerator Physics Course, edited by Turner S., CERN-1994-001 (CERN, Geneva, 1994). <http://dx.doi.org/10.5170/CERN-1994-001.1>
- [7] HERA Design Team, HERA: a proposal for a large electron proton colliding beam facility at DESY, DESY-HERA-81/10 (1981).
- [8] J. Rossbach and P. Schmueser, in Proceedings of the CAS-CERN Accelerator School: 5th General Accelerator Physics Course, edited by Turner S., CERN-1994-001 (CERN, Geneva, 1994). <http://dx.doi.org/10.5170/CERN-1994-001.17>
- [9] W. Herr, in Proceedings of the CAS-CERN Accelerator School: Intermediate Course on Accelerator Physics, edited by Brandt D., CERN-2006-002 (CERN, Geneva, 2006). <http://dx.doi.org/10.5170/CERN-2006-002.379>
- [10] F. Zimmermann, Challenges for highest energy circular colliders, in Proc. 5th International Particle Accelerator Conf. (IPAC 2014, Dresden), in press.
- [11] A multi-TeV linear collider based on CLIC technology: CLIC conceptual design report, edited by M. Aicheler et al., CERN-2012-007 (CERN, Geneva, 2012). <http://dx.doi.org/10.5170/CERN-2012-007>
- [12] A. Palaia, Beam momentum changes due to discharge in high-gradient accelerator structures, Ph.D. thesis, Uppsala University, 2013; ISBN 978-91-554-8802-4.
- [13] W.P. Leemans et al., *Phys. Rev. Lett.* **113** (2014) 245002. <http://dx.doi.org/10.1103/PhysRevLett.113.245002>
- [14] A. Cho, *Science* **337** (2012) 141. <http://dx.doi.org/10.1126/science.337.6091.141>