Version 0.3

Ezio.Todesco@cern.ch

Chapter 1

Magnet requirements in circular accelerators: the arc, beam energy and beam size

Plan of the chapter
In this chapter we discuss the principles of a synchrotron (section 1.1). After a short recall of relativistic mechanics (section 1.2), we introduce in section 1.3 the equations that relate the momentum, the curvature radius and the energy in a circular storage ring. The linear stability of the motion is treated in section 1.4, where we show how quadrupoles stabilize the particles around the reference orbit. In section 1.5 the definition of the optical functions and the requirements on the magnet aperture related to the transverse beam size are given. The quadrupole strength and the beta functions in an alternating gradient lattice (FODO cell) are given in section 1.6. We summarize in 1.7 the flowchart that determines the magnet requirements for dipoles and quadrupoles in the arc.

1.1 Principles of a synchrotron

Charged particles interact with electromagnetic fields through the electromagnetic force, also known as Lorentz force

$$\vec{F} = e\vec{E} + e\vec{v} \times \vec{B} \,. \tag{1.1}$$

In this text we will use international system (IS) units, i.e. meters, kilograms, seconds and ampere unless explicitly stated (as for instance GeV for particle energy in some equations). In a circular particle accelerator, a section is devoted to the accelerating part, where the electric field is used to increase the beam momentum. The rest of the accelerator is used to steer the particle beam back to the entrance of the accelerating structure. The bending is given by the magnetic field B, and under some approximations discussed in the next section, the relation between the curvature radius ρ, the magnetic field B, and the particle momentum p (see Fig. 1.1) is

$$p = eB\rho \,. \tag{1.2}$$

During the acceleration, the increase of the momentum p is "synchronized" with the increase of the magnetic field B so that the particle trajectories in the dipoles have a fixed curvature radius ρ. This gives the name "synchrotron" to this type of accelerator. For comparison, a "cyclotron" is a particle accelerator where the magnetic field is kept constant, and the curvature radius increases with the energy, producing particle trajectories that spiral out of the magnet (see [1], first chapter).

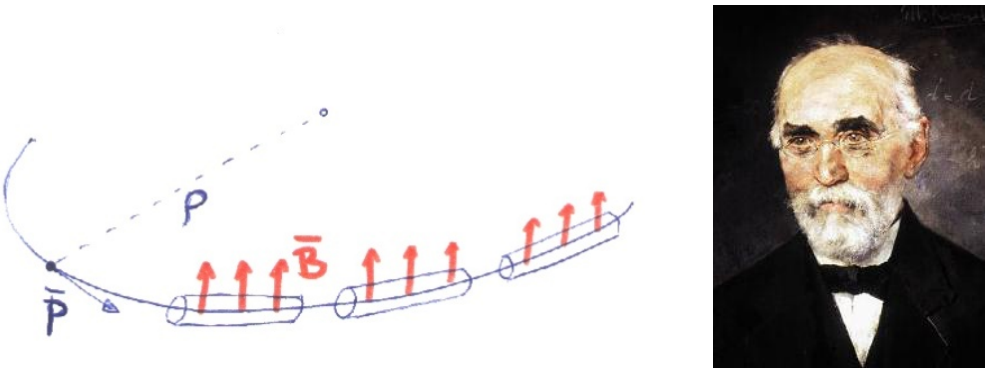


Fig. 1.1: Curvature radius, momentum and magnetic field (left); portrait of Hendrik Antoon Lorentz (18 July 1853 - 4 February 1928), painted by Menso Kamerlingh Onnes, brother Heike Kamerlingh Onnes, who discovered superconductivity in 1911 (right).

In a synchrotron, the energy of the particles increases at each turn, and the limit to the energy increase is the maximum magnetic field provided by the magnets. Therefore, the synchrotron contains two paradoxes:

- The accelerating section occupies a minor part of the accelerator, both in terms of space and cost; most of the machine is devoted not to increase the energy, but to bend the particle trajectories.
- Even though the magnetic field does not increase the particle energy, the maximum energy is fixed by the maximum magnetic field.

This is why magnets are so important in a circular particle accelerator, both in terms of performance, space and cost. The dipoles occupy most of the arcs, where the particle trajectories are bent. Long straight sections (LSS) are needed to have a region free from accelerator magnets to house the detectors (see Fig. 1.2). They are also needed to house the accelerating system, to inject and dump the beam, and to have special regions to remove from the beam the particles that are diffusing far from the reference orbit.

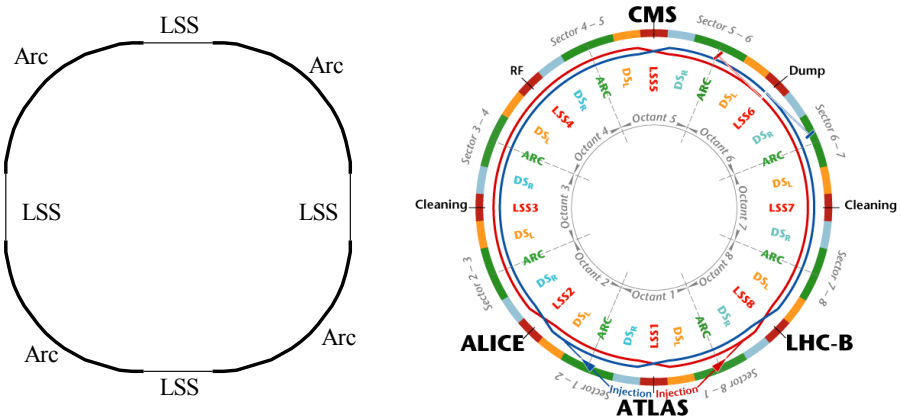


Fig. 1.2: Schematic of a synchrotron with arcs and long straight sections (LSS), left; layout out of the LHC (right) with four experiments CMS, LHC-B, ATLAS and ALICE, cleaning insertions, dump, and accelerating section (RF)

The accelerator can be seen as a "concentrator of energy", i.e. takes the electric energy from the grid and, through the magnet and the accelerating systems, concentrates it in a small quantity of charged particles. Typically an accelerator beam has ~10^{13} particles, i.e. a nanogram of matter. The concentration process is obviously far from been effective (a rough estimate for the LHC, given in

chapter 2, gives a factor 1000, i.e. if 1000 J are taken from the power grid, only 1 J is transferred to the protons). Since a lot of energy is put on a few particles, the total energy stored in the beam is relatively small (this is also why an accelerator cannot "explode").

1.2 Basics of relativistic mechanics

The kinematics of circular motion states that the acceleration is proportional to the square of the velocity divided by the curvature radius ρ (see Fig. 1.1) – this equation is the same both in the classical and in the relativistic mechanics

$$\left|\frac{d\vec{v}}{dt}\right| = \frac{v^2}{\rho}; \tag{1.3}$$

according to classical mechanics the momentum is

$$\vec{p} \equiv m\vec{v}, \tag{1.4}$$

and the Newton law is

$$F = \frac{d\vec{p}}{dt}. \tag{1.5}$$

In relativistic mechanics the momentum has the additional term γ_r (see for instance [2], chapters 11 and 12 for a general description of special relativity, and [1], chapter 1 for a short summary of what needed in our case)

$$\gamma_r \equiv \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \qquad \vec{p} \equiv m\gamma_r\vec{v} \tag{1.6}$$

which diverges for velocities approaching the speed of light:

$$\gamma_r(v) \xrightarrow[v\to c]{} \infty \qquad \vec{p} \xrightarrow[v\to c]{} \infty. \tag{1.7}$$

The dynamics is still given by the Newton law with this new definition of momentum

$$\vec{F} = \frac{d\vec{p}}{dt} = m\frac{d(\gamma_r\vec{v})}{dt}. \tag{1.8}$$

Please note that the mass m is a scalar invariant, related to the norm of the quadrivector $(E/c,p)$

$$\frac{E^2}{c^2} - p^2 = m^2c^2. \tag{1.9}$$

The above equation, giving the preservation of the norm of the quadrivector, is equivalent to defining the energy as

$$E = m\gamma_r c^2. \tag{1.10}$$

In fact from (1.9) one can derive (1.10) according to

$$E = c\sqrt{p^2+m^2c^2} = c\sqrt{m^2{\gamma_r}^2v^2+m^2c^2} = mc^2\sqrt{{\gamma_r}^2\frac{v^2}{c^2}+1} =$$
$$= mc^2\sqrt{\frac{v^2}{c^2-v^2}+1} = \sqrt{\frac{c^2}{c^2-v^2}}mc^2 = \sqrt{\frac{1}{1-\frac{v^2}{c^2}}}mc^2 = m\gamma_r c^2. \tag{1.11}$$

This gives the well-known definition of rest energy at v=0, associated to each body for the fact of having a mass

$$E(v=0) \equiv E_0 = mc^2. \tag{1.12}$$

For small velocities one can expand the square root finding the classical definition of kinetic energy

$$E = \sqrt{m^2c^4 + p^2c^2} = mc^2\sqrt{1+\frac{p^2c^2}{m^2c^4}} \sim mc^2\left(1+\frac{p^2c^2}{2m^2c^4}\right) = mc^2 + \frac{p^2}{2m} = E_0 + \frac{p^2}{2m} \tag{1.13}$$

where we used the Taylor expansion, valid for small a

$$\sqrt{1+a} \sim 1+\frac{a}{2}. \tag{1.14}$$

In the ultrarelativistic regime the rest energy (given by the mass) is negligible with respect to the kinetic energy and (1.9) becomes

$$E = \sqrt{m^2c^4 + p^2c^2} \sim \sqrt{p^2c^2} = pc \qquad p >> mc\,. \tag{1.15}$$

1.3 How to keep charged particles on a circular orbit: relation between field, energy and accelerator size

We now derive the relation (1.2) between momentum, magnetic field and curvature radius (see also [1], chapter 2): in absence of electric field the electromagnetic force (1.1) becomes

$$\vec{F} = e\vec{v}\times\vec{B}\,. \tag{1.16}$$

We assume that that the longitudinal acceleration (increasing the modulus of velocity, i.e., increasing the energy of the particle) is much smaller that the transverse acceleration (given by the change of direction of the velocity):

$$m\gamma_r\left|\frac{d\vec{v}}{dt}\right| >> m\frac{d\gamma_r}{dt}\left|\vec{v}\right|; \tag{1.17}$$

therefore using the relativistic mechanics (1.3), (1.5) and (1.6) one has

$$\left|\vec{F}\right| = \left|\frac{d\vec{p}}{dt}\right| = m\left|\frac{d(\gamma_r\vec{v})}{dt}\right| \approx m\gamma_r\left|\frac{d\vec{v}}{dt}\right| = m\gamma_r\frac{v^2}{\rho} \tag{1.18}$$

and using (1.16) one obtains

$$\left|\vec{F}\right| = \left|e\vec{v}\times\vec{B}\right| = evB = m\gamma_r\frac{v^2}{\rho} \qquad eB = m\gamma_r\frac{v}{\rho} = \frac{p}{\rho}. \tag{1.19}$$

The last term is the famous equation (1.2) for the magnetic field B needed to keep a particle of momentum p on a curvature radius ρ, given at the beginning of the chapter:

$$p = eB\rho\,. \tag{1.20}$$

Note that in the ultra-relativistic regime (1.15) one has

$$E = pc = ecB\rho\,. \tag{1.21}$$

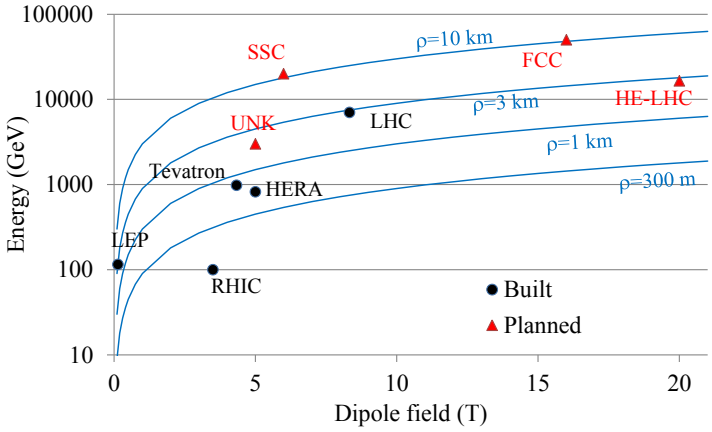
In particle physics, the unit of energy commonly used is the elettronvolt eV, related to the Joule through the electron charge e=1.602×10^{-19} C:

$$eE[\mathrm{eV}] = E[\mathrm{J}] \qquad 1\,\mathrm{eV} = 1.602\times10^{-19}\ \mathrm{J}. \tag{1.22}$$

Therefore the above relation (1.20) can be rewritten in the handy format, with the energy in GeV, the field expressed in Tesla and the curvature radius in meters, and the speed of light is rounded at 3.00×10^8 m/s

$$E[GeV] = \frac{10^{-9}}{e}ecB\rho = 10^{-9}\times3\times10^8\times B\rho = 0.3\times B\rho \quad . \tag{1.23}$$

The curvatures, energies and magnetic fields of a few accelerators, planned or built, are given in Fig. 1.3, with a logarithmic scale for the energies (see [3], chapter 1.6.3, for a parameter list of particle accelerators).



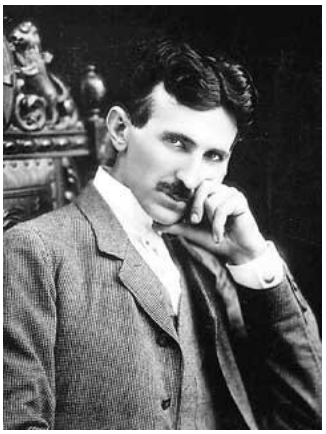


Fig. 1.3: Left: Curvature radius, energy and magnetic field in particle accelerators, built or planned, in semilog scale. Right: Nikola Tesla (10 July 1856 – 7 January 1943).

Example 1.1: Find the relation between γ_r and the energy in GeV for protons and electrons.
For a proton one has

$$E = m_p \gamma_r c^2 \qquad \gamma_r = \frac{E}{m_p c^2} = \frac{10^9 \times 1.602 \times 10^{-19}}{1.672 \times 10^{-27} \times (3 \times 10^8)^2} E[\text{GeV}] = 1.064 \times E[\text{GeV}]. \tag{1.24}$$

Therefore for a proton γ_r is 6.4% larger than the energy in GeV. For instance, a proton at 7 TeV has γ_r=1.064×7000=7452.
For an electron one has

$$\gamma_r = \frac{E}{m_e c^2} = \frac{10^9 \times 1.602 \times 10^{-19}}{9.109 \times 10^{-31} \times (3 \times 10^8)^2} E[\text{GeV}] = 1954 \times E[\text{GeV}]. \tag{1.25}$$

Therefore for an electron γ_r is approximately 2000 times the energy in GeV. For instance, an electron at 100 GeV has γ_r=1954×115=225000 (and therefore is faster than a proton at 7 TeV).

Note that in several texts the mass is called "rest mass", and denoted by m_0, and a "relativistic mass" m is defined

$$m \equiv m_0 \gamma_r . \tag{1.26}$$

This allows to write both the momentum definition and the Newton law as in the classical form

$$\vec{p} \equiv m\vec{v} \tag{1.27}$$

$$F = \frac{d\vec{p}}{dt} . \tag{1.28}$$

According to this interpretation, an accelerator can be considered as a tool to "increase the mass of particles". This statement is controversial, and the concept itself of "relativistic mass" was not proposed by Einstein. Actually, Einstein was far from being in favour of this idea [4]:

"It is not good to introduce the concept of the mass of a moving body for which no clear definition can be given. It is better to introduce no other mass concept than the 'rest mass' m. Instead of introducing M it is better to mention the expression for the momentum and energy of a body in motion."

and the owerwhelming majority of theoretical physicists oppose to this concept, see for instance [5]:
"The concept of relativistic mass is subject to misunderstanding. That's why we don't use it. First, it applies the name mass - belonging to the magnitude of a 4-vector - to a very different concept, the time component of a 4-vector. Second, it makes increase of energy of an object with velocity or momentum appear to be connected with some change in internal structure of the object. In reality, the

increase of energy with velocity originates not in the object but in the geometric properties of spacetime itself."

How much we manage to approach the speed of light in particle accelerators? In Fig. 1.4 we give an overview plot. The gamma factor is directly linked to the velocity, but the energy is proportional to the gamma factor times the particle mass (see Eq. 1.10). This is why the speed record does not belong to the LHC, accelerating protons to 7 TeV, but to LEP, accelerating electrons at 120 GeV. The speed of electrons in LEP differs from the speed of light only in its 11th digit.

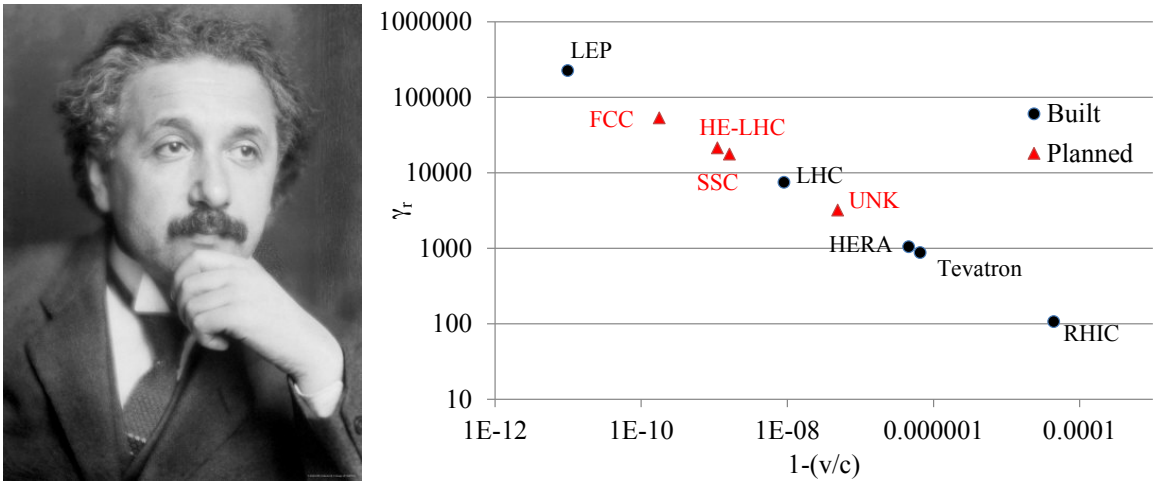


Fig. 1.4: Albert Einstein (14 March 1879 – 18 April 1955), left; relative difference to the speed of light and relativistic factor for some particle accelerators, planned or built.

Example 1.2: LHC has a nominal energy of 7 TeV and a nominal field of 8.33 T. Knowing that the magnetic length of the dipoles is L_d=14.3 m, how many dipoles are in the accelerator? What is curvature radius of the dipoles? Knowing that the total length of the tunnel is approximately L_T=27 km, what is the fraction of the tunnel occupied by dipoles?

Using the Eq. (1.23) we obtain the curvature radius

$$\rho = \frac{E[GeV]}{0.3 \times B} = \frac{7000}{0.3 \times 8.3} = 2801 \text{ m}. \tag{1.29}$$

The number of the dipoles is approximately

$$N_d = \frac{2\pi\rho}{L_d} = \frac{2 \times 3.14 \times 2801}{14.3} = 1230 \tag{1.30}$$

(note that the actual number is 1232), and the fraction of the accelerator covered by the dipoles is

$$f = \frac{N_d L_d}{L_T} = \frac{1230 \times 14.3}{27000} = 65\%. \tag{1.31}$$

Example 1.3: The FCC (Future Circular Collider) study proposed in the mid '10s a 50 TeV hadron collider, with a dipole nominal field of 16 T. Knowing that the total length of the tunnel is approximately L_T=100 km, what is the fraction of the tunnel occupied by dipoles? What is the gamma factor and the speed of the protons? Is it larger than the electrons in LEP?

The curvature radius is

$$\rho = \frac{E[GeV]}{0.3 \times B} = \frac{50000}{0.3 \times 16} = 10400 \text{ m}. \tag{1.32}$$

and

$$f = \frac{2\pi\rho}{L_T} = \frac{2\times 3.14\times 10400}{100000} = 65\%. \tag{1.33}$$

The gamma factor is given by

$$\gamma_r = \frac{E}{mc^2} = \frac{50\times 10^{12}\times 1.608\times 10^{-19}}{1.672\times 10^{-27}\times\left(3\times 10^8\right)^2} = 53200. \tag{1.34}$$

The proton speed will be close to the speed of light by less than one part per billionth

$$\frac{v}{c} = \sqrt{1-\frac{1}{{\gamma_r}^2}} \approx 1-\frac{1}{2{\gamma_r}^2} = 1-\frac{1}{2\times(53200)^2} = 1-\frac{1}{5.66\times 10^9} = 1-1.77\times 10^{-10} \tag{1.35}$$

so the speed is still slower than in LEP electrons, whose gamma factor is ~200000, see example 1.1.

Example 1.4: The kinetic energy of a fly is 1 TeV, and its mass is 1 g. What is its speed?
Using the Newton equations (whose validity is verified a posteriori) we compute

$$v = \sqrt{\frac{2E}{m}} = \sqrt{\frac{2\times 10^{12}\times 1.6\times 10^{-19}}{0.001}} = 0.018. \tag{1.36}$$

i.e. ~2 cm/s, which does not require the relativistic equations.

1.4 Linear stability and quadrupoles

We now consider a particle on a circular orbit matching the condition (1.2) between energy curvature, radius and momentum. We define the following coordinate system: s is the coordinate along the orbit, and (x,y) is the transverse plane, where x is on the plane of the orbit (horizontal axis, see Fig. 1.5). We also define a quadrupole as a magnetic field in the transverse plane, proportional to the distance to the center of the reference orbit, over the quadrupole length l_q

$$\begin{cases} B_y(s) = Gx \\ B_x(s) = Gy \\ B_s(s) = 0 \end{cases} \qquad 0 < s < l_q \tag{1.37}$$

where G, expressed in T/m, is the quadrupole gradient.

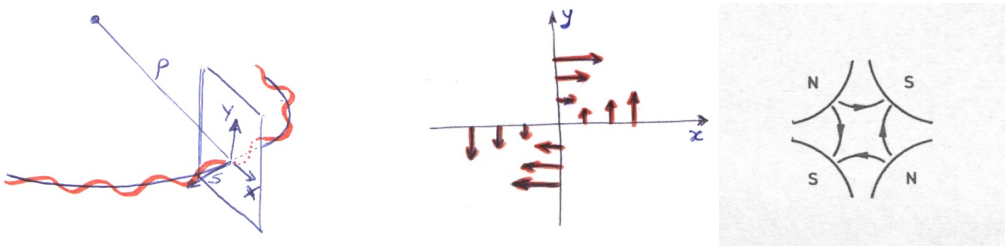


Fig. 1.5: Left: the particle oscillating around the reference orbit, and the transverse coordinate system; center and right: field lines of a quadrupole

A quadrupole provides a spring-like force, which ensures that particles deviating from the reference orbit are brought back to it (see Fig. 1.5). In fact the Lorentz force reads

$$F_x = e\left[\vec{v}\times\vec{B}\right]_x = e\left(\frac{dy}{dt}B_s - \frac{ds}{dt}B_y\right) = -e\frac{ds}{dt}B_y \ . \tag{1.38}$$

Therefore the force in the x plane is proportional to x:

$$F_x = -e\frac{ds}{dt}B_y = -e\frac{ds}{dt}Gx \tag{1.39}$$

and the motion becomes "like" the harmonic oscillator (we will clarify how much it is similar to an harmonic oscillator in the coming lines). In the *y* plane we have

$$F_y = e\left[\vec{v}\times\vec{B}\right]_y = e\left(\frac{ds}{dt}B_x - \frac{dx}{dt}B_s\right) = e\frac{ds}{dt}Gy \tag{1.40}$$

Note that the sign is different, so a quadrupole gives linear stability in one plane and linear instability in the other one. For $G>0$ (what is called a "focusing quadrupole") motion is stable in *x* and unstable in *y*. For the defocusing case $G<0$, and the motion is unstable in *x* and stable in *y*.

We now consider the equation for the *x* coordinate, replacing the derivative with respect to time with the derivatives with respect to the *s* coordinate:

$$F_x = m\frac{d}{dt}\gamma_r v_x \sim m\gamma_r\frac{d^2x}{dt^2} = m\gamma_r\frac{d^2x}{ds^2}\left(\frac{ds}{dt}\right)^2 \tag{1.41}$$

Putting together the two equations for the force in the *x* coordinte (1.39) and (1.41) together we obtain

$$m\gamma_r\frac{d^2x}{ds^2}\left(\frac{ds}{dt}\right)^2 = -e\frac{ds}{dt}Gx \tag{1.42}$$

and therefore we have

$$\frac{d^2x}{ds^2} + \frac{eG}{m\gamma_r\dfrac{ds}{dt}}x = 0 \tag{1.43}$$

and since the momentum is mainly carried by the component along the orbit

$$p = m\gamma_r\sqrt{\left(\frac{ds}{dt}\right)^2 + \left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} \sim m\gamma_r\frac{ds}{dt} \qquad \frac{d^2x}{ds^2} + \frac{eG}{p}x = 0. \tag{1.44}$$

Using the relation between curvature radius (1.2), dipole field and momentum one obtains

$$\frac{d^2x}{ds^2} + \frac{G}{B\rho}x = 0. \tag{1.45}$$

In the vertical coordinate *y* one has

$$\frac{d^2y}{ds^2} - \frac{G}{B\rho}y = 0. \tag{1.46}$$

A constant quadrupole field along the orbit $G(s)=G_0$ cannot provide stable motion in both *x* and *y*; on the other hand, if the quadrupolar field varies along the orbit, one can find cases where the linear stability is both in *x* and *y*. The linear equations therefore read

$$\begin{cases} \dfrac{d^2x}{ds^2} + K_1(s)x = 0 \\ \dfrac{d^2y}{ds^2} - K_1(s)y = 0 \end{cases} \tag{1.47}$$

with K_1 is defined as

$$K_1(s) \equiv \frac{G(s)}{B\rho}. \tag{1.48}$$

Please note that K is called the gradient in the jargon of accelerator physicists, and that G is called the gradient in the jargon of magnet engineers. The dimension of K_1 is [1/m^2], whereas the dimension of G is [T/m], so be sure to understand which gradient are you dealing with, depending on who are you talking to. Here we will follow the magnet engineer convention, so the gradient is G and we use no name for K_1.

Equations (1.47) are called Hill equations; they were written well before the construction of synchrotrons by George William Hill, astronomer, at the end of the 19th century (see Fig. 1.5). Originally developed for celestial mechanics, they apply to several other fields, as atomic physics and accelerator physics.

1.5 Transverse beam size, beta functions and magnet aperture

Hill equations are more complicated than the equation of the harmonic oscillator

$$\frac{d^2x}{ds^2} + \omega^2 x = 0 \tag{1.49}$$

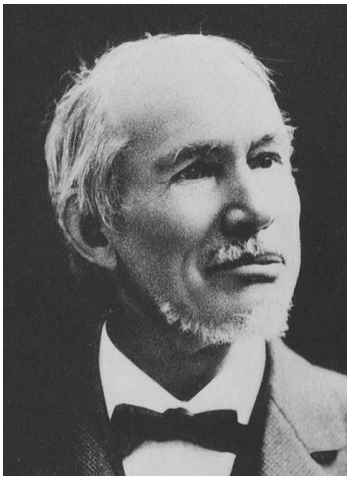
since the frequency $\omega=\sqrt{K_1}$ depends on the variable s. This reflects the fact that the accelerator is made by a sequence of different magnets, and therefore the gradient is not constant along the whole ring: infact $K_1(s)$ is a discontinuous function, zero in the dipoles and $G/B\rho$ in the quadrupoles. The other difference is that in Hill equations applied to accelerators, the time variable is replaced by the longitudinal coordinate s, which is periodic.

The equation for the harmonic oscillator has the well-known solution

$$x(t) = A\cos(\omega t + \omega_0)\,. \tag{1.50}$$

Hill equations, representing the motion in the transverse plane of the particles in a circular accelerator, has a solution similar to the harmonic oscillator, but where both the amplitude and the phase are s-dependent (see Fig. 1.6, right), and both functions are periodic in s over the accelerator length l_T.

$$x(s) = A(s)\cos(\psi(s) + \psi_0) \qquad \psi(s + l_T) = \psi(s) \qquad A(s + l_T) = A(s)\,. \tag{1.51}$$



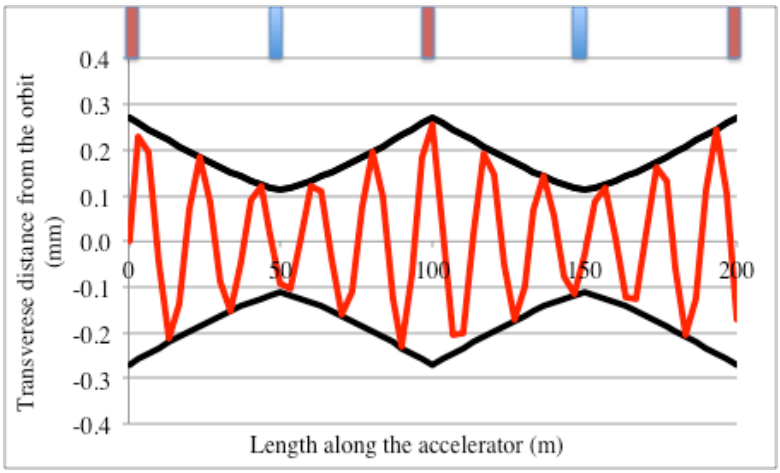


Fig. 1.6: Left: George William Hill, American (3 March 1838 – 16 April 1914). Right: a solution of the Hills equation (in red), with amplitude (in black) and period of the oscillations varying along time.

The usual convention is to express the amplitude as the square root of a function dependent on s, which is called the beta function

$$x(s) = \sqrt{a\beta(s)}\cos(\psi(s) + \psi_0)\,. \tag{1.52}$$

The advantage of using (1.52) with respect to (1.51), is that the function ψ (called the phase advance) has a simple relation to the beta function: it is the integral of the inverse of the beta function

$$\psi(s) = \int_0^s \frac{1}{\beta(t)} dt \,. \tag{1.53}$$

This equation is proved in Appendix A, where we discuss in more detail the mathematical methods to solve Hills equations, and we give explicit solutions for a sequence of quadrupoles.

The quantity a in (1.52) is invariant along the ring, and one can prove that it is inverse proportional to the energy of the particle

$$x(s) = \sqrt{\frac{\varepsilon_n \beta(s)}{\gamma_r}} \cos(\psi(s) + \psi_0) \,. \tag{1.54}$$

So that during the acceleration the invariant ε_n called normalized emittance, is preserved (for a perfect chain of accelerators). For an ensemble of particles, the emittance is usually defined as the standard deviation of the distribution of the emittances of each particle. The above equation is very relevant, since it gives the beam size variation along the accelerator. The requirements on the magnet aperture ϕ are related to the beam size:

$$\phi = 2n\left|x(s)\right| + \phi_0 = 2n\sqrt{\frac{\varepsilon_n \beta(s)}{\gamma_r}} + \phi_0 \tag{1.55}$$

where the factor 2 is given by the convention of indicating magnet apertures as diameters, and not as radius, and n is the number of sigma taken for the beam (typically 15 sigma are taken). There is an additional margin ϕ_0 required by the tolerances and by two components needed inside the magnet. The first one is the bore separating the coil, immersed in the cryogenic fluid, from the beam aperture in ultra-high vacuum, usually called cold bore. The second one is the beam screen used to maintain the ultra-high vacuum in the arc and to absorb a good fraction of the particles, thus reducing the heat deposition in the magnets.

Summarizing, the aperture requirement for a magnet in the arcs (dipole or quadrupole) related to the beam size depends on three terms:

- The aperture is proportional to the square root of the inverse of the beam energy: larger energies will give smaller beam size and therefore require smaller magnets; low energy machines will have wider aperture magnets;
- The aperture is proportional to the square root of the normalized emittance, which is a property of the injected beam and (in the linear case) is preserved along the ring and through acceleration;
- The aperture is proportional to the square root of the beta function.

1.6 The FODO lattice

We now discuss the beta functions for an accelerator based on the alternating gradient magnetic lattice. One can prove that a sequence of quadrupoles, equally spaced by a distance L, which alternate opposite gradients $\pm G$, gives a motion that is linearly stable in the two transverse coordinates.This sequence is also called FODO cell, or alternating gradient optics [6,7,8]. The length of the cell, i.e. its periodicity, is $2L$. The case is fully discussed in the Appendix A, and here we give the results that are relevant for the magnet specifications.

Since the particle motion is the product of two uncoupled harmonic oscillators, one in x and one in y (whose frequency and amplitude depend on s), one can define the number of oscillations made by the particle after one turn around the accelerator, in horizontal and vertical coordinates

$$\begin{aligned} Q_x &= \psi_x(l_T) \\ Q_y &= \psi_y(l_T) \end{aligned} \tag{1.56}$$

as the linear tunes of the accelerator. The number of oscillations over one cell

$$\begin{aligned} \psi_{xL} &= \psi_x(s+2L) - \psi_x(s) \\ \psi_{yL} &= \psi_y(s+2L) - \psi_y(s) \end{aligned} \tag{1.57}$$

is called the cell phase advance (usually the same in both planes).Varying the quadrupole gradient and the quadrupole spacing one can build cells with different phase advances.

One can prove that the beta functions between the quadrupoles are a second order polynomial in the longitudinal variable s and that peaks are reached in the successive quadrupoles (see Fig. 1.7).

In the case of thin lens $l_q \ll L$ (quadrupole with small length with respect to the quadrupole spacing), one can prove that the phase advance per cell is related to the quadrupole spacing L and to the focusing strength of the quadrupole through

$$\cos\psi_L = \frac{2 - L^2 f^{-2}}{2} \qquad f = \frac{L}{\sqrt{2 - 2\cos\psi_L}} \qquad Gl_q = \frac{B\rho}{f} = B\rho\frac{\sqrt{2 - 2\cos\psi_L}}{L} \tag{1.58}$$

and the beta functions in the focusing and defocusing quadrupole are

$$\beta_{F,D} = \frac{2L \pm L^2 f^{-1}}{\sin\psi}. \tag{1.59}$$

In general the sequence of magnets aims at having a cell phase advance close to a fraction of the integer, i.e. the oscillations are much slower than what shown in Fig. 1.6, right. In a 90° cell phase advance the particle makes one full oscillation over four cells. This is one of the most optimized solutions in terms of gradient and aperture. Since $\psi_L = \pi/4$, the integrated gradient is

$$f = \frac{L}{\sqrt{2}} \qquad Gl_q = \frac{\sqrt{2}B\rho}{L} \tag{1.60}$$

and the beta functions in the quadrupole oscillate between the two peaks

$$\beta_F = (2+\sqrt{2})L \qquad \beta_D = (2-\sqrt{2})L. \tag{1.61}$$

The spacing of the quadrupoles L is the free parameter of the lattice. A longer spacing allows to have less quadrupoles, leaving more space to the dipoles to bend the particles (i.e. giving more energy for the same tunnel length), and requiring less strength in the quadrupoles. But it produces larger beta functions, i.e. requires larger apertures in the magnets (see Fig. 1.7). Typically the quadrupole spacing is of the order of 10-100 m, depending on the accelerator size; longer accelerators have longer cells.

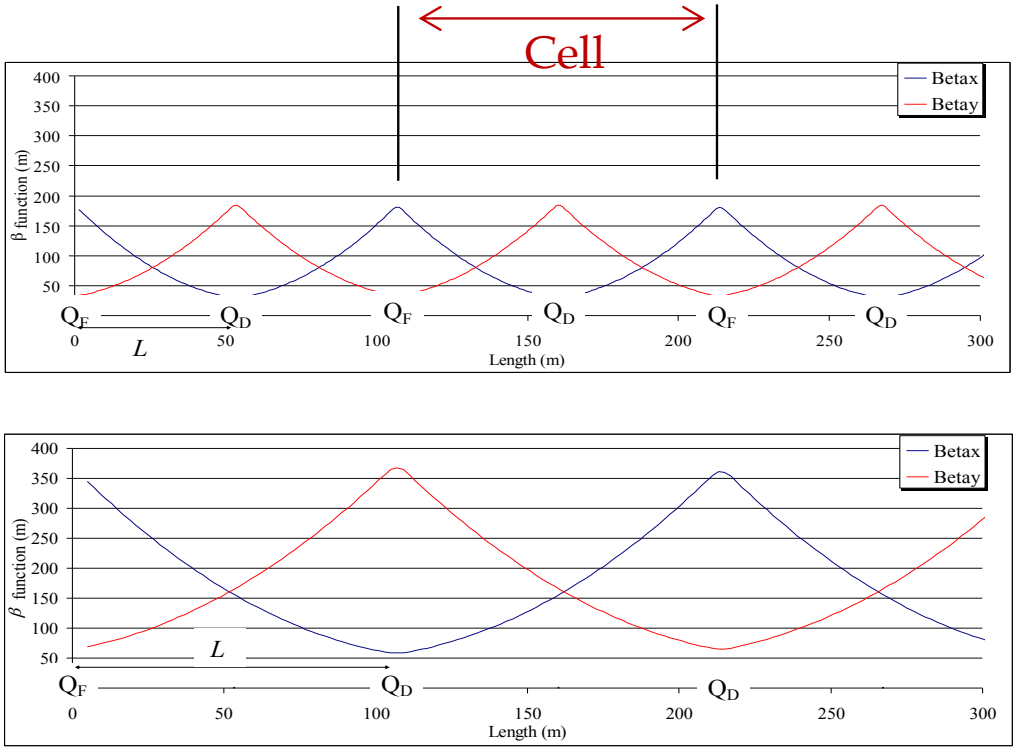


Fig. 1.7: Beta functions in an alternating gradient lattice with 50 m (upper part) and 100 m quadrupole spacing (lower part).

1.7 A flowchart for the magnet specifications in the arc

We summarize here a possible flowchart to determine the parameters of the arc magnets in a circular accelerator. We refer to Fig. 1.8, where the input parameters are shown in blue (collision and injection energy), the constraints from the technologies are in red (maximum field, related to the magnet technology, and beam emittance given by the injectors).

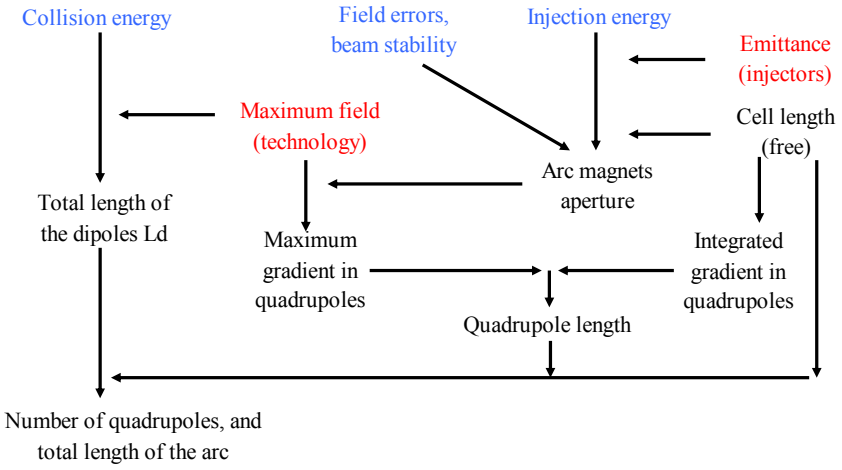


Fig. 1.8: Flowchart to determine the magnet specifications in the arc: constraints from the technologies in red, input parameters in blue, outcome in black.

Step 1: The collision energy E and the maximum operational field B fix the total length of the dipoles L_{Td} according to (1.23)

$$E[GeV] = 0.3B\left[T\right]\rho\left[m\right] \tag{1.62}$$

and therefore

$$L_{Td} = 2\pi\rho = \frac{2\pi E[GeV]}{0.3\times B}. \tag{1.63}$$

Step 2: the injection energy E_i and the normalized beam emittance ε_n are used to compute the magnet aperture, using the additional input of the cell length L. The injection energy gives the gamma factor at injection (see 1.9)

$$\gamma_{r,i} = \frac{E_i}{mc^2} \tag{1.64}$$

(remember to use the energy in J and the mass of the accelerated particle in kg, or the rules given in section 1.2). The cell length is a free parameter of the accelerator design and determines the maximal beta function; if the cell has a 90° phase advance one has

$$\beta_F = (2+\sqrt{2})L; \tag{1.65}$$

finally one gets the aperture through Eq. (1.55), where n is the number of sigma of the beam (typically 15 to 20); for the offset ϕ_0 one has to make an educated guess

$$\phi = 2n\left|x(s)\right| + \phi_0 = 2n\sqrt{\frac{\varepsilon(2+\sqrt{2})L}{\gamma_{r,i}}} + \phi_0. \tag{1.66}$$

Step 3: we can now determine the quadrupoles parameters required to provide the linear stability. Having the quadrupole aperture ϕ, and the maximum field B, the gradient is limited by the field at the edge of the aperture

$$\frac{G\varphi}{2\lambda} \le B. \tag{1.67}$$

Due to magnet design features, which will be discussed in Chapter 5, the above equation requires an additional coefficient λ describing the overshooting of the field in the coil, which for quadrupole can be of the order of 1.15. Therefore the maximum gradient is given by

$$G = \frac{2B}{\lambda\varphi}. \tag{1.68}$$

The integrated gradient for a 90° phase advance cell is given by (1.48):

$$Gl_q = \frac{\sqrt{2}B\rho}{L} = \frac{\sqrt{2}}{0.3L}E\left[GeV\right] \tag{1.69}$$

and therefore the length of the quadrupoles is

$$l_q = \frac{\sqrt{2}}{0.3GL}E\left[GeV\right]. \tag{1.70}$$

Step 4: We can compute now all the elements of the arc. Assuming that the length of the dipoles in the half cell is l_d, and their number is n_d, we have

$$L = l_q + n_d l_d + l_{i,c} \tag{1.71}$$

where the residual length $l_{i,c}$ is needed for interconnections and correctors, and the filling factor, i.e. the fraction of the arc covered by the dipoles, is

$$\chi = \frac{L - l_q - l_{i,c}}{L} = 1 - \frac{l_q + l_{i,c}}{L}. \tag{1.72}$$

This number is typically between 0.8 and 0.85 for high energy circular collider, so that most of the arc is dedicated to bend particles and only a small fraction is required for the beam focusing. In the interaction regions the optics follows totally different rules and requirements: this will be the subject of chapter 2.

We now give some examples to clarify the procedure and the application of the equations.

Example 1.5: LHC has a quadrupole spacing of L=50 m. Knowing that the arc quadrupoles have a gradient of 220 T/m, compute the quadrupole length. The emittance is 3.75×10^{-6} m mrad (one sigma of the distribution). Assuming that the magnet aperture should be able to accommodate 15 sigma beam plus 50% margin at the injection energy of 450 GeV, estimate the aperture requirement.

The integrated gradient is given by (1.60), where the magnetic field is 8.33 T and the curvature radius 2801 m (see also Example 1.2)

$$Gl_q = \frac{\sqrt{2}B\rho}{L} = \frac{\sqrt{2}\times 8.33\times 2801}{50} = 660\,\mathrm{T}, \tag{1.73}$$

so the quadrupole is 3 m long. This means that it occupies 6% of the cell length, i.e. the thin lens approximation is satisfied. The aperture is given by (1.55)

$$\phi \propto |x(s)| = \sqrt{\frac{\varepsilon\beta(s)}{\gamma_{r,i}}} \tag{1.74}$$

where the gamma factor at injection is

$$\gamma_{r,i} = \frac{E}{mc^2} = \frac{450\times10^9\times1.602\times10^{-19}}{1.672\times10^{-27}\times\left(3\times10^8\right)^2} = 480. \tag{1.75}$$

The aperture for 15 σ will be

$$\varphi = 2\times15\sqrt{\frac{\varepsilon\beta(s)}{\gamma_{r,i}}} = 30\sqrt{\frac{3.75\times10^{-6}\times(2+\sqrt{2})\times50}{480}} = 0.035\ \mathrm{m} \tag{1.76}$$

so the aperture should be larger than 52 mm, accounting for 50% margin. LHC quadrupoles and dipoles have an aperture (diameter) of 56 mm.

Example 1.6: Using the LHC parameters, rescale them to get the quadrupole aperture and integrated strength for the FCC study of a proton collider, with 50 TeV energy, assuming an injection energy of 3 TeV and a L=50 m of quadrupole spacing. Assuming a field on the coil B_m=12 T, compute the quadrupole length and the fraction of cell covered by quadrupoles.

The integrated gradient scales with the energy, the curvature radius, and the inverse of the cell length.

$$Gl_q = \frac{\sqrt{2}B\rho}{L} = \frac{\sqrt{2}E[\mathrm{GeV}]}{0.3L} = \frac{\sqrt{2}\times50000}{0.3\times50} = 4700\,\mathrm{T}. \tag{1.77}$$

So the required strength is ~7 times larger than in the LHC. The aperture scales with inverse the square root of the energy and inverse of the beta function

$$A_{FCC} = \sqrt{\frac{450}{3000}}A_{LHC} = \frac{56}{2.58} = 22\ \mathrm{mm}. \tag{1.78}$$

Note that the scaling is far from being exact since there is a component which is a constant, so the real aperture requirement will be much larger than 22 mm. Having said this, with a 22 mm aperture quadrupole, we can make the exercise of having 12 T peak field, and the gradient can be estimated as

$$G = \frac{2B_m}{\lambda\phi} = \frac{2\times 12}{1.15\times 0.022} = 950\,\mathrm{T} \tag{1.79}$$

and the length is 4700/950=5.0 m.

Example 1.7: Rediscuss the above case doubling the cell length. Having L=100 m, the integrated strength will decrease by a factor two

$$Gl_q = \frac{\sqrt{2}B\rho}{L} = \frac{\sqrt{2}E[\mathrm{GeV}]}{0.3L} = \frac{\sqrt{2}\times 50000}{0.3\times 100} = 2360\,\mathrm{T} \tag{1.80}$$

but the aperture will get larger

$$\phi_{FCC} = \sqrt{\frac{450}{3000}\frac{100}{50}}\phi_{LHC} = \frac{56}{1.82} = 0.031\,\mathrm{m}. \tag{1.81}$$

With a 31 mm aperture quadrupole, having 12 T on the pole, the gradient can be estimated as

$$G = \frac{2B}{\lambda\phi} = \frac{2\times 12}{1.15\times 0.031} = 670\,\mathrm{T} \tag{1.82}$$

and the length is shorter 2350/670=3.5 m. The advantage is to have half of the quadrupoles, with 30% shorter length, at the price of a 50% larger aperture.

Acknowledgements

I wish to acknowledge W. Scandale for having introduced me to the concepts of beam optics in the arcs of hadron colliders, during the studies done on the LHC nonlinear dynamics the end of the 80's.

References

1. M. Conte, W. McKay, "An introduction to the physics of particle accelerators" World Scientific, Singapore (1991)
2. C. Kittel, W. D. Knight, M. A. Ruderman, "Berkeley physics course, Vol I: Mechanics" McGraw-Hill, London (1973)
3. A. Chao, M. Tigner, "Handbook of accelerator physics and engineering" World Scientific, Singapore (1999)
4. A. Einstein in letter to Lincoln Barnett, 19 June 1948, quote from L. B. Okun, "The concept of mass" Phys. Today 42 (1989) 31-36
5. E. F. Taylor, J. A. Wheeler, "Spacetime Physics" W.H. Freeman and Company, New York (1992)
6. E. D. Courant, H. S. Snyder, "Theory of alternating gradient synchrotron" Ann. Phys. **3** (1958) 1-48
7. E. Courant, M. S. Livingston, and H. Snyder, "The strong focusing synchrotron – a new high energy accelerator" Phys. Rev. **88** (1952) 1190
8. N. C. Christofilos, unpublished manuscript (1950)

Version 0.2

Ezio.Todesco@cern.ch

Chapter 2

Magnet requirements in circular accelerators: the interaction regions

Plan of the chapter
In this chapter we discuss the magnet requirements in the interaction regions. We first introduce the definition of luminosity (section 2.1), summarizing the main physical effects that sets limits to high luminosities (section 2.2). In section 2.3 we outline the requirements on the experimental magnets; the longitudinal size of the experiment is an essential parameter for the optics around the interaction region. In section 2.4 we present the layout for the separation and recombination dipoles in a collider with separate beam pipes. We then discuss the need of a crossing angle to avoid parasitic beam collisions in section 2.5, and its adverse effect on luminosity via the geometric reduction factor. Finally, in section 2.6 we discuss the optics around the collision points, showing the requirements in terms of gradients, aperture and length for a triplet system aiming at minimizing the beam size in the interaction point. The whole chapter is focussed on the case of LHC and HL-LHC.

2.1 Luminosity of a collider

Besides the energy, the other relevant parameter of a circular collider is the luminosity, which is related to the quantity of collisions per second. Luminosity L is expressed in $cm^{-2}\ s^{-1}$, and not simply in s^{-1} (collisions per second), since for a given collider events with larger probability will give more collisions [1]. The probability associated to a given type of collisions is expressed as a cross-section, that intituively can be imagined as the surface of the colliding particle. The number of events per second is then the product between the luminosity and the cross-section of that event. Cross-sections are expressed in barns, where 1 barn= 10^{-24} cm^{-2}, i.e. is a square of 10^{-14} m. Typically for the proton-proton cross-section at the LHC 7+7 TeV energy the cross-section is of the order of 0.1 barn: therefore the colliding particles have a "size" that is 0.3×10^{-14} m. In Appendix B we will show that the size of the atom is of the order of 0.5×10^{-10} m: therefore we are four order of magnitudes below the atom size. Indeed, 0.3×10^{-14} m is the scale of the classical radius of the electron, defined as the size of a charged particle whose electrostatic energy is equal to the rest energy mc^2. For the LHC [2], the inelastic cross-section of proton collisions at 7 TeV is ~0.06 barn, the nominal luminosity is 10^{34} $cm^{-2}\ s^{-1}$, and therefore we have about 6 billions of collisions per second

The luminosity is a characteristic of the accelerator, independent of the collision type. It is given by

$$L = \frac{n_b N_b^2 f_{rev} \gamma_r}{4\pi\varepsilon_n \beta^*} F(\beta^*)\,; \tag{2.1}$$

where we have

- Main parameters of the collider: γ_r is the relativistic factor of the collision energy, and f_{rev} is the frequency of the accelerator, i.e. the particle velocity divided by the accelerator length;
- Beam parameters: the beam is not continuous, but separated in n_b bunches, each containing N_b particles per bunch;
- Injector features: ε_n is the emittance of the beam in collision, which has been introduced in chapter 1. It is given by the injection emittance, plus some increase (emittance blow-up) given by the dynamics of the collider;
- Collider optics: β^* is the beta function in the interaction point (the so called beta-star, see section 2.6). F is a geometrical factor that will be discussed in section 2.5.

Luminosity is usually given in CGS units, i.e. $cm^{-2}\ s^{-1}$, whereas engineers and applied physicists usually use International System (SI) or mixed units, typically either m or cm for β^*, and mm mrad for the emittance. The safest choice is to use SI units and to avoid mixing units when possible, as we do here - when we had exceptions as for the energy in GeV in the previous chapter, we explicitly state the units. So, luminosity of Eq. (2.1) provides values in $m^{-2}\ s^{-1}$ and not in $cm^{-2}\ s^{-1}$.

To increase luminosity there are two main paths: (i) collide more particles; note that the dependence on particles per bunch N_b is quadratic; (ii) reduce the size of the beam in the collision, i.e. $\varepsilon_n\beta^*$; this can be done either through decreasing the emittance of the injectors or by having an optics that allows smaller beta functions in the interaction points. Note that there is also a dependence on the main collider parameters: luminosity is (i) proportional to the energy through the gamma factor, and (ii) is inverse proportional to the collider size, through the frequency. A summary of the luminosity parameters for several colliders is given in Table 2.1.

Table 2.1: Summary of main luminosity parameters of some colliders

		Tevatron		HERA		LHC nom.	LHC 2013	LHC 2020	HL-LHC
Particle		p+	p-	p+	e-	p+	p+	p+	p+
E	(GeV)	900		820	30	7000	4000	6800	7000
l	(km)	6.28		6.4		27.6	27.6	27.6	27.6
N_b	(adim)	7.00E+10	2.90E+10	3.60E+10	1.01E+11	1.15E+11	1.35E+11	1.40E+11	2.20E+11
n_b	(adim)	6	6	210	210	2808	1404	2500	2760
γ_r	(adim)	958		872	55080	7448	4256	7235	7448
f_{rev}	(s^{-1})	4.78E+04		4.70E+04		1.09E+04	1.09E+04	1.09E+04	1.09E+04
β^*	(m)	0.55		2.2	10	0.55	0.6	0.3	0.15
F	(adim)	0.71		1	1	0.86	0.9	0.54	0.716
ε_n	(m)	4.20E-06	3.00E-06	1.00E-05	1.00E-05	3.75E-06	2.20E-06	2.20E-06	2.50E-06
L	($cm^{-2}\ s^{-1}$)	1.6E+30		5.30E+36		1.0E+34	6.42E+33	2.5E+34	1.6E+35
L_{level}	($cm^{-2}\ s^{-1}$)							2.5E+34	5.0E+34
θ	(μrad)					150	150		300
σ_{in}	(mbarn)					60	60	60	60
Pile-up	(adim)					20	25	55	100

Accelerators can collide the same particle type, as the LHC: in this case the counter-rotating beams require opposite magnetic fields, i.e the magnets shall have a double aperture, and the dipoles shall provide opposite fields in each aperture. This is doubling the magnet system. The other option is to collide particle and antiparticle (as Tevatron [3]): in this case the two beams can circulate in the same pipe, and profit of the same magnetic field. They rotate in opposite directions since their charge is opposite. The two beams are sent along spirals, centered on the magnet axis, so to avoid that they

collide in each section of the machine. The spirals are collapsed on each other in the experiments. Luminosity equation reads

$$L = \frac{n_b N_b^+ N_b^- f_{rev} \gamma_r}{4\pi\sqrt{\varepsilon_n^+ \beta^{+*}}\sqrt{\varepsilon_n^- \beta^{-*}}} F(\beta^{*+}, \beta^{*-}) ; \tag{2.2}$$

where N^+ is the number of particles per bunch and N^- is the number of antiparticles, and emittance and β^* can have different values for the two types of beams. The disadvantage is that generating high intensity bunches of antiparticles is not trivial; Tevatron luminosity has always been limited by the number of antiprotons.

Example 2.1: Compute the collisions per second in the LHC, if it was filled with cherries instead of protons, assuming the nominal luminosity 10^{34} cm^{-2} s^{-1}.
Assuming for the cherry cross-section 1 cm^2, one would have 10^{34} collisions per second.

Example 2.2: Compute the energy stored in one nominal LHC beam; express this energy in units of chemical energy stored in a BigMac.

The energy of the beam is given by

$$U_{beam} = n_b N_b E = 2808 \times 1.15 \times 10^{11} \times 7 \times 10^{12} \times 1.602 \times 10^{-19} = 360 \text{ MJ} \tag{2.3}$$

Since one BigMac has ~500 kcal, i.e., ~2 MJ (1 cal=4.18 J), the beam is equivalent to ~180 BigMacs.

Example 2.3: Considering that the LHC complex has a power consuption of a maximum of 200 MW, and that the accelerator is refilled every 12 hours, estimate the efficiency of the LHC in concentrating the grid power in protons. Estimate the maximum efficiency, considering a continuous refilling and acceleration.

The power given to the protons (two beams) is

$$P_{beam} = \frac{2E_{beam}}{12 \text{ hours}} = \frac{2 \times 360 \times 10^6}{43200} = 17 \text{ kW} . \tag{2.4}$$

So since the LHC complex consumes ~200 MW, the efficiency is the energy concentration is of the order of 1:10000. Considering that the acceleration takes about 1000 s, and the injection and ramp down another 2000 s, in principle one could accelerate every hour, and the power given to the beams would be 200 kW. In these conditions, there is a factor 1:1000 in the efficiency.

2.2 Luminosity limitations

There are several physical phenomena that limit luminosity; before focusing on the requirements on the insertion region magnets we will briefly recall a few of them.

Limitations due to beam-beam tune shift; when the beams collide, one proton sees the electrostatic field of the other bunch, which is proportional to the density of charges in the transverse plane. Beyond a threshold of this density, the Coulomb interaction creates a tuneshift that makes the beam unstable: this puts an limit on the ratio between the bunch intensity and the emittance N_b/ε_n (defined as the beam brilliance). The beam-beam tune shift parameter [1] is defined as

$$\xi \equiv n_{IP} \frac{r_p N_b}{4\pi\varepsilon_n} ; \tag{2.5}$$

where r_p is the classical radius of the proton (see also Appendix B for the definition),

$$r_p \equiv \frac{e^2}{4\pi\varepsilon_0 m_p c^2} = \frac{\left(1.602\times10^{-19}\right)^2}{4\pi\times8.85\times10^{-12}\times1.672\times10^{-27}\times\left(3\times10^8\right)^2} = 1.53\times10^{-18}\ \text{m} \qquad (2.6)$$

and n_{IP} is the number of interaction point. This empirical limit is usually set at 0.01, and depends on nonlinearities in the transverse motion (the less the nonlinearities, the larger the beam-beam limit). Please note that we are using the slightly misleading notation ε_n for transverse beam emittance, and ε_0 for the vacuum permittivity.

Example 2.4: Estimate the beam-beam tune shift for the nominal LHC and for the 2015 run [4], using data of Table 1.

The beam-beam tune shift for the nominal parameters (see Table 1) is given by

$$\xi \equiv n_{IP}\frac{r_p N_b}{4\pi\varepsilon_n} = 2\frac{1.53\times10^{-18}}{4\pi}\frac{1.15\times10^{11}}{3.75\times10^{-6}} = 0.0075 \qquad (2.7)$$

and becomes 0.02 for the parameters used in the 2015 run, thanks to the 30% smaller emittance and to the 50% larger beam population:

$$\xi = 2\frac{1.53\times10^{-18}}{4\pi}\frac{1.7\times10^{11}}{2.2\times10^{-6}} = 0.019. \qquad (2.8)$$

This is twice what is considered as a safe limit for beam-beam. The LHC succesfully operated with this high brilliance thanks to the low level and to the good control of the accelerator nonlinearities.

Limitations due to the electron cloud: when the bunches are too closely spaced, the electrons of the beam screen that are extracted by the particles in the beam halo can be accelerated by the next bunch, thus extracting other electrons and creating an electron cloud in the beam pipe through an avalanche effect (see Fig. 2.1). This effect, discovered in the 80's [5], is related to the bunch intensity and to the number of bunches; it can be partially cured by a treatment of the beam screen surface to reduce the probability of extracting electrons. The electron cloud sets a lower bound on the bunch spacing.

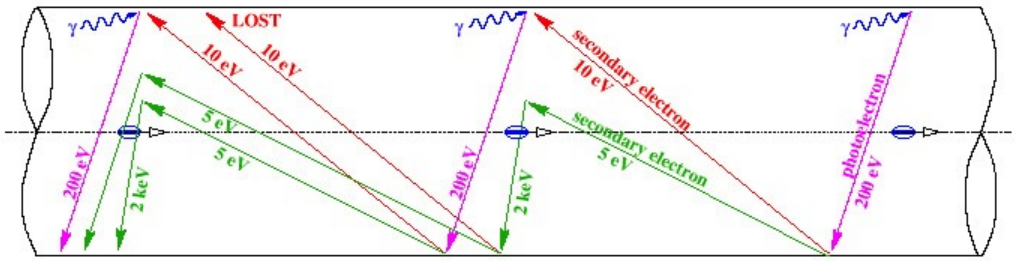


Fig. 2.1: Mechanism of electron cloud formation

Limitations due to the synchrotron radiation: a charged particle going through a bending dipole emits radiation, and the energy loss per turn is given by

$$E_s = \frac{e^2}{3\varepsilon_0}\frac{\gamma_r^4}{\rho} \qquad (2.9)$$

where ρ is the dipole bending radius. The energy loss per turn scales with the fourth power of the relativistic gamma and with the inverse of the curvature radius. In many texts you will see equations where the gamma factor is replaced by the ratio between the particle energy and the mass.

$$E_s = \frac{e^2}{3\varepsilon_0}\frac{E^4}{\left(mc^2\right)^4\rho}. \tag{2.10}$$

This is shows that for a given energy the power loss scales with the inverse of the fourth power of the mass. This is why proton colliders present the advantage with respect to electron colliders of having ~10^{13} smaller energy loss. A relevant quantity for the cryogenics is the power loss per meter:

$$P_s = n_b N_b \frac{E_s f}{L_T} \approx n_b N_b \frac{E_s c}{L_T^2} = n_b N_b \frac{e^2 c}{3\varepsilon_0 L_T^2}\frac{\gamma_r^4}{\rho}. \tag{2.11}$$

The main limitation here is given by the cryogenic power required to remove this heat from the beam screen and from the magnet.

Example 2.5: Compute the energy loss per turn and per particle, and the power loss per meter in the LHC [2], in the FCC-hh study [6], and in the LEP [7].

In the LHC the energy loss per turn is given by

$$E_s = \frac{\left(1.602\times10^{-19}\right)^2}{3\times8.85\times10^{-12}}\frac{7450^4}{2801} = 1.06\times10^{-15}\ \text{J} \tag{2.12}$$

and the power loss per meter is

$$P_s = 2808\times1.15\times10^{11}\frac{1.06\times10^{-15}\times3\times10^8}{26700^2} = 0.144\ \text{W/m}. \tag{2.13}$$

For the FCC the energy loss per particle per turn is ~700 times larger, since the length is four times larger, and the energy 7 times larger (scales with γ_r^4/L_T)

$$E_s = \frac{\left(1.602\times10^{-19}\right)^2}{3\times8.85\times10^{-12}}\frac{53200^4}{10400} = 7.4\times10^{-13}\ \text{J} \tag{2.14}$$

and the power loss per meter is ~170 times larger (scales with γ_r^4/L_T^2)

$$P_s = 10400\times1.15\times10^{11}\frac{7.4\times10^{-13}\times3\times10^8}{26700^2} = 26.7\ \text{W/m}. \tag{2.15}$$

For the LEP, accelerating electrons at 115 GeV, one has

$$E_s = \frac{\left(1.602\times10^{-19}\right)^2}{3\times8.85\times10^{-12}}\frac{225000^4}{3030} = 8.13\times10^{-10}\ \text{J} \tag{2.16}$$

and the power loss per meter is

$$P_s = 3030\times2\times10^{11}\frac{8.13\times10^{-10}\times3\times10^8}{26700^2} = 207\ \text{W/m}. \tag{2.17}$$

Limitations due to the beta function in the interaction point: one can prove (see Appendix A) that in the experimental region around the interaction point, where no transverse fields are present, the beta function has the following parabolic form

$$\beta(s) = \beta^* + \frac{s^2}{\beta^*} \tag{2.18}$$

where s=0 is the interaction point in the centre of the experiment. Therefore in the first magnets seen by the beam after the collisions, which are at a distance l^*, the beta function is proportional to the inverse of β^*.

$$\beta(l^*) = \beta^* + \frac{l^{*2}}{\beta^*} \approx \frac{l^{*2}}{\beta^*}. \tag{2.19}$$

This relation has a deep physical meaning: if we want to reduce the beam size in the interaction point via a reduction of the beta function, we must have very large aperture magnets close to the experiments. The possibility of approaching the first magnets (reducing l^*) is not viable, since the experimental magnets need some space, as we will outline in the next section. The relation between l^*, β^* and magnet aperture will be discussed in more detail in section 2.5. Moreover, the luminosity increase due to the reduction of the beam size in the interaction point is limited by the crossing angle and by the geometric reduction factor F (see Eq. 2.1). This will be discussed in section 2.4.

Limitations due to the number of events per crossing (aka pile-up). It is defined as the number of events that take place at the same time in the detector, and is given by the luminosity times the cross-section, divided by the revolution frequency and the number of bunches:

$$p_u \equiv \frac{L}{n_b f_{rev}} \sigma_{in}. \tag{2.20}$$

The challenge for the detector hardware and for the analysis software is to be able to separate the different events. In the nominal LHC the pile-up is ~20, i.e., 20 events are present at the same time in the detector traces. To limit the pile-up keeping a large luminosity one has to increase the number of bunches (i.e., decrease the bunch spacing) rather than increase the number of particles per bunch. So the pile-up sets a limit on the "easy" way of increasing luminosity through increasing N_b. On the other hand, the increase of number of bunches is in conflict with the electron cloud requirements on the minimal bunch spacing.

2.3 Experiment requirements

The experimental magnets are spectrometers that bend the collision debris allowing to reconstruct the momentum/charge ratio from the deviation of the particle from a straight trajectory. There are different configurations, as a simple solenoidal field (see Fig. 2.2 left) or a toroidal field (see Fig. 2.2, center); here we will focus on the solenoidal configuration, that can be approximated by a constant magnet field parallel to the beam over a cylinder of radius R_t and semi-length l*.

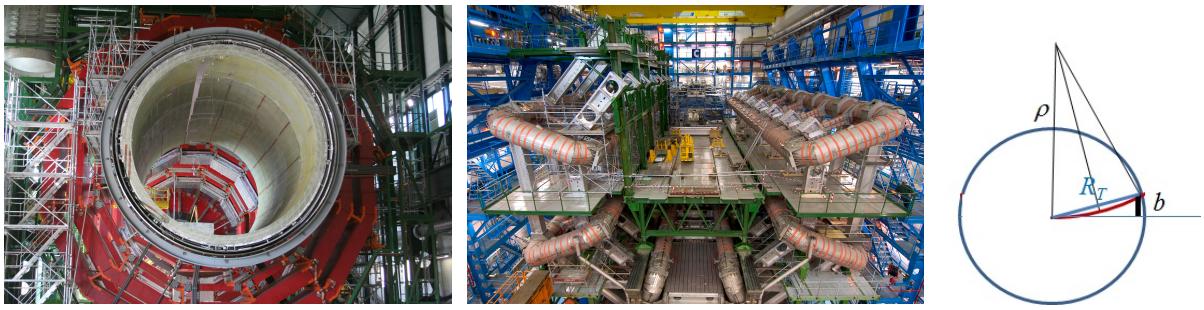


Fig. 2.2: the CMS solenoid (left) and the toroidal coils of ATLAS (center), and schematics of the trajectory of a charged collision debris (in red) in the plane transverse to the beam, where R_t is the experimental solenoid radius and ρ is the curvature radius induced by the solenoid magnetic field.

As shown in the previous chapter, a charged collision debris, with momentum p in the plane transverse to the beam, will be curved by the magnetic field of the solenoid B on a curvature radius ρ according to

$$p = eB\rho \,. \tag{2.21}$$

Assuming that the particle is perpendicular to the beam trajectory, it will reach the outer radius of the detector solenoid R_T with a deviation b from the straight trajectory (see Fig. 2.2, right). Since the 90° triangles made with b and R_T,, and made with R_T /2 and ρ are similar, one has

$$\frac{b}{R_T} = \frac{R_T / 2}{\rho} \qquad\qquad b = \frac{R_T^2}{2\rho} = \frac{eBR_T^2}{2p} = 0.15\frac{BR_T^2}{E\left[\text{GeV}\right]}. \tag{2.22}$$

Therefore, the precision in measuring the particle momentum is proportional to R_t^2B: to improve the precision it is much more effective to increase the detector transverse radius than its magnetic field. This is the reason behind experimental detector sizes of the order of 10 m. To have the same deviation b with the same particle energy, reducing the radius of a factor ten would require increasing the magnetic field by a factor 100.

Example 2.6: In the ALEPH experiment [8] we have a solenoid with 1.5 T magnetic field and 2.65 m radius and 6.5 m longitudinal length. What is the deviation b for a 115 GeV particle at the exit of the solenoid?

$$b = 0.15\frac{BR_T^2}{E\left[\text{GeV}\right]} = 0.15\frac{1.5\times 2.65^2}{115} = 13.7 \text{ mm}\,. \tag{2.23}$$

So considering a resolution in the detector of 0.1 mm, one can measure the momentum with 1% precision.

Example 2.7: Use a scaling of ALEPH detector to propose a solenoid for the LHC.

Since the energy increase is 61 (from 115 GeV to 7 TeV), we can share this increase between a factor 3 in the magnetic field (i.e. going to 4.5 T) and the remaining factor 61/3 =20 should be covered by a factor ~4.5 increase in solenoid radius, i.e. from 2.65 m to 12 m.

Example 2.8: In LHC we have a distance of the first magnet to the interaction point l^*=23 m. What could be used for the FCC, which has a ~7 times larger energy?

Since $7^{1/3}$=1.9, one can use a 1.9 larger radius (and length of the experiment), and a 1.9 larger magnetic field. Therefore, l^*~23×1.9=44 m, and the field in the solenoid would increase from 4 T to 8 T.

2.4 Separation and recombination dipoles for colliders with two beam pipes

We have shown in Chapter 1 that a collider based on two beams of identical particles require two beam pipes, with the main dipole fields in opposite direction. Until LHC, the magnetic system of a collider was based on independent magnets for each beam pipe. In the LHC, the design has been based on a dipole providing opposite field in two apertures, separated by a minimal distance (~200 mm) sufficient to avoid a large electromagnetic coupling between the two beam pipes. In both cases, the beams travel in separate apertures along most of the accelerator and then they are put on the same path via a separation and a recombination dipole [2], as shown in Fig. 2.3. The recombination dipole is a double aperture dipole, with field in the same direction, making converge the beams up to the separation dipole, where they get a second dipolar kick to place them on the same path towards the experiment. After the separation dipole one has the triplet of quadrupoles, i.e. the magnets closer to

the experiment. In Section 2.6 we will show the interest in placing the triplet as close as possible to the experiment to get the smallest possible beam size in the collision point without requiring too large apertures for the magnets. Other layouts are possible, as placing the dipoles closer to the experiment, and then the quadrupoles (dipole-first [9] layout), but are less effective in reducing the IP beam size.

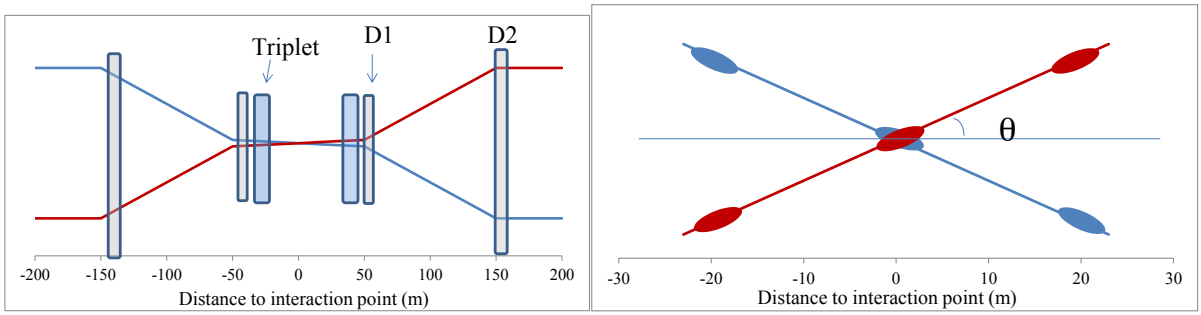


Fig. 2.3: Schematic of the orbit in a layout with a triplet, a separation D1 and a recombination D2 dipole (left) and magnification of the crossing angle around the interaction point (right).

Example 2.9: The integrated field of the separation and the recombination dipole in the LHC around ATLAS and CMS is 27 T m at 7 TeV energy. Compute the angular kick given by each dipole, and the distance between the baricentre of the dipoles knowing that the beam interdistance is 192 mm.

Since the curvature radius of LHC is 2808 m and the dipole field is 8.3 T, the integrated required field to have a 2π rotation is 2×π×2808×8.3=146 kT m. Therefore 1 mrad kick is given by 146/2π=23.3 kT m. Since the dipole have a 27 T m integrated field, they will change the beam angle trajectory by 27/23.3=1.16 mrad. This kick shall displace the beam by half of the distance between the beam apertures in the arc, i.e. 192/2=96 mm. Therefore the distance between the D1 and D2 baricentre shall be 96/1.16= 82 m.

Example 2.10: In FCC-hh collider [6], with 50 TeV beam energy, the beam separation is 250 mm. Assuming a distance between D1 and D2 of 100 m, estimate the integrated field in D1 and D2.

The dipole kick should be 250/2/100= 1.25 mrad (not far from the LHC values). For a 50 TeV collider as FCC-hh, 1 mrad kick is given by

$$1 \text{ mrad} = \frac{\rho B}{1000} = \frac{E\left[GeV\right]}{1000 \times 0.3} = \frac{50}{0.3} = 167 \text{ T m}\,. \tag{2.25}$$

Therefore the integrated field of the separation/recombination dipoles shall be 167×1.25=208 T m. Using Nb-Ti, with a 8 T field one could use two units of 13-m-long dipoles.

2.5 Crossing angle and geometric reduction factor

When the bunch spacing is smaller than the longitudinal size of the experiment, one has to avoid the collisions taking place not in the center of the detector (usually called parasitic collisions). For instance, in the LHC the bunch spacing is 7.5 m, the longitudinal size of the experiments is $2l^*$=46 m, so one has six additional points inside the experiment where bunches collide. The solution is to have the beams colliding not exactly face to face, but with a small crossing angle θ (see Fig. 2.3, right).

Let us require that the separation of the beams at a distance s of the interaction point is equal to n sigma of the beam (for the LHC one takes n~10)

$$2s\theta = n\sqrt{\frac{\varepsilon_n \beta(s)}{\gamma_r}}\,. \tag{2.26}$$

Since close to the interaction point the beta function behaves as s^2/β^* (see Eq. 2.19), one has

$$2s\theta = n\sqrt{\frac{\varepsilon_n s^2}{\beta^*\gamma_r}} \qquad\qquad \theta = \frac{n}{2}\sqrt{\frac{\varepsilon_n}{\beta^*\gamma_r}} \tag{2.27}$$

so to keep this distance between the beams, the crossing angle has to scale with the inverse of the square root of β^*.

The crossing angle gives a reduction of the luminosity of a factor F (see Eq. 2.1) given by

$$F(\theta) = \sqrt{\frac{1}{1-\left(\frac{\sigma_z}{\sigma_x}\frac{\theta}{2}\right)^2}} \approx \sqrt{\frac{1}{1-\frac{\gamma_r}{\varepsilon_n\beta^*}\left(\frac{\theta}{2\sigma_x}\right)^2}} \tag{2.28}$$

where σ_z is the longitudinal length of the bunch. Here we do not give any proof of this equation: this reduction is due to the fact that the bunches are not colliding head on, and therefore the area where the beam intersect is reduced by the crossing angle (see Fig. 2.3, right). At the same time, the reduction of β^* requires a larger crossing angle as given in (2.27) to avoid parasitic collisions. One can observe that in the limit of small β^* (i.e. very large magnet aperture close to the interaction regions), the geometric factor becomes proportional to β^*

$$F(\theta) = \sqrt{\frac{1}{1-\frac{\gamma_r}{\varepsilon_n\beta^*}\left(\frac{\theta}{2\sigma_x}\right)^2}} \xrightarrow[\beta^*\to 0]{} \beta^* . \tag{2.29}$$

In this limit, the F factor perfectly compensates for the luminosity increase proportional to $1/\beta^*$, the positive effect of reducing β^* vanishes and luminosity saturates at a constant value. In Fig. 2.4 we give the dependence of crossing angle and geometric reduction as a function of β^*, and the values chosen for the LHC operation. Note that the use of crab cavities [3] allow to tilt the bunches and collide them head on even with a nonzero crossing angle, i.e., to remove the geometric reduction factor. This option will be implemented in the LHC upgrade, aiming at a strong reduction of β^*.

Example 2.11: Compute the crossing angle for the LHC, for the HL-LHC project [10] and for the FCC study, requiring a 10 sigma separation.

For the LHC one has

$$\theta = \frac{10}{2}\sqrt{\frac{3.75\times 10^{-6}}{0.55\times 7450}} = 150\ \mu\text{rad} . \tag{2.30}$$

The four times smaller beta function, foreseen for HL-LHC, gives a twice larger crossing angle:

$$\theta = \frac{10}{2}\sqrt{\frac{3.75\times 10^{-6}}{0.15\times 7450}} = 290\ \mu\text{rad} . \tag{2.31}$$

The crossing angle proportional to the inverse of the square root of the energy. Therefore for a case as the FCC-hh study, having 50 TeV beam energy and a β^* =0.3 m, the crossing angle can be smaller

$$\theta = \frac{10}{2}\sqrt{\frac{3.75\times 10^{-6}}{0.3\times 53200}} = 77\ \mu\text{rad} . \tag{2.32}$$

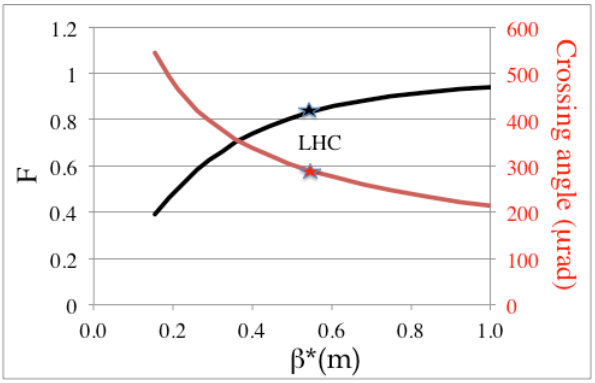


Fig. 2.4: Dependence of the crossing angle and of the geometric reduction factor on β*, and the case of the LHC

2.6 Optics around the interaction regions and triplet requirements

As outline in Section 2.2, and discussed in more detail in Appendix A.7, in a region free of magnetic field the beta function is quadratic in the longitudinal coordinate:

$$\beta(s) = \beta^* + \frac{s^2}{\beta^*} \tag{2.33}$$

where s is the distance along the beam axis to the interaction point, and β^* is the beta function in the interaction point. The consequence of this equation is that to obtain small β^*, the size of the beam in the first magnet placed after the experimental area at a distance l^* from the interaction point, becomes very large (see Fig. 2.5)

$$\beta(l^*) = \beta^* + \frac{l^{*2}}{\beta^*} \approx \frac{l^{*2}}{\beta^*}. \tag{2.34}$$

The final focus quadrupoles are used to bend the parabolic beta functions in the experiment and to bring them down to values that can be matched to the beta functions in the arcs. To give an order of magnitude in the LHC, l^* is 23 m, nominal β^* is 0.55 m, and using (2.34) $\beta^*(l^*)$ is 1050 m, whereas in the arc it ranges between 30 and 170 m. The reduction of these large beta functions at the boundary of the experiment is done through a series of quadrupoles (3 in the LHC); one needs at least two quadrupoles to focus the beam in both planes, since a single quadrupole is focusing in one plane but defocusing in the other one. A triplet configuration F-D-F is very effective to steer the beta functions down to matchable values for the arc, and it is shown in Fig. 2.5 for the LHC case.

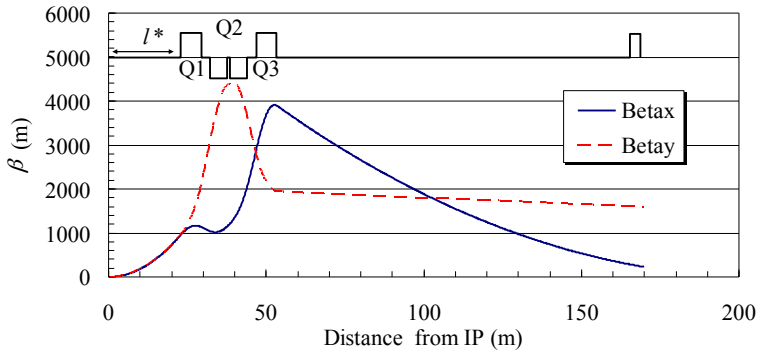


Fig. 2.5: Beta functions close the LHC interaction point, and position of the triplet and of the quadrupole Q4.

In a triplet the focusing force is approximately (*–f*, *2f*, *f*) and having the same gradients in each quadrupole, the ratio between the lengths of Q1/Q3 and Q2 is around a factor 2. Q2 is usually split in two magnets, giving the paradox that a triplet is usually composed of four magnets. The ratio between the Q1/Q3 length and the Q2 length is used to fine tune the values of the maximum beta function in both planes, making them equal, so to best exploit the precious magnet aperture. This optimal ratio is related to the gaps between the magnets, to the total length of the triplet l_t, and to l^*. Note that in the LHC baseline shown in Fig. 2.5 this optimization is not perfectly realized since the gaps between magnets were changed after setting the quadrupole lengths.

What is the required focusing strength of the quadrupole triplet ? Keeping the optics analogy, one has

$$\frac{Gl_q}{B\rho} = \frac{1}{f} \approx \frac{1}{l^*} \tag{2.35}$$

and therefore for the whole triplet

$$Gl_t \approx \frac{4B\rho}{l^*} . \tag{2.36}$$

For instance, in the LHC one has Q1 and Q3 length of 6.23 m, and Q2 of 11 m, split in two 5.5-m-long magnets, with a 200 T/m gradient. This gives an integrated gradient of 4700 T, to be compared to the right-hand side of (2.38) that gives 1000 T.

$$Gl_t = 200 \times (6.23 \times 2 + 11) = 4700 \text{ T} \qquad \frac{4B\rho}{l^*} = \frac{4 \times 8.3 \times 2808}{23} = 4070 \text{ T} . \tag{2.37}$$

However, the maximum beta function is much larger than the beta function at the entrance of the Q1 given by Eq. (2.35) as the triplet is far from being a thin lens: in the LHC case shown in Fig. 2.5 the total length of the triplet (23.5 m) is similar to the distance to the interaction point (23 m). Therefore the scaling of (2.36) can be improved to

$$Gl_t \approx k \frac{4B\rho}{l^* + \frac{l_t + l_g}{2}} . \tag{2.38}$$

where l_g is the total length of gaps in the triplet, and the additional factor in the denominator accounts for the real baricentre of the triplet, and κ accounts for the deviations from thin lens optics, with $\kappa \sim 1.8$ in the range $l^* < l_t < 2l^*$.

Note that when the beam is injected, the optical functions in the interaction points are kept at values which are close to the minimum beta function in the arcs. After the acceleration to nominal energy, the β^* is squeezed, i.e., it is reduced to its nominal value (needed for producing a large luminosity). During the squeeze, the gradient of the triplet is constant, since as discussed in this chapter and mor ein detail in the Appendix A, according to the optics analogy, the gradient is given by the triplet focal length; the squeeze is performed using the quadrupoles in the matching section and in the dispersion suppressor, that lay between the triplet and the arcs. Therefore one the following statements are true:

- Increasing the triplet gradients does not further reduce β^*, but just makes the triplet "astigmatic", i.e. the optics is not closed and the beam does not circulate any more;
- If the triplet does not reach the nominal gradient the optics cannot be produced at any β^*, and the only way to have circulating beams is to reduce the beam energy;
- When the triplet aperture is at the limit of the beam size imposed by β^*, any further reduction of β^* has to go through a larger triplet aperture increase, and not through a larger gradient.

We can now define an algorithm to estimate a layout around the interaction point, and its minimum β^*.

- Select a distance between the IP and the first quadrupole l^*, according to the requirements of the experiments.

- Select the triplet length l_t and the length of the gaps l_g between the magnets. The triplet length together with the magnet technology will give the β^* at the end of the process. Then, one can select a longer length to increase β^* or a shorter length to reduce it, and iterate the process.
- Having fixed the layout, one can estimate the ratio r between the maximum beta function in the triplet and the beta function at the entrance of the triplet. This gives the maximum beta function that can be achieved in the triplet

$$\beta_{\max} = r\frac{l^{*2}}{\beta^*}. \tag{2.39}$$

  Since the triplet is thick, r is larger than one. When the triplet length is between l^* and $2l^*$ one has the approximate relation

$$r \approx 1 + a\frac{l_t}{l^*}. \tag{2.40}$$

  with a~3.6 for gaps of the order of 1 m. This means that if the triplet length is equal to the distance of the first magnet to the interaction point, as in the LHC, the beta function inside the triplet is ~4.6 times larger than the beta function at the entrance.
- The optics computation also gives the required gradient G to match the optics according to the scaling given in Eq. (2.38). For longer l^* and for longer l_t, the required integrated gradient becomes smaller.
- Select a magnet technology: this will give the maximum triplet aperture ϕ that is possible to achieve through the equation

$$\lambda\frac{G\phi}{2} < B_t \tag{2.41}$$

  where B_t is the maximum field provided by the technology. The peak field is given by the gradient times the aperture radius, plus an additional term λ accounting for the overshooting of the peak field in a quadrupole (this issue is discussed in detail in Section 8.8, and in typical cases is λ~1.15).
- We can now compute the minimum β^* by equating the aperture given by Eq. (2.43) to the aperture requirements given by Eq. (1.52) in Chapter 1.

$$\frac{\phi}{2} = n\sqrt{\frac{\varepsilon_n\beta_{\max}}{\gamma_r}}. \tag{2.42}$$

This intricated set of equations and constraints will become more clear in the following examples.

Example 2.12 (LHC case): In the LHC [2], the distance of the triplet to the interaction point is l^*=23 m. The triplet is made by Q1 and Q3, whose length is 6.3 m, and Q2 is split in two magnets of 5.5 m. The triplet coils have a 70 mm aperture. Compute: (i) the quadrupole gradient and check if it can be in Nb-Ti, (ii) the maximum beta function in the triplet for a β^* of 0.55 m, (iii) the aperture required by the beam assuming that the triplet must house two beams of 10 sigma, separated by 15 sigma. (iv) Is there enough space in the magnet to reach β^*=0.25 m? (v) If not, can you reach it by reducing the emittance by 30%?

(i) The length of the triplet is l_t=6.3×2+11=23.6 m, and applying Eq. (2.38) we obtain

$$Gl_t \approx k\frac{4B\rho}{l^* + \frac{l_t}{2}} = 1.8\frac{4\times 8.3\times 2808}{23 + \frac{23.6}{2}} = 4820 \text{ T} \qquad G \approx \frac{4820}{23.6} = 205 \text{ T/m} \tag{2.43}$$

The peak field is estimated using (2.41):

$$B_t = \lambda \frac{G\phi}{2} = 1.15 \times 205 \times 0.035 = 8.3 \text{ T}. \tag{2.44}$$

So the coil is at the limit of the Nb-Ti technology, whose operational fields should not exceed 8.5-9 T.

(ii) The maximum beta function is given by (2.39)

$$\beta_{\max} = r\frac{l^{*2}}{\beta^*} = \left(1 + a\frac{l_t}{l^*}\right)\frac{l^{*2}}{\beta^*} = \left(1 + 3.6\frac{23.6}{23}\right)\frac{23^2}{0.55} = 4400 \text{ m}. \tag{2.45}$$

(iii) The required aperture of 10+15+10=35 sigma is

$$\phi = n\sqrt{\frac{\varepsilon_n \beta_{\max}}{\gamma_r}} = 35\sqrt{\frac{3.75 \times 10^{-6} \times 4400}{7450}} = 52 \text{ mm}. \tag{2.46}$$

So the beam fits within the 70 mm aperture of the magnet.

(iv) If we decrease β^* to 0.25 m, β_{max} increases to 9700 m according to (2.39), and the now beams requires an aperture of

$$\phi = n\sqrt{\frac{\varepsilon_n \beta_{\max}}{\gamma_r}} = 35\sqrt{\frac{3.75 \times 10^{-6} \times 9700}{7450}} = 77 \text{ mm} \tag{2.47}$$

so the beam does not fit in the 70 mm magnet aperture.

(v) If the beam emittance is reduced from 3.75 to 2.2 10^{-6} m, then

$$\phi = n\sqrt{\frac{\varepsilon_n \beta_{\max}}{\gamma_r}} = 35\sqrt{\frac{2.2 \times 10^{-6} \times 9700}{7450}} = 59 \text{ mm} \tag{2.48}$$

and therefore β^*=0.25 m can be achieved; this configuration was used in the RunII of the LHC.

Example 2.13 (LHC upgrade with Nb-Ti): Consider a new layout of the LHC interaction region with longer magnets, $l_{Q1}=l_{Q3}$=9.2 m and $l_{Q2a}=l_{Q2B}$=7.8 m, based on Nb-Ti [11]. Compute (i) the quadrupole gradient, (ii) the aperture that can be provided by Nb-Ti technology and (iii) the minimum β^* that can be reached, assuming that the magnet aperture must exceed the 35 sigma by 20%.

(i) The length of the triplet is l_t=9.2×2+7.8×2=34 m, plus 3 m for gaps, and applying Eq. (2.16) we obtain

$$G \approx \frac{k}{l_t}\frac{4B\rho}{l^* + \frac{l_t + l_g}{2}} = \frac{1.8}{34}\frac{4 \times 8.3 \times 2808}{23 + \frac{34+3}{2}} = 123 \text{ T/m} \tag{2.49}$$

(ii) Assuming a 8.5 T peak field, the maximum aperture is:

$$\phi = \frac{2B_t}{\lambda G} = \frac{2 \times 8.5}{1.15 \times 123} = 120 \text{ mm}. \tag{2.50}$$

(iii) The requirement on the aperture is to have 20% more than 35 sigma:

$$\phi = 1.2 \times 35 \sqrt{\frac{\varepsilon_n \beta_{\max}}{\gamma_r}} \tag{2.51}$$

and therefore the maximum β_{max} is

$$\beta_{\max} = \frac{\gamma_r}{\varepsilon_n}\left(\frac{\phi}{1.2 \times 35}\right)^2 = \frac{7450}{3.75 \times 10^{-6}}\left(\frac{0.120}{1.2 \times 35}\right)^2 = 16200 \text{ m}. \tag{2.52}$$

Using Eq. (2.34) one can compute the β^* reach

$$\beta^* = r\frac{l^{*2}}{\beta_{\max}} = \left(1 + a\frac{l_t}{l^*}\right)\frac{l^{*2}}{\beta_{\max}} = \left(1 + 3.6\frac{34}{23}\right)\frac{23^2}{16200} = 0.021 \text{ m}. \tag{2.53}$$

The above exercise shows a fundamental feature of the triplet optics: *given a magnet technology, one can always reach lower β^* by making the triplet longer, larger and with smaller gradient*.

Example 2.14 (HL-LHC case): consider a new layout of the LHC interaction region in Nb_3Sn, assuming 11.5 T operational peak field and a 150 mm aperture [10]. Compute (i) the gradient and the triplet length, (ii) the minimum β^* that can be reached, assuming that the magnet aperture must exceed the 35 sigma by 20%.

The gradient is given by Eq. (2.43)

$$G = \frac{2B_t}{\lambda\phi} = \frac{2 \times 11.5}{1.15 \times 150} = 133 \text{ T/m}. \tag{2.54}$$

To estimate the triplet length one has to invert Eq. (2.16)

$$Gl_t = k\frac{4B\rho}{l^* + l_t/2}. \tag{2.55}$$

$$l_t = \frac{-2l^* + \sqrt{\left(2l^*\right)^2 + 32kB\rho/G}}{2} = \frac{-46 + \sqrt{46^2 + 32 \times 1.8 \times 8.3 \times 2808/133}}{2} = 32 \text{ m} \tag{2.56}$$

and verify a posteriori that the application range of (2.38) $l^* < l_t < 2l^*$ is satisfied. The β_{max} is

$$\beta_{\max} = \frac{\gamma_r}{\varepsilon_n}\left(\frac{\phi}{1.2 \times 35}\right)^2 = \frac{7450}{3.75 \times 10^{-6}}\left(\frac{0.150}{1.2 \times 35}\right)^2 = 25400 \text{ m}. \tag{2.57}$$

and the minimum β^* is

$$\beta^* = r\frac{l^{*2}}{\beta_{\max}} = \left(1 + a\frac{l_t}{l^*}\right)\frac{l^{*2}}{\beta_{\max}} = \left(1 + 3.6\frac{30}{23}\right)\frac{23^2}{25400} = 0.12 \text{ m}. \tag{2.58}$$

Example 2.15: Consider the layout of the FCC-hh interaction region [6], with 50 TeV energy, and 62 m distance between the interaction point and the first triplet magnet. Estimate the triplet length and gradient for a 140 mm aperture, assuming a Nb_3Sn technology with 12 T peak operational field, and compute the β^* reach.

The gradient is given by Eq. (2.43)

$$G = \frac{2B_t}{\lambda\phi} = \frac{2\times 12}{1.15\times 0.140} = 150 \text{ T/m}. \quad (2.59)$$

To estimate the triplet length one has to invert Eq. (2.16) as in the previous case

$$l_t = \frac{-2l^* + \sqrt{\left(2l^*\right)^2 + 32kB\rho/G}}{2} = \frac{-62 + \sqrt{62^2 + 32\times 1.8\times 16\times 10400/150}}{2} = 98 \text{ m}. \quad (2.60)$$

With such a long triplet, the magnets will be split in several units. The β_{max} can be estimated as

$$\beta_{\max} = \frac{\gamma_r}{\varepsilon_n}\left(\frac{\phi}{1.2\times 35}\right)^2 = \frac{53200}{3.75\times 10^{-6}}\left(\frac{0.140}{1.2\times 35}\right)^2 = 157 \text{ km}. \quad (2.61)$$

and the minimum β^* is

$$\beta^* = r\frac{l^{*2}}{\beta_{\max}} = \left(1 + a\frac{l_t}{l^*}\right)\frac{l^{*2}}{\beta_{\max}} = \left(1 + 3.6\frac{98}{62}\right)\frac{62^2}{157000} = 0.16 \text{ m}. \quad (2.62)$$

Acknowledgements

I wish to thank J. P. Koutchouk for introducing me to the beam optics in the interaction regions, and for the preparatory work done together for the LHC upgrade.

References

1. A. Chao, M. Tigner, "Handbook of accelerator physics and engineering" World Scientific, Singapore (1999)
2. O. Bruning, et al., "LHC design report" CERN-2004-003 (2004)
3. S. Holmes, R. Moore and V. Shiltsev, "Overview of Tevatron collider complex: goals, operations and performance" J. Instrum. 6 (2011)
4. M. Lamont, "Status of the LHC" J. Phys.: Conf. Series 455 (2013) 012001
5. F. Zimmermann, et al., "Review of single bunch instabilities driven by electron cloud" Phys. Rev. ST Accel. Beams 7 (2004) 124801
6. A. Abada, et al., "FCC-hh: the hadron collider" Eur. Phys. J. Spec. Top. 228 (2019) 755-1107
7. A.A.V.V. "LEP design report, v.2: the main ring" CERN-LEP-84-01 (1984)
8. J. M. Baze, et al., "Design, construction and test of the large superconducting solenoid ALEPH" IEEE Trans. Magn. 24 (1988) 1260-1263
9. O. Bruning, et al., "LHC luminosity and energy ugrade: a feasibility study" LHC project report 626 (2002) p. 27
10. O. Aberle, et al., "HL-LHC technical design report" CERN-2020-010 (2020)
11. J. P. Koutchouk, L. Rossi, E. Todesco, "A solution for phase-one upgrade of the LHC low-beta quadrupoles based on Nb-Ti" LHC project report 1000 (2007)

Ezio.Todesco@cern.ch

Chapter 3

Multipolar expansion of magnetic field, and specifications of corrector magnets

Plan of the chapter
Mawell equations impose a very specific constraint on the magnetic field in a charge-free region. As described in the first section of this chapter, this allows defininining an harmonic series, also known as multipole expansion, of the magnetic field. In section 3.2 we consider the expansion in power series of the function 1/(1-*x*), that is the perfect paradigm to understand the field harmonics of a current line; we give the validity limits of the expansion and the level of approximation that can be reached when truncating the series after few orders. In section 3.3, we derive the magnetic field given by an infinite current line and by a current loop. Since the field given by a infinite current line is exactly in the form 1/(1-*x*) using the section 3.2 results we compute in section 3.4 the multipolar expansion of a current line. In section 3.5 we outline how measurements are usually related to integral values along the magnet length, and the conditions necessary to apply the multipolar expansion. The multipoles decay according a very simple law; this can be used to validate the results of field measurements, and to find their precision. Moreover, in case of field anomalies the rate of decay of the multipoles can locate the origin of the anomaly. We describe in section 3.5 these heuristic techniques, showing how they can be applied to analyse specific cases. Finally, in section 3.6 we outline how the strength of corrector magnets is determined in the specific case of the LHC.

3.1 Harmonics of magnetic field

The Maxwell equations can be written as

$$\begin{aligned} \nabla E &= \frac{\rho}{\varepsilon_0} \qquad & \nabla B &= 0 \\ \nabla \times E &= -\frac{\partial B}{\partial t} \qquad & \nabla \times B &= \mu_0 j + \frac{1}{c^2}\frac{\partial E}{\partial t} \end{aligned} \tag{3.1}$$

where as usual we use the international system units; this is not the original form of Maxwell of 1865, but rather the expression reworked by Heaviside in 1884. In a domain free of charges one has

$$\begin{aligned} \nabla B &= 0 \\ \nabla \times B &= 0. \end{aligned} \tag{3.2}$$

Using a cartesian system, the above conditions can be explicitely written as

$$\frac{\partial B_x}{\partial x}+\frac{\partial B_y}{\partial y}+\frac{\partial B_z}{\partial z}=0$$
$$\frac{\partial B_y}{\partial z}-\frac{\partial B_z}{\partial y}=0 \qquad \frac{\partial B_z}{\partial x}-\frac{\partial B_x}{\partial z}=0 \qquad \frac{\partial B_x}{\partial y}-\frac{\partial B_y}{\partial x}=0. \tag{3.3}$$

Let us consider the magnetic field inside the aperture of a magnet, modeled as a cylinder whose length along z is much larger than its diameter in the x,y plane. The above equations can be written as

$$\frac{\partial B_x}{\partial x}+\frac{\partial B_y}{\partial y}=0$$
$$\frac{\partial B_y}{\partial z}=0 \qquad \frac{\partial B_x}{\partial z}=0 \qquad \frac{\partial B_x}{\partial y}-\frac{\partial B_y}{\partial x}=0. \tag{3.4}$$

According to the Cauchy-Riemann theorem [1], a function of a complex variable

$$C(z)=f_x(x,y)+if_y(x,y) \qquad z=x+iy \qquad i\equiv\sqrt{-1} \tag{3.5}$$

can be written as a power series in the complex variable z, over a given domain D

$$C(z)=\sum_{n=0}^{\infty}C_n z^n=\sum_{n=0}^{\infty}C_n(x+iy)^n \qquad z\in D \tag{3.6}$$

if and only if the following conditions on the deritative of the components hold

$$\frac{\partial f_x}{\partial y}+\frac{\partial f_y}{\partial x}=0$$
$$\frac{\partial f_x}{\partial x}-\frac{\partial f_y}{\partial y}=0 \tag{3.7}$$

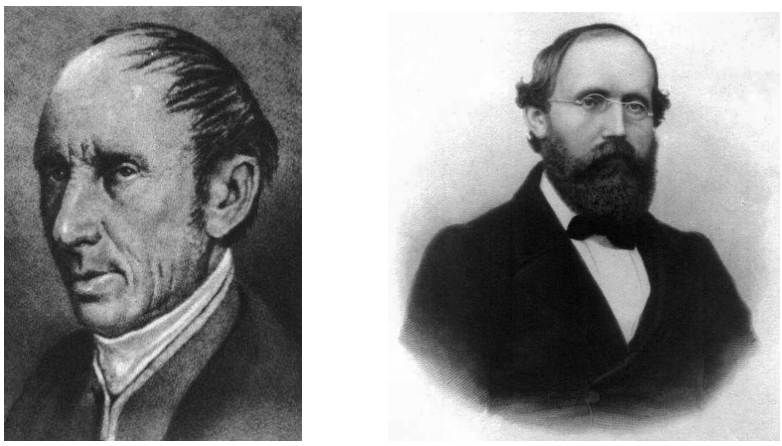


Fig. 3.1: Augustine Louis Cauchy 1789-1857, French (left), and Georg Riemann 1826-1866, German (right)

If we now consider the complex function whose real component is B_y and the imaginary component is B_x, it satisfies the conditions (3.7) and therefore

$$C(z) = B_y(x,y) + iB_x(x,y) = \sum_{n=0}^{\infty} C_n z^n \qquad z \in D \ . \tag{3.8}$$

The coefficients C_n are the multipolar expansion of the magnetic field, widely used in the literature, see for instance [2] for the relation to beam dynamics and magnet construction, and [3] on how to measure them. Their real and imaginary parts are defined as the normal and skew multipoles:

$$C_n = B_n + iA_n \tag{3.9}$$

that is B_1 corresponds to a vertical dipolar field, A_1 to a horizontal dipolar field, B_2 to a quadrupole, A_2 to a skew quadrupole, and so on (see Fig. 3.2). This way of expressing the magnetic field corresponds to a huge reduction of the quantity of information needed to describe it: instead of a continuous function from $\mathbf{R}^2$ (the (x,y) space) to $\mathbf{R}^2$ (the components of magnetic field in x and y), we can use a infinite sequence of numerable complex coefficients C_1, C_2, ...

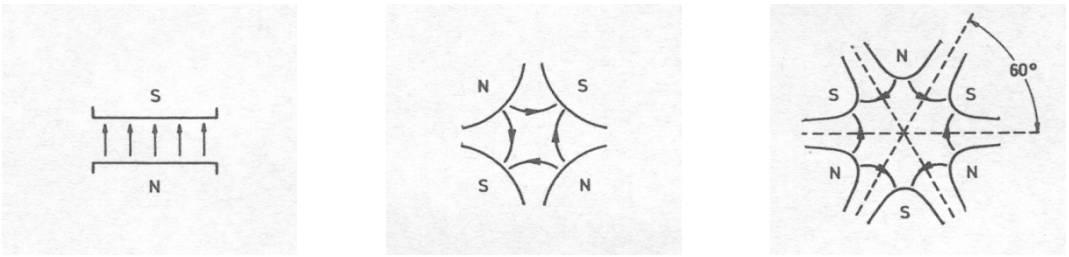


Fig. 3.2: Multipolar expansion of field: normal dipole (left), normal quadrupole (centre) and normal sextupole (right), from [2].

But the story does not ends here: the series (3.8) is rapidly converging, and therefore one can limit to consider the first 10 to 15 coefficients of the expansion, neglecting the higher orders. We will discuss the fundamental topic of the speed of convergence, of errors induced by truncations and of convergence domains in the next section. Moreover, we will see that the symmetry induced by the sequence of poles needed to create a pure multipolar field cancels most of the coefficients C_n: therefore the problem of getting a pure multipolar field with a current distribution is reduced to the optimization of a few parameters, that can be counted on the fingers of one hand. This will be discussed in the next chapter.

However, there are also drawbacks of this approach: the convergence domain of the multipole expansion is much smaller than the domain of definition of the original function, and the expansion has to be used with some care, as discussed in Section 3.2.

Since the magnet is build to create a main component, it is customary to factorize it and write the magnetic field inside the magnet aperture as (we give the case of a dipole)

$$C(z) = B_y(x,y) + iB_x(x,y) = 10^{-4} B_1 \sum_{n=0}^{\infty} \left(b_n + ia_n\right) \left(\frac{z}{R_{ref}}\right)^{n-1} \qquad b_1 \equiv 10^4 \tag{3.10}$$

where a reference radius is introduced to have adimensional multipoles. The reference radius has no physical meaning, and it is a selection of a unit of distance to express the relative weight of the field harmonics. Usually the reference radius is taken at 2/3 of the magnet aperture; with this choice, $b_2=b_3$ means that the second and third harmonics are the same at 2/3 of distance to the centre of the magnet

$$b_2 + ia_2 = b_3 + ia_3 \Rightarrow C_2(R_{ref}) = C_3(R_{ref}) \ . \tag{3.11}$$

Using a smaller reference radius give smaller multipole coefficients; to be more quantitative, using half of the reference radius will give the same dipolar normalized multipole, half of the quadrupolar

normalized multipole b_2, ¼ of the sextupolar normalized multipole b_3, … and so on. Therefore when comparing the field quality in different magnets via normalized multipoles one has to be sure that the ratio between reference radius and magnet aperture is the same, otherwise the magnet using a smaller reference radius will appear to have a much better field quality.

3.2 A digression on perturbative expansions: the case of 1/(1-x)

Let us consider the function expressed by the power series

$$\sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \ldots \tag{3.12}$$

One can observe that

$$\frac{\sum_{n=0}^{\infty} x^n - 1}{x} = \frac{1 + x + x^2 + x^3 + \ldots - 1}{x} = 1 + x + x^2 + x^3 + \ldots = \sum_{n=0}^{\infty} x^n \tag{3.13}$$

and therefore

$$\sum_{n=0}^{\infty} x^n - 1 = x \sum_{n=0}^{\infty} x^n \qquad\qquad \sum_{n=0}^{\infty} x^n = \frac{1}{1-x}\,. \tag{3.14}$$

Note that if x=1/2, the right hand side gives

$$\frac{1}{1-1/2} = \frac{1}{1/2} = 2 \tag{3.15}$$

and the left hand side is

$$\sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^n = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \ldots = 2. \tag{3.16}$$

Now, let us imagine that we are a kid who has not yet been taugth how to make quotients, but who masters well non-integer numbers, sums and multiplications. Having to compute the function 1/(1-x), in x=0.5, she uses the power series (3.16), and being quite motivated she stops the computation at the fourth term

$$\sum_{n=0}^{\infty} \left(\frac{1}{2}\right)^n \approx 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = 1.875, \tag{3.17}$$

getting the correct result with an error smaller than 10%. Perturbative expansions are a fundamental tool addressing the problem of having an equation to solve (on our case, computing 1/(1-x)), without having the required mathematical tools (in our case, the concept of quotient) and finding an approximated solution on the base of known tools (in our case, sums and multiplications).

For instance, let us consider that the function 1/(1-x) is the solution of the set of differential equations given by Newton laws for three bodies: the sun, Jupiter, and an unknown planet at a distance larger than Saturn. Indeed, the equations do not have a solution that can be written analytically according to our present knowledge. In the 19th century, astronomers and mathematicians develop a pertubative approach to find approximate solutions; matching the approximate solution to the observed trajectory of Jupiter, they work out the orbital parameters of a seventh planet. Powerful telescopes are pointed in the direction suggested by the computations, and Neptune is discovered by J. Galle in 1846 [4].

Actually, this story has an interesting counterpart in the search for planet X by P. Lowell and in the accidental discovery of Pluto in 1931 [5].

Perturbative expansions are very precise even if we compute only a few orders (as the kid of our example) in the case of a small perturbative parameter. In our case the perturbative parameter is the absolute value of x, and therefore the use of the trunacted series becomes more and more precise when x approaches small values. For instance, if x=0.01 one has

$$\frac{1}{1-0.01} = 1.010101... \tag{3.18}$$

and using the series, with two terms one has a one per mil precision:

$$\frac{1}{1-0.01} \approx 1+0.01 = 1.01 \tag{3.19}$$

and with three terms one has a precision of 1 part par million:

$$\frac{1}{1-0.01} \approx 1+0.01+0.0001 = 1.0101. \tag{3.20}$$

In the case of Neptune discovery, the perturbative parameter is the relavitely small mass of Jupiter with respect to the Sun.

Indeed, when we replace the right hand side of the second equation in (3.14) with the left hand side, we gain a way to solve "impossible" problems with our limited tools, but we also lose a precious part of information. Let us consider again the case of 1/(1-x), this time estimated in x=-1: over there, the function is perfectly well defined:

$$\frac{1}{1-(-1)} = \frac{1}{2}, \tag{3.21}$$

but the series gives

$$\sum_{n=0}^{\infty}\left(-1\right)^n = 1-1+1-1+1-1+... = ?. \tag{3.22}$$

In the case of x=-2 one has

$$\frac{1}{1-(-2)} = \frac{1}{3} \tag{3.23}$$

but the series gives

$$. \sum_{n=0}^{\infty}\left(-2\right)^n = 1-2+4-8+16-32+... = ?. \tag{3.24}$$

Using the power series, we lose the value of the function 1/(1-x) outside the convergence domain of, which corresponds to numbers whose absolute value are smaller than 1. Therefore equation (3.14) is wrong and we should have written

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x} \quad x \in D, \;\; D = \left\{\left|x\right| < 1\right\}. \tag{3.25}$$

If making a pertubative expansion can be straightforward, finding the validity limits (i.e. the domain D) is usually a much more difficult exercise. In many field of physics we face situations in which the exact equations are replaced by power series, and then are applied outside the validity limits as in

(3.21) and (3.23), giving divergent series. If one is so clever to recognize the original function from the pertubative expansion, one can resum the divergent series, and therefore giving sense to the rather awkward equation

$$\sum_{n=0}^{\infty}(-2)^n = 1-2+4-8+16-32+\ldots = \frac{1}{3}. \tag{3.26}$$

Power series are convergent on a circular domain, and therefore the first singular value (as x=1) limits the domain also on the other side (x=-1) where the original function 1/(1-x) has no singularities. Given a distribution of current lines to produce a multipolar component (for instance the quadrupole of figure 3.3, centre and right), the validity limit of the multipolar expansion is the circle whose radius is the closer distance of any current line. Therefore, never use the multipolar expansion outside this validity limit, and in particular to estimate the field inside the coil. For further reading we refer to the first chapters of [6], one of the main references in the topic of divergent series.

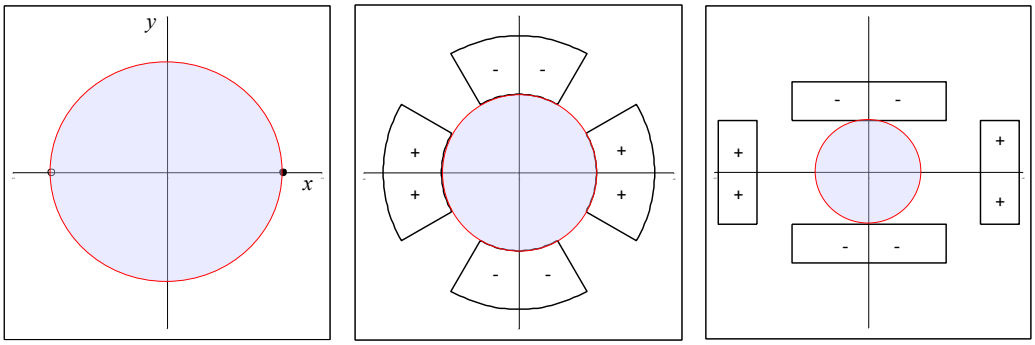


Fig. 3.3: Convergence radius of the power series of 1/(1-x), defined by the first singularity (black spot) (left); convergence radius of multipolar expansion in a sector coil (centre) and in a block coil with rectangular aperture (right).

3.3 Field of an infinite current line and of a current loop

The field given by an infinitesimal element of current (see Fig. 3.4, left) is given by

$$d\vec{B}(\vec{r}) = \frac{\mu_0 \vec{I} \times \vec{r}}{4\pi |r|^3}. \tag{3.27}$$

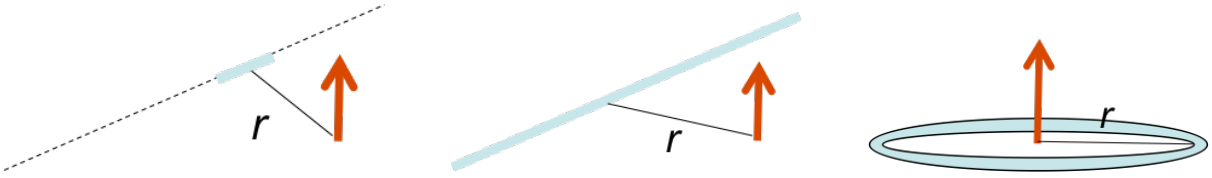


Fig. 3.4: Field given by an element of current line (left), by in infinite straight current line (centre) and by a current loop.

The expression is given for the international system, with distances in m, current in A, and the permeability constant μ_0=4π10^{-7} T m/A. For an infinite straight current line (see Fig. 3.4, centre) the integral gives

$$\vec{B}(\vec{r}) = \frac{\mu_0 I}{4\pi}\int_{-\infty}^{\infty}\frac{ds}{s^2+r^2} = \frac{\mu_0 I}{4\pi r}\int_{-\infty}^{\infty}\frac{d\left(s/r\right)}{\left(s/r\right)^2+1} = \frac{\mu_0 I}{4\pi r}\int_{-\infty}^{\infty}\frac{dy}{y^2+1} = \frac{\mu_0 I}{2\pi r} \tag{3.28}$$

and in the case of a current loop (see Fig. 3.4, right), the field in the centre of the loop is given by

$$B(r) = \frac{\mu_0 I}{4\pi}\int_0^{2\pi r}\frac{ds}{r^2} = \frac{\mu_0 I}{2r} \tag{3.29}$$

These equations are known as Biot-Savart laws.

Fig. 3.5: Felix Savart 1791-1841, French (left), and Jean-Baptiste Biot 1774-1862, French (right)

3.4 Field harmonics of an infinite current line

The field in (x,y) given by in infinite current line in the position (x_0, y_0) (see Fig. 3.6) has the following components:

$$B_x = -\frac{\mu_0 I}{2\pi}\frac{y-y_0}{\left(x-x_0\right)^2+\left(y-y_0\right)^2} \qquad B_y = \frac{\mu_0 I}{2\pi}\frac{x-x_0}{\left(x-x_0\right)^2+\left(y-y_0\right)^2} \tag{3.30}$$

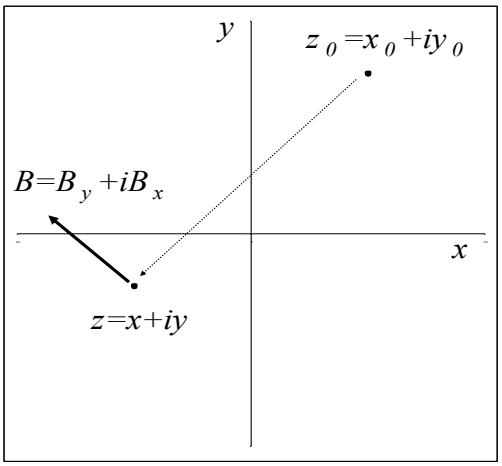


Fig. 3.6: Field in the position (x,y) by an infinite current line in (x_0,y_0) perpendicular to the (x,y) plane

Therefore using the complex notation one has

$$B_y + iB_x = \frac{\mu_0 I}{2\pi} \frac{\left(x - x_0\right) - i\left(y - y_0\right)}{\left(x - x_0\right)^2 + \left(y - y_0\right)^2}. \tag{3.31}$$

Since for a complex number

$$\frac{a - ib}{a^2 + b^2} = \frac{a - ib}{\left(a + ib\right)\left(a - ib\right)} = \frac{1}{a + ib}, \tag{3.32}$$

one can rewrite (3.31) as

$$B_y + iB_x = \frac{\mu_0 I}{2\pi} \frac{1}{\left(x - x_0\right) + i\left(y - y_0\right)} = \frac{\mu_0 I}{2\pi} \frac{1}{z - z_0}. \tag{3.33}$$

We now use what learnt in section 3.2:

$$B_y + iB_x = -\frac{\mu_0 I}{2\pi} \frac{1}{z_0} \frac{1}{1 - z / z_0} = -\frac{\mu_0 I}{2\pi z_0} \sum_{n=0}^{\infty} \left(\frac{z}{z_0}\right)^n = -\frac{\mu_0 I}{2\pi z_0} \sum_{n=1}^{\infty} \left(\frac{z}{z_0}\right)^{n-1} \tag{3.34}$$

and comparing to the definition of normalized multipoles (3.10)

$$B_y(x,y) + iB_x(x,y) = 10^{-4} B_1 \sum_{n=1}^{\infty} \left(b_n + ia_n\right) \left(\frac{z}{R_{ref}}\right)^{n-1} \tag{3.35}$$

we finally obtain

$$\left(b_n + ia_n\right) = -\frac{10^4}{B_1} \frac{\mu_0 I}{2\pi R_{ref}} \left(\frac{R_{ref}}{z_0}\right)^n. \tag{3.36}$$

The above expression shows that the normalized multipoles decay as a power law (see Fig. 3.7) and the slope in the semi-logarithmic plot is given by the logarithm of the ratio between the reference radius and the distance to the current line:

$$\ln\left|b_n + ia_n\right| = \ln\left(\frac{10^4}{B_1} \frac{\mu_0 I}{2\pi R_{ref}}\right) + n \ln\left(\frac{R_{ref}}{z_0}\right). \tag{3.37}$$

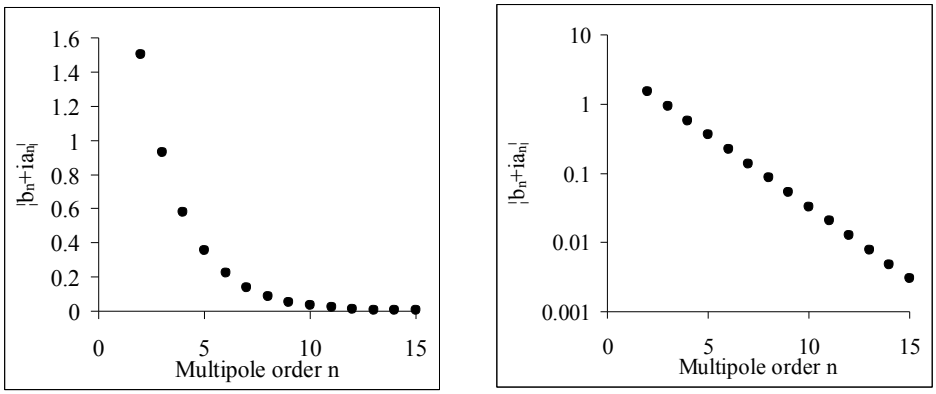


Fig. 3.7: Decay of normalized multipoles in linear scale (centre) and in semilogarithmic scale (right).

3.5 Integral multipoles

We have shown that the multipolar expansion is valid locally in case of absence of longitudinal components (see Eq. 3.4). Multipolar expansions can be also defined as an integral over a given length of the magnet: one can prove that the condition necessary to apply the expansion is that the beginning and the end of the interval is in a region where there is no variation of field (see Fig. 3.8). Therefore, when measuring the field in the coil ends, one can use the multipolar expansion if one end of the measuring device is well inside the magnet (in the region where the decay due to the presence of the end is absent) and the other end shall be well out of the magnet. Any measuring position is characterized by a field B_1^k, and multipoles (b_n^k, a_n^k). The integral mutipoles will be defined as the average of the multipoles in each position weighted via the main field

$$b_n \equiv \frac{\sum_{k=1}^{N} B_1^k b_n^k}{\sum_{k=1}^{N} B_1^k} \qquad a_n \equiv \frac{\sum_{k=1}^{N} B_1^k a_n^k}{\sum_{k=1}^{N} B_1^k} \tag{3.38}$$

It is also customary to define a central field, given by the average of the field in the central positions (excluding the first and the last where one has the coil ends)

$$B_1 \equiv \frac{1}{N-2} \sum_{k=2}^{N-1} B_1^k , \tag{3.39}$$

and finally one defines the magnetic length l_m as the length of the magnet if the field decay in the ends were sharp as a theta function (see Fig. 3.8, right)

$$l_m \equiv l_c \frac{\sum_{k=1}^{N} B_1^k}{B_1} \tag{3.40}$$

where l_c is the length of the measuring coils.

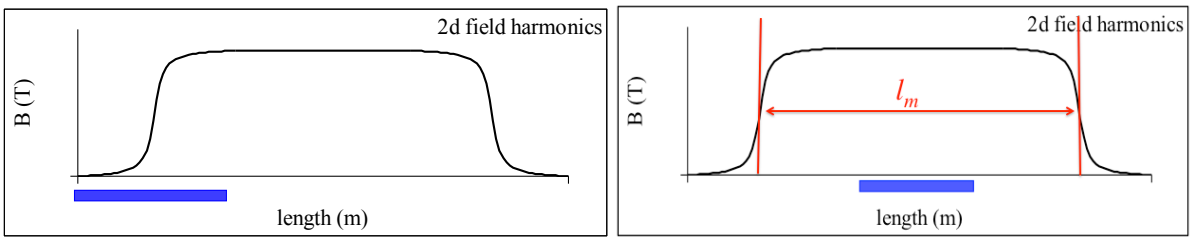


Fig. 3.8: Possible positions of the rotating coil device (in blue) to obtain integral field that can be expressed with multipolar expansion in the case of a dipole; the continuous line is the value of the central magnetic field. Left part: measure in coil heads; right part: measure in the centre, and indication of the magnetic length.

3.6 Applications of multipole decay

We will outline here two powerful applications of the multipole decay. The first one is to use Eq. (3.37) to estimate the precision of magnetic measurements; long magnets are usually measured by several consecutive rotating coils that cover the full length of the magnet. The standard deviation of the measured multipoles along the magnet axis (i.e., of the different measurements in each rotating coils) have to follow the multipolar decay, with a slope corresponding to (3.37); this means that if we

use a reference radius as 2/3 of the aperture radius, every next order should be reduced by a factor 2/3: if $\sigma(b_3)$=1, $\sigma(b_4)$~0.66, $\sigma(b_5)$~0.44, … In practice, every 5-6 orders one loses one order of magnitude in the multipole. This is well seen in the standard deviation of the magnetic measurements along the axis (see Fig. 3.9). However in the figure one can see that around order 25 the multipoles saturate at values of the order of 10^{-4}: this is the precision of our measuring system.

To summarize, plotting the standard deviation of the measured multipoles along the magnet axis in semilogarithmic scale and making a linear regression allows to verify the consistency of the measurements (the slope must be compatible with the ratio between the reference radius and the aperture radius) and to estimate the precision of the measuring system (10^{-4} in the case shown in Fig. 3.9 for the multipoles, i.e. an amazing 10^{-8} of the main component).

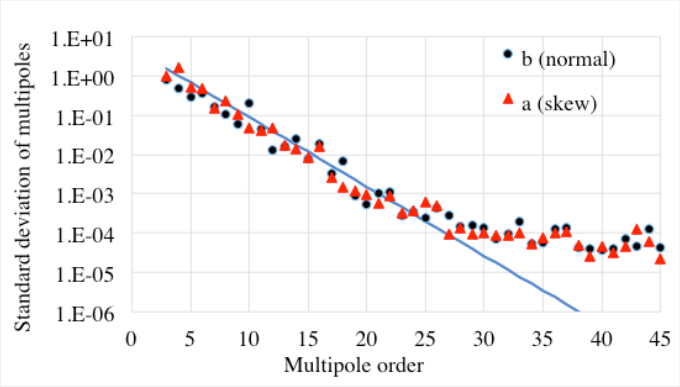


Fig. 3.9: Standard deviation of normalized multipoles measured in different positions along the magnet axis (case of the HL-LHC quadrupole MQXF, courtesy of L. Fiscarelli).

The second powerful tool is the detection of field anomalies and its attribution to misplaced current lines, finding out the distance of the source of field anomaly, i.e. the assembly or the design error. Here we will give two examples for the LHC dipole [7], whose active part is made by two layers of a 15 mm thick cable. All 15-m-long magnets had room temperature magnetic measurements with a 750 mm long rotating coil, giving 20 sets of measurement along the magnet axis. The reference radius was set at 17 mm, i.e. slightly lower than the 2/3 choice that would give 18.67 mm. Field anomalies appear as off-range values in one or more consecutive positions: their value, after substraction of the average of the normal positions, are plotted in semilogarithmic scale and the slope gives an estimate of the distance d of the assembly error from the magnet centre via the equation

$$d \cong \frac{R_{ref}}{\exp(k)}, \tag{3.41}$$

where k is the coefficient of the logarithmic fit (note that a k=0, i.e. the same impact on all multipoles, implies an error on the reference radius).

The first example is showing a field anomaly in the middle of the magnet; when plotting the diffference between measurements in this field position and the average of the other positions, this quantity decays with a slope corresponding to exp(-1.273)=0.28 (see Fig. 3.10, left). Note that the linear regression in semilogarithmic scale (i.e. what is done in Excel via the option exponential fit) has to be done on multipoles up to order 8, since at higher orders we see the plateau related to the precision of the measurements previously discussed. Therefore the anomaly is located at ~17/0.28 mm= 61 mm, i.e. just outside the outer layer. Disassembly showed the presence of a double 0.5 mm thick coil protection sheet between the outer layer and the collars (see Fig. 3.10, right).

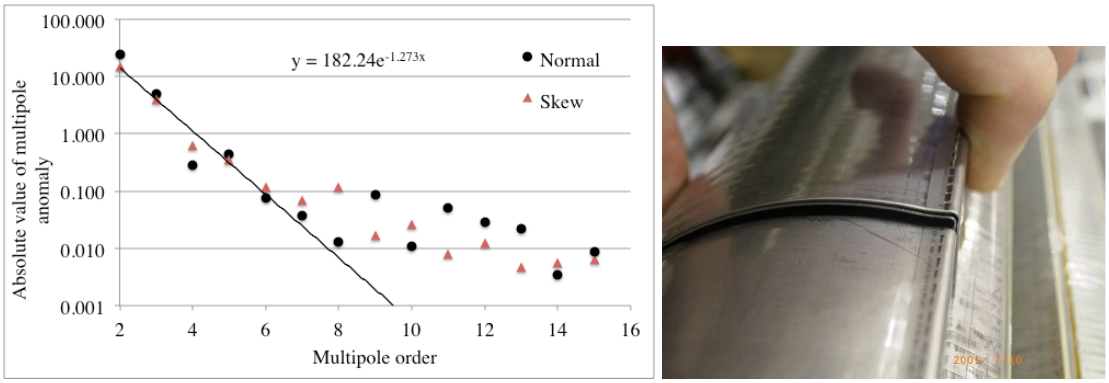


Fig. 3.10: Field anomaly of LHC dipole 2002 in semilogarithmic scale, and linear regression made on orders 2 to 8 (right); double coil protection sheet found during disassembly (right).

The second case we discuss is rather puzzling, since the multipole decay of the anomaly is extremely slow, pointing out to a error source at a distance smaller than the inner edge of the coil. As shown in Fig. 3.11, the slope of the exponential with was location the error at 17/exp(-0.358)~24.3 mm for the normal multipoles, and 17/exp(-0.415)~25.7 mm for the skew multipoles, in both cases well inside the magnet aperture of 28 mm. After some futher analysis, it was shown that the error was not related to the coil manufacturing, but the source was the cold bore, a stainless steel tube inside the magnet that separates the LHe circulating in the coils from the vacuum where the beam circulates. Its radius was 26 mm, and the cold bore was found to have had a non-conform stainless steel, with inclusions of ferromagnetic material.

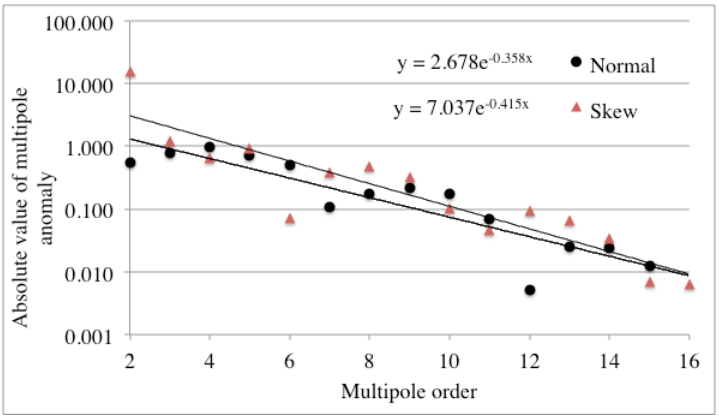


Fig. 3.11: Cross-section of the LHC coil (left) and field anomaly of LHC dipole 1251 in semilogarithmic scale, and linear regression made on normal and skew multipoles.

3.6 Specifications of correctors: the LHC case

We will conclude this chapter about multipolar expansions with an attempt of claryfing how to specify the corrector magnets and what are the requirements imposed by beam dynamics. This interdisciplinary subject is rarely covered in the literature in a systematic way; since there is a strong dependence on the type of lattice and on the features of the accelerator, here we will address this topic via a discussion of the LHC case. With this section we conclude the review of the requirements on the accelerator magnets (main magnets in the arcs in Chapter 1, main magnet in the interaction regions in Chapter 2, and correctors here) imposed by beam dynamics.

The LHC arc [8], as already discussed in Chapter 1, has a main dipole giving 8.3 T over 14.3 m length, for a total integrated field of 119 T m, and a cell quadrupole of 220 T/m over 3.15 m length, for a total intergrated gradient of 693 T. Integrated fields are also given at the reference radius of

17 mm, and therefore the quadrupole has a integrated field of 693×0.017=11.8 T m. This way of expressing magnet of different multipolar order at the reference radius allows to give a comparison of the impact of each magnet for a particle that is circulating at a distance from the magnet centre equal to the reference radius.

The cell is of alternating sign quadrupoles (focusing and defocusing) spaced by 53 m, with three dipoles between each quadrupole, as discussed in Chapter 1. The layout is shown in Fig. 3.12. Several types of correctors are present.

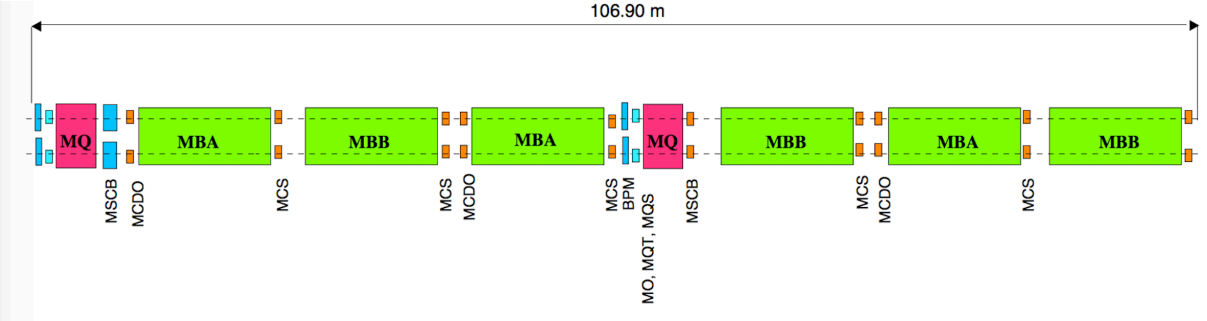


Fig. 3.12: Lay-out of the LHC cell, including the correctors.

Table. 3.1: Parameters of the LHC correctors.

Type	Acronym	Magnetic length (m)	Field (T)	Gradient (T/m)	Sextupole (T/m^2)	Integrated strength at Rref (T m)	Number per cell
Orbit corrector	MCBV, MCBH	0.647	2.93			1.88	1 1
Tuning quadrupole	MQTH, MQTV	0.380		123		0.8	1 1
Skew quadrupole	MQS						
Lattice sextupole	MS	0.369			4430		2
Sextupole spool piece	MCS	0.110			1630	0.052	3

Orbit correctors: in the LHC there are two orbit corrector families MCBV and MCBH, one with horizontal field (called V, since it gives a kick in the vertical direction) placed close to the defocusing quadrupole, i.e. where the beta function in y is larger. Similarly, there is a corrector giving a vertical field (called H, since it gives horizontal kick) placed close to the focusing quadrupole, see Fig. 3.12. Both dipole correctors have a dipole field of 2.93 T, a magnetic length of 0.647 m, providing an integrated field of 1.9 T m, see Table 3.1. This field is 0.3% of the field of six dipoles, that is 8.3×14.3×6=720 T m.

The need of orbit correctors comes from two sources:

- To compensate the dipole kick given by a misalignent of the beam in the cell quadrupole.
- To compensate the variation of the integrated field from dipole to dipole;

For the first source, let us consider a misalignent of ~0.5 mm in the quadrupole: at that distance a gradient of 220 T/m produces a dipolar field of ~0.11 T. Having two quadrupoles of 3.15 m length, this gives over the cell a kick with an integrated field of ~0.7 T m. A 1.9 T m integrated field H/V dipole at each second quadrupole allows the correction a 0.5 mm systematic misaligment of the orbit in the quadrupoles, still leaving more than half of the force free for other scopes.

Tuning quadrupoles: in the LHC there are two families of tuning quadrupoles MQT, with 123 T/m gradient and 0.38 m magnetic length, placed close to each cell quadrupole. The integrated field of the tuning quadrupole at the LHC reference radius of 17 mm is 123×0.38×0.017= 0.80 T· m. This is about 6% of the integrated strenght of one quadrupole 11.80 T· m. The need of tuning quadrupoles comes from the requirement of changing the tune in x and y planes, giving flexibility to machine operation. The minimal amount of change is ±0.5, to be able to span the whole space of resonances: considering that the LHC tune is ~60, this would require tuning quadrupole whose strength is 1.2% of the main quadrupoles. The MQT are five times stronger, and they allow changing the tunes of ~5 integers.

Sextupole spool pieces: in the LHC there are sextupoles correctors MCS placed close to each dipole. These magnets have a 1630 T/m^2 gradient over a 0.11 m magnetic length. The integrated gradient at reference field can be obtained by 1630×0.11×0.017^2= 0.052 T· m. These magnets are mainly used to correct the variation of b_3 in the main dipoles caused by the superconductor magnetization. To convert the sextupole integrated gradient in units of b_3 in the dipole one has to compute the integrated field at reference radius of the sextupolar component in one dipole: one unit of sextupole is 10^{-4} the integrated field of the dipole, i.e. 0.012 T m. Therefore the spool piece can correct 0.052/0.012= 4.35 units of b_3 in the dipoles at high field.

Chromatic sextupoles: in the LHC there are sextupoles correctors MS placed close to each quadrupole. These magnets have a 4430 T/m^2 gradient over a 0.369 m magnetic length, and therefore are 9 times stronger than the spool pieces MCS. Their function is to correct the chromaticity of the lattice. Its computation goes beyond the horizon of these notes.

References

1. L. V. Ahlfors, "Complex analysis: an introduction to the theory of analytic functions of one complex variable", McGraw.Hill, New York (1956)
2. K. H. Mess, P. Schmüser, S. Wolff, "Superconducting Accelerator Magnets", World Scientific, Singapore (1996)
3. A. Jain, "Basic theory of magnets", CERN 98-05 (1998) 1-26
4. N. A. Kollerstrom, "A Neptune Discovery Chronology", University College London (2001)
5. T. Standage, "The Neptune file : a story of Astronomical Rivalry and the pioneers of planet hunting", Walker, New York (2000)
6. G. H. Hardy, "Divergent series", Claredon Press, Oxford (1956)
7. C. Vollinger, E. Todesco, "Identification of Assembly Faults through the Detection of Magnetic Field Anomalies in the Production of the LHC Dipoles", IEEE Trans. Appl. Supercond. 16 (2006) 204-206
8. O. Bruning, et al., "LHC design report" CERN-2004-003 (2004)

Ezio.Todesco@cern.ch

Chapter 4

How to design dipole and quadrupole magnets with current lines

Plan of the chapter
A few coil lay-outs provide a perfect dipolar field; in the first section we describe these configurations, computing the field as a function of the geometrical parameters, and discussing the possible use for magnet design. However, a more powerful model can be used: the approximation based on sector coils. We describe this configuration in Section 4.2, giving the main dependence of the field on the coil width, and in Section 4.3 we discuss how to optimize field quality with few wedges at the level required by accelerator magnets. A system of equations can be derived to solve the field quality optimization, and this allows having an understanding of the reasons for the layout of several dipole cross-sections. In Section 4.4 we show how the sector coil can be used to benchmark the efficiency of any coil layout, making comparisons to the ideal layouts based on intersecting ellipses; moreover we introduce the most optimized layout found with numerical methods. The efficiency of block and double helix layouts are discussed in section 4.5. In Section 4.6 we consider the field quality optimization for a quadrupole, showing that the equations and the solutions can be derived by a simple rescaling of the dipole case. We conclude this chapter with some comments on how to carry out field quality optimization.

4.1 Lay-outs generating perfect dipole fields

In an infinte solenoid of coil width w and overall current density j (see Fig. 4.1), the magnetic field is

$$B = \mu_0 j w \,. \tag{4.1}$$

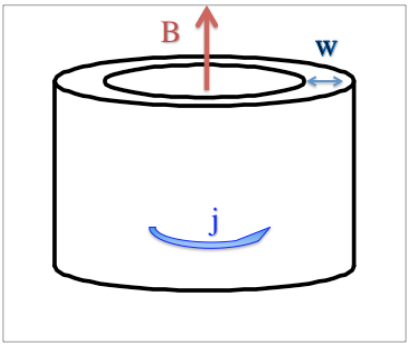


Fig. 4.1: Solenoid configuration, winding width, and field direction.

Note that by overall current density we denote the current density in the insulated coil. Using the practical units of mm for the coil width and A/mm^2 for the current density, one has

$$B\left[\mathrm{T}\right] = 4\pi 10^{-4}\, j\left[\mathrm{A/mm^2}\right] w\left[\mathrm{mm}\right] \approx 0.00126\, j\left[\mathrm{A/mm^2}\right] w\left[\mathrm{mm}\right]; \tag{4.2}$$

this implies that each 10 mm of coil width with 400 A/mm^2 overall current density give ~5 T.

Accelerator dipoles have to provide a field that is not parallel to the beam direction, but perpendicular to it. Let us start with the configuration of two walls of current, of width w and distance $2r$ (see Fig. 4.2). Integrating the contribution of the current lines, one finds

$$B = 2\frac{\mu_0}{2\pi}\int_r^{r+w}\int_{-\infty}^{\infty}\frac{jx}{x^2+y^2}\,dy\,dx = \frac{\mu_0 j}{\pi}\int_r^{r+w} x\int_{-\infty}^{\infty}\frac{dy}{x^2+y^2}\,dx = \frac{\mu_0 j}{\pi}\int_r^{r+w}\int_{-\infty}^{\infty}\frac{dz}{1+z^2}\,dx = \mu_0 jw\,, \tag{4.3}$$

and the proportionality constant between the field and the width times the current density is as in the solenoid, see Eq. (4.1). One can prove that the field inside the walls is uniform, and outside is zero. This configuration providing a perfect dipolar field is of little interest since the layout has infinite size.

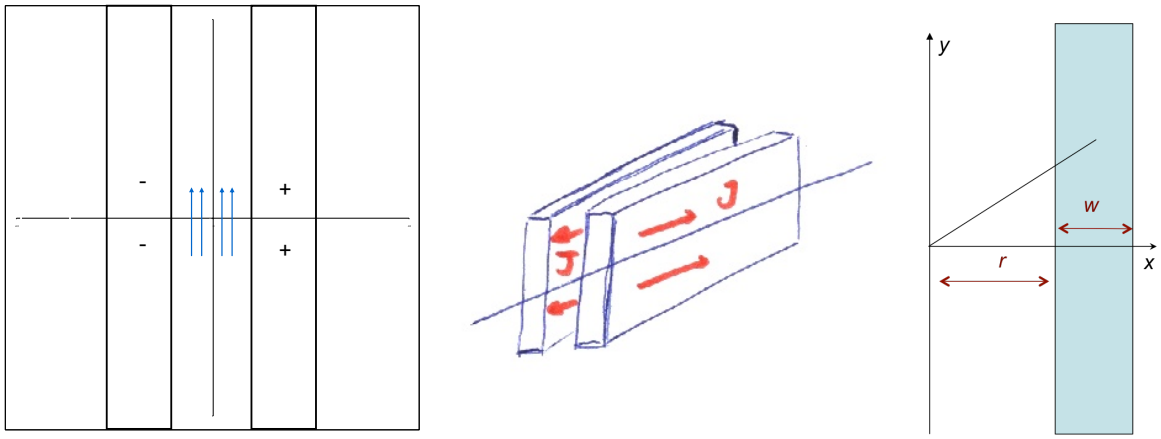


Fig. 4.2: Wall dipole sketch (center), cross-section (left), and coordinates to compute the field (right).

For more realistic cases (i.e. not involvingan infinite coil) we will show that dipoles provide half of the field of a solenoid, for the same coil width w and current density j. Let us consider the case of intersecting ellipses (see Fig. 4.3, left): this is a better solution, since the layout is finite, even though the aperture is not round. Inside the aperture there field is uniform and is given by (see [1], pg 31)

$$B = \mu_0 jw\frac{a_V}{a_H + a_V} \tag{4.4}$$

where w is the width of the coil in the midplane, a_H is the horizontal semiaxis of the ellipse, and a_V is the vertical semiaxis. In the case of an ellipse with zero eccentricity a_H=a_V one has

$$B = \frac{\mu_0 jw}{2} \qquad\qquad B\left[T\right] = 0.000628\, j\left[\mathrm{A/mm^2}\right] w\left[\mathrm{mm}\right]. \tag{4.5}$$

We then consider a cos theta layout, is a ring where the current density depends on the cosine of the angle to the midplane (see Fig. 4.3, right). Here we gain a round aperture, but we lose the uniform current density. Also in this case one has

$$B = 4\frac{\mu_0 j}{2\pi}\int_0^{\pi/2}\int_r^{r+w}\cos^2\theta\, d\theta = 4\frac{\mu_0 j}{2\pi}\frac{\pi}{4}w = \frac{\mu_0 jw}{2} \tag{4.6}$$

where w is the thickness of the ring and j the maximum current density, occurring in the midplane. The cos theta layout has been widely used in the literature to provide analytical estimates of magnet design (see for instance [1,2]). However, a better analytical model can be derived, closer to what is

finally achieved during winding: the sector coil [3,4]. This is what we will discuss it in the next section.

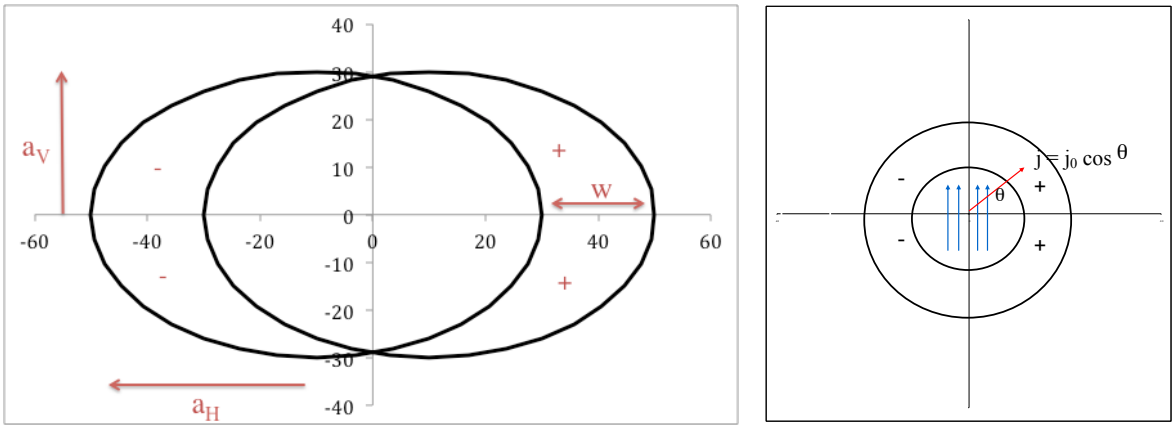


Fig. 4.3: Coil cross-section based on intersecting ellipses (left) and on cos theta current density (right).

We conclude this section by presenting a fourth ideal layout, derived from our starting point (the solenoid): the double helix [5], also know in the literature as tilted solenoid or canted cos theta (CCT). Here the conductor is wound along a solenoid, tilted with an angle α with respect to the beam axis, in two successive layers, with opposite orientation of the angle with respect to the beam direction (see Fig. 4.4). The resulting field inside the aperture is a pure dipole, the aperture is round, and the current density is constant.

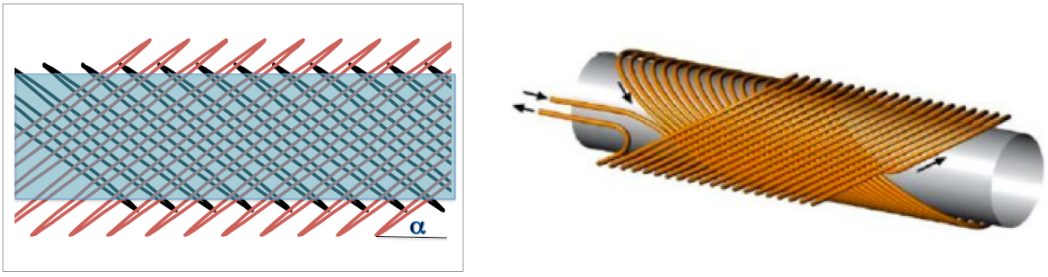


Fig. 4.4: Tilted solenoid configuration, side view with inner winding in black, outer winding in red, and bore tube in blue (left), and 3D view (right).

The conductor path is given by the equations parametrized by t

$$\begin{aligned} x(t) &= r\cos t \\ y(t) &= r\sin t \\ z(t) &= pt + r\tan\alpha\sin t \end{aligned} \qquad \begin{aligned} x(t) &= \left(r+a\right)\cos t \\ y(t) &= \left(r+a\right)\sin t \\ z(t) &= pt - \left(r+a\right)\tan\alpha\sin t \end{aligned} \tag{4.7}$$

where p is the advancement of the spiral along z in each turn. The inner tilted solenoid is wound on theinner radius r, and the outer titled solenoid is wound on the outer radius $(r+a)$, with opposite angle. The rather peculiar shape of the winding can be kept by machining grooves in a supporting structure. The disdvantage of this layout is a lower efficiency, reduced by at least a factor $\cos\alpha$: a fraction of conductor is used to create a solenoidal field that is canceled by the outer winding. Typical angles range between 15 and 30 degrees. Inn section 4.4 we give a case study to compute the efficiency of this layout in section 4.4.

4.2 Field in a dipole magnets based on sector coils

We will now analyze a sector coil with windings having a constant current density (see Fig. 4.5 for a 60° sector without wedges). Contrary to the wall dipole, to the intersepting ellispses and to the cos theta, this lay-out is the only one that can be manufactured based on a rectangular cable with a slight keystone to follow the sector shape; the only approximation is that for small apertures (order of 50 mm diameter) as used in arc magnets of very high energy accelerators, the required keystone angle becomes too large and therefore wedges have to be used to arrange the cable stacks perpendicular to the magnet aperture. However, we will show in the next section that a few wedges are anyway needed to get an adequate field quality.

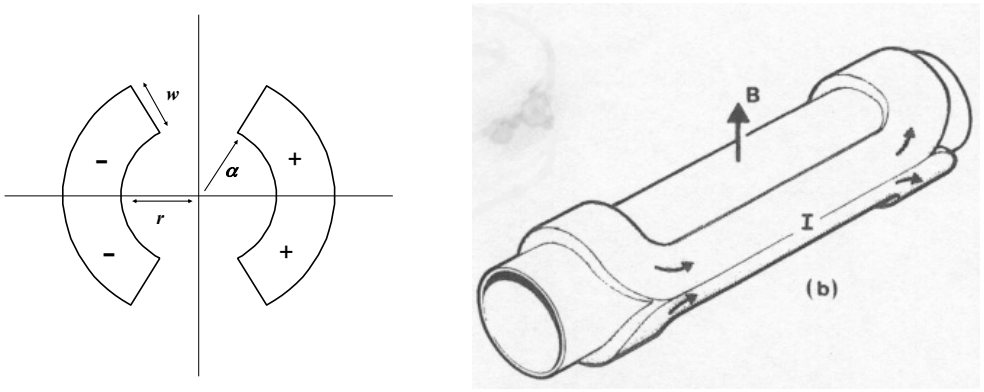


Fig. 4.5: Cross-section of a dipole based on a sector coil (left) and 3D view of the coil lay-out (right).

Let us estimate the main dipolar field given by a sector coil: the contribution to the magnetic field in the centre of the reference system by a line of current placed in (x_0,y_0) is given by (see Eq. 3.34)

$$B = -\frac{\mu_0 I}{2\pi}\mathrm{Re}\left(\frac{1}{z_0}\right) = -\frac{\mu_0 I}{2\pi}\frac{\cos\theta}{|z_0|} \qquad z_0 = x_0 + iy_0. \tag{4.8}$$

Replacing the current with current density

$$I \rightarrow j\rho d\rho d\theta \tag{4.9}$$

and performing the integration over a sector coil of aperture radius r, coil width w and angular aperture α (see Fig. 4.5) we obtain

$$B = -2\frac{\mu_0 j}{2\pi}\int_{-\alpha}^{\alpha}\int_{r}^{r+w}\frac{\cos\theta}{\rho}\rho d\rho d\theta = -\frac{2\mu_0 j}{\pi} w\sin\alpha\,. \tag{4.10}$$

For a 60° degrees sector coil, one finds the simple expression

$$B = -\frac{\sqrt{3}}{\pi}\mu_0 j w \tag{4.11}$$

and making use of the practical units, one obtains

$$B = -\frac{\sqrt{3}}{\pi}4\pi\times10^{-7}\times10^{6}\, j\left[\mathrm{A/mm^2}\right]10^{-3} w\left[\mathrm{mm}\right] \approx 6.9\times10^{-4}\, j\left[\mathrm{A/mm^2}\right] w\left[\mathrm{mm}\right]. \tag{4.12}$$

This gives ~10% higher field with respect to the cos theta layout. If we consider an overall current density j of 400 A/mm^2, each 10 mm of coil width gives 2.8 T of dipolar field. Using this rough estimate for the LHC dipoles, having 30 mm coil width, one finds 8.1 T, close to the actual value of 8.3 T.

Summarizing, one finds that

- The dipolar field is proportional to the current density;
- The dipolar field is independent of the aperture radius *r*;
- The dipolar field is proportional to the coil width *w*.

4.3 Harmonic optimization in a dipole magnets based on sector coils: the low cost of field quality

Using the expression computed thanks to the multipolar expansion of a field line, see Chapter 3, Eq. (3.36) and integrating over a sector coil as shown in Fig. 4.5, one finds

$$C_n = -2\frac{\mu_0 j R_{ref}^{n-1}}{2\pi}\int_{-\alpha}^{\alpha}\int_{r}^{r+w}\frac{\exp(-in\theta)}{\rho^n}\rho d\rho d\theta = -\frac{\mu_0 j R_{ref}^{n-1}}{\pi}\int_{-\alpha}^{\alpha}\exp(-in\theta)d\theta\int_{r}^{r+w}\frac{d\rho}{\rho^{n-1}}. \tag{4.13}$$

For the second order harmonic one has a logarithmic dependence

$$C_2 = -\frac{2\mu_0 j R_{ref}}{\pi}\sin(2\alpha)\ln\left(1+\frac{w}{r}\right) \tag{4.14}$$

and for the higher orders one has

$$C_n = -\frac{\mu_0 j R_{ref}^{n-1}}{\pi}\frac{2\sin(n\alpha)}{n}\frac{(r+w)^{2-n}-r^{2-n}}{2-n}. \tag{4.15}$$

Summarizing, one finds that

- The unnormalized field harmonics are proportional to the current density;
- The unnormalized field harmonics of order *n* are proportional to the sinus of *n* times the angle of the sector; therefore they can be set to zero with an selection of the sector angle;
- The unnormalized field harmonics of order *n* are proportional to the inverse of a power *n*-2 of the location of the current lines in the sector; therefore higher and higher order multipoles are less and less affected by the parts of the coil that are far from the centre.

The 60° degrees sector coil cancels the first order harmonic C_3 since sin(3π/3) = 0. To cancel out the order five one needs either 36° or 72° degrees sector coil. Obviously, having only one angle to play with, one can cancel only one harmonic. This is why we have to go for sector coils with wedges.

Let us consider a sector coil with one wedge, see Fig. 4.6; the unnormalized multipoles are given by

$$C_n = -\frac{\mu_0 j R_{ref}^{n-1}}{\pi}\frac{(r+w)^{2-n}-r^{2-n}}{2-n}\frac{2}{n}\left[\sin(n\alpha_3)-\sin(n\alpha_2)+\sin(n\alpha_1)\right] \tag{4.16}$$

and therefore there are solutions that can cancel both order 3 and order 5:

$$\begin{aligned}\sin(3\alpha_3)-\sin(3\alpha_2)+\sin(3\alpha_1)&=0\\ \sin(5\alpha_3)-\sin(5\alpha_2)+\sin(5\alpha_1)&=0\end{aligned}. \tag{4.17}$$

Four solutions with integer angles exist, see Fig. 4.6 right:

- (0°-24°), (36°-60°) this has the lower angular width (total of 48°);
- (0°-36°), (44°-64°) this has the lower angular width of the wedge (8°);

- (0°-48°), (60°-72°) this is widely used by magnet designers;
- (0°-52°),(72°-88°) this is of little practical interest since the winding extends to angles which give very small contribution to vertical field, i.e. is far from being effective in the use of the conductor. Moreover having a winding that extends over nearly 90° poses other challenges in terms of manufacturing and assembly.

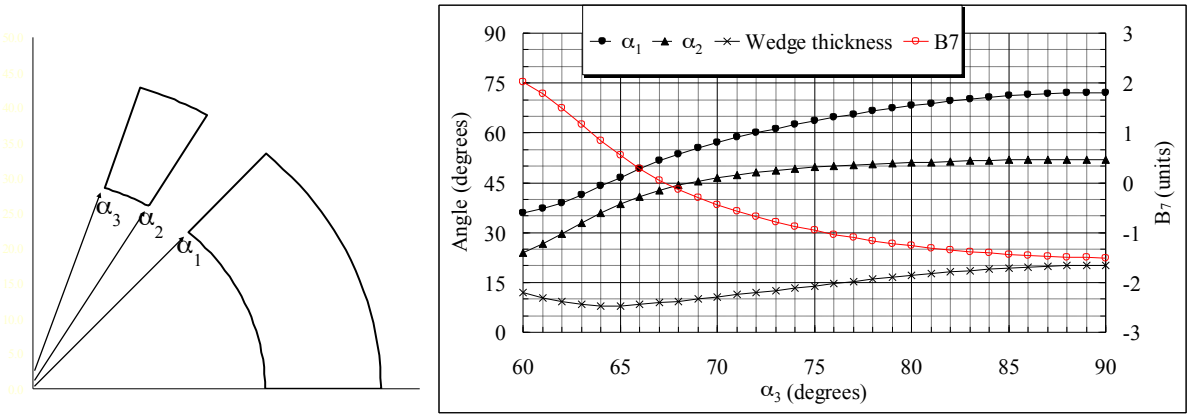


Fig. 4.6: Sector coil with one wedge (left) and solutions of the angles α_1 α_2 as a function of α_3 that set to zero b_3 and b_5 (right).

The values of α_1 and α_2, as a function of α_3, that solve the system (4.18) are given in Fig. 4.6 right, where also the thickness of the wedge is given, and B_7. There is one value that cancels also B_7: it corresponds to the angles ~(0°-43.2°),(53.2°-67.3°); the layout is shown in Fig. 4.7.

Now it is easy to generalize to another wedge, to cancel also B_9 and B_{11}: we have to solve the system

$$\begin{aligned}
\sin(3\alpha_5)-\sin(3\alpha_4)+\sin(3\alpha_3)-\sin(3\alpha_2)+\sin(3\alpha_1)&=0\\
\sin(5\alpha_5)-\sin(5\alpha_4)+\sin(5\alpha_3)-\sin(5\alpha_2)+\sin(5\alpha_1)&=0\\
\sin(7\alpha_5)-\sin(7\alpha_4)+\sin(7\alpha_3)-\sin(7\alpha_2)+\sin(7\alpha_1)&=0\\
\sin(9\alpha_5)-\sin(9\alpha_4)+\sin(9\alpha_3)-\sin(9\alpha_2)+\sin(9\alpha_1)&=0\\
\sin(11\alpha_5)-\sin(11\alpha_4)+\sin(11\alpha_3)-\sin(11\alpha_2)+\sin(11\alpha_1)&=0
\end{aligned} \qquad (4.18)$$

that may appear pretty complex, but that can be solved with numerical tools, and has one solution shown in Fig. 4.8 ~(0°-33.3°),(37.1°-53.1°),(63.4°-71.8°). These simple equations already give a good analytical insight on the angles of the coil selected for the RHIC dipole (see Fig. 4.8), which is a large aperture (80 mm diameter) magnet based on a 10-mm-width sector coil [6]. The angles of the wedges are not far from the analytical solution; moreover the first block is split in two to be able to recover a prependicularity between cables and magnet aperture.

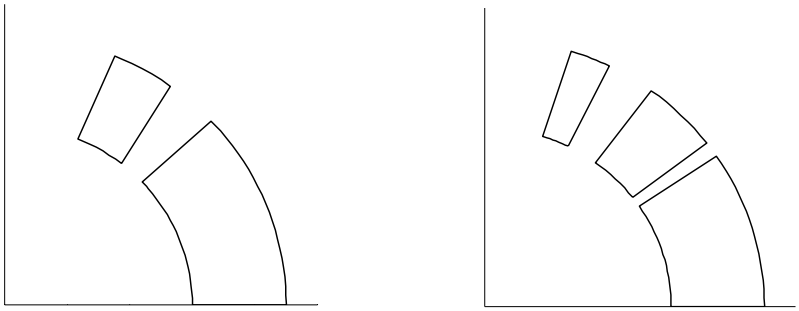


Fig. 4.7: 1-wedge sector canceling up to b_7 (left) and 2-wedges sector canceling up to b_{11} (right).

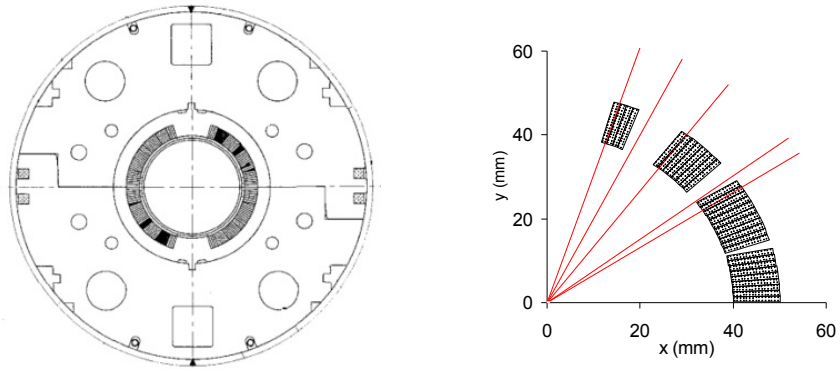


Fig. 4.8: RHIC dipole magnet cross-section (left) and cross-section of the coil (one quarter shown, right) with red lines indicating the angles of coil shown in Fig. 4.7, right.

For a two-layer magnet things become a bit more complex, since there is a dependency also on radial parameters which is not present in the case of a one layer coil. Let us carry out the computations for a two layer coil, without wedges; we will have two free parameters, namely the angle of the first layer and the angle of the second one (see Fig. 4.9). In this case, the integration gives

$$C_n = -\frac{2\mu_0 j R_{ref}^{n-1}}{\pi n(2-n)}\left(\left[\frac{1}{(r+w)^{n-2}} - \frac{1}{r^{n-2}}\right]\sin(n\alpha_1) + \left[\frac{1}{(r+2w)^{n-2}} - \frac{1}{(r+w)^{n-2}}\right]\sin(n\alpha_2)\right) \quad (4.19)$$

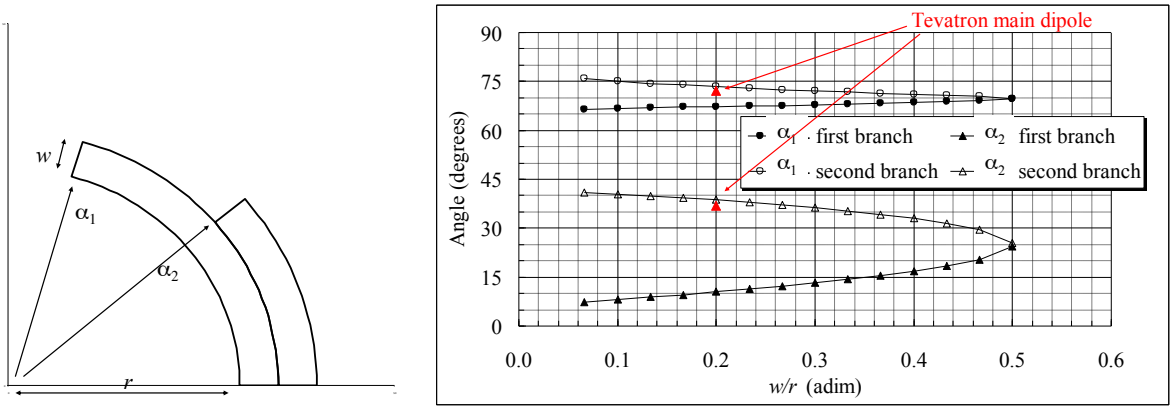


Fig. 4.9: Two-layer coil withhout wedges (left) and solutions setting to zero b_3 and b_5 as a function of the ratio between the layer width and the aperture radius (right)

and the equations to set order B_3 and B_5 to zero are

$$\begin{aligned}&\left[\frac{1}{r+w} - \frac{1}{r}\right]\sin(3\alpha_1) + \left[\frac{1}{r+2w} - \frac{1}{r+w}\right]\sin(3\alpha_2)\\ &\left[\frac{1}{(r+w)^3} - \frac{1}{r^3}\right]\sin(5\alpha_1) + \left[\frac{1}{(r+2w)^3} - \frac{1}{(r+w)^3}\right]\sin(5\alpha_2)\end{aligned} \quad (4.20)$$

that can be rewritten as

$$\begin{aligned}&\left[\frac{1}{1+w/r} - 1\right]\sin(3\alpha_1) + \left[\frac{1}{1+2w/r} - \frac{1}{1+w/r}\right]\sin(3\alpha_2)\\ &\left[\frac{1}{(1+w/r)^3} - 1\right]\sin(5\alpha_1) + \left[\frac{1}{(1+2w/r)^3} - \frac{1}{(1+w/r)^3}\right]\sin(5\alpha_2)\end{aligned} \quad (4.21)$$

The numerical solutions are shown in Fig. 4.9, right. For $w/r>0.5$ there are no solutions, since the second layer is too far to inflence B_5. Below this limit there are two solutions, the most interesting of which is the upper branch. Note that Tevatron dipole (see 4.10) has r=43 mm and w=8, i.e. w/r=0.2, and the analytical solution on the upper branch corresponds to ~73° and ~37° for inner and outer layer respectively [7], in good agreement with the the actual design of the magnet coil, see Fig. 4.10.

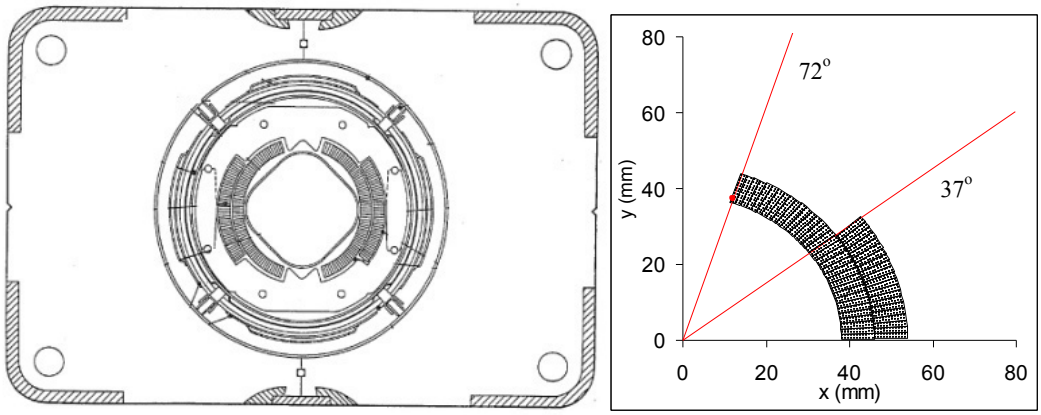


Fig. 4.10: Tevatron dipole magnet cross-section (left) and cross-section of the coil (one quarter shown, right) with angles shown.

In Table 4.1 we compare the proportionality factors between field and coil width times current density. As expected, layouts with larger angular width (i.e. more conductor) provide more field. Having a closer look at the comparison between the (0°-60°) and the (0°-48°),(60°-72°), we see that they have the same angular width (same quantity of conductor) and therefore a direct comparison is possible; setting also b_5 to zero as in the (0°-48°),(60°-72°) solution costs about 5% in terms of field obtained by the same quantity of conductor. The sector coils setting to zero up to order 7 and up to order 11 have a 3% lower constant, but with a 4.5% lower angular width. Note the (0°-52°),(72°-88°) lay-out: the B/jw is the same of the (0°-48°),(60°-72°), layout, but using a 10% larger angle as conductors at very large angles give smaller contribution to the main field.

Table 4.1: Ratio between field and coil width times current density, angular widths and field quality.

Lay-out	B/jw (T mm/A)	Angular width	Field quality
Solenoid	12.56×10^{-4}		all b_n=0
Wall dipole	12.56×10^{-4}		all b_n=0
Cos theta	6.28×10^{-4}		all b_n=0
Intersecting circles	6.28×10^{-4}		all b_n=0
Sector coil (0°-60°)	6.93×10^{-4}	60°	b_3=0
Sector coil (0°-24°),(36°-60°)	5.48×10^{-4}	48°	b_n=0 up to n=5
Sector coil (0°-36°),(44°-64°)	6.34×10^{-4}	56°	b_n=0 up to n=5
Sector coil (0°-48°),(60°-72°)	6.63×10^{-4}	60°	b_n=0 up to n=5
Sector coil (0°-52°),(72°-88°)	6.69×10^{-4}	66°	b_n=0 up to n=5
Sector coil ~(0°-43.2°),(53.2°-67.3°)	6.45×10^{-4}	~57.3°	b_n=0 up to n=7
Sector coil ~(0°-33.3°),(37.1°-53.1°),(63.4°-71.8°)	6.41×10^{-4}	~57.5°	b_n=0 up to n=11

4.4 Electromagnetic efficiency of coil layouts

In the previous section we compared the electromagnetic efficiency of a coil layout computing the ratio between the field and the current density times the coil width $B/(jw)$, see Table 4.1, but we also showed that layouts with same coil widths can have different quantities of conductor, according to the angular width of the sector and of the wedges; for this reason, we propose to define an equivalent coil width by computing the width of a 60° sector coil whose area is the same of the area of the layout that we are considering. This quantity will be used to define a coil efficiancy, i.e. the ability of a given quantity of conductor to produce field given an overall current density; the equivalent coil width concept also allows to compute coil efficiencies for other coil layouts, as block of tilted solenoids [8]. Since a 60° sector coil has a cross-sectional surface A given by

$$A = \frac{2\pi}{3}\left[\left(r+w\right)^2 - r^2\right], \tag{4.22}$$

one can invert the above relation according to

$$\frac{3A}{2\pi r^2} = \left[\left(1+\frac{w}{r}\right)^2 - 1\right] \qquad \sqrt{1+\frac{3A}{2\pi r^2}} - 1 = \frac{w}{r}, \tag{4.23}$$

and for any coil with cross-sectional surface A we can define the equivalent coil width as

$$w_{eq} \equiv r\left(\sqrt{1+\frac{3A}{2\pi r^2}} - 1\right) \tag{4.24}$$

i.e. the width of the 60° sector coil having the same aperture and the same quantity of conductor. Once we have computed this parameter, we can define the coil efficiency as

$$\gamma_c \equiv \frac{B}{jw_{eq}}. \tag{4.25}$$

The coil efficiency for a coil with one wedge, total angular width 60°, setting to zero first two order harmonics, and having a wedge between 48° and 60° is 6.63×10^{-4} T mm/A: this is a good reference value to benchmark the other layouts. Let us start the exercise with the layouts given in the section 4.1. The wall dipole cannot be treated in this way, since it requires an infinite coil. Considering the intersecting ellipses, we have to compute the area of the coil, find the equivalent coil width and estimate the coefficient γ_c using the relation for the magnetic field (4.4). The layout depends on two parameters: the ratio between the axis and the width of the coil in the midplane. Results are shown in Fig. 4.11, left, for an aperture radius of 25 mm: the intersecting ellipses layout can provide values between 6.5×10^{-4} and 7.0×10^{-4} T mm/A, i.e. up to 6% larger than our reference value. The price to pay for this larger efficiency is a larger coil width in the midplane (see case in Fig. 4.12, refering to 40 mm equivalent coil width, and giving a 50 mm coil width in the midplane) and the presence of conductor at angles close to 90°, not trivial to integrate in a mechanical structure.

The second case is the cosθ layout. Here in principle one should take the whole ring as conductor area, but the current density is not uniform; therefore, we assume that the full current density is gathered in a sector of angle α,

$$\int_0^{\pi} \cos\theta\, d\theta = \int_0^{\alpha} d\theta \tag{4.26}$$

thus giving α=1 (in radiants), i.e. a 5% smaller angular width of $\pi/3$. Estimating the efficiency for a few coil widths, and a aperture radius of 25 mm, we find values around 6.5×10^{-4} T mm/A, i.e. very close to our reference value of a sector coil with one wedge.

Using the same approach, we can estimate the efficiencies of the layouts whose total angular width is not exactly 60°; here the efficiency computation weakly depends on the ratio between the coil width and the aperture radius – we present in Table 4.2 few cases. Values are very similar, and for the cases with field quality correction up to order 7 or 11 the efficiency is between 6.6 and 6.7×10^{-4} T mm/A.

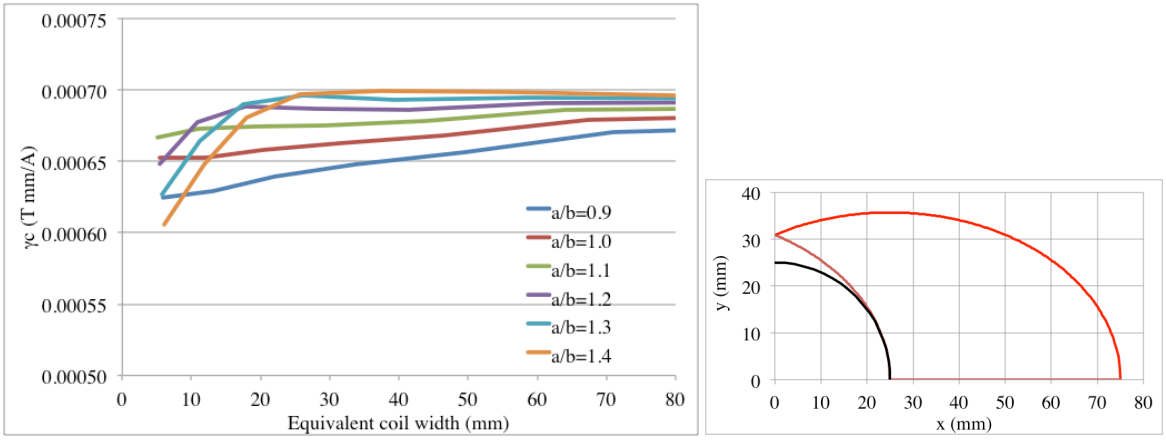


Fig. 4.11: Coil efficiency for intersecting ellipses lay-outs (left) and optimized case for 40 mm coil width and 25 mm aperture.

Table 4.2: Coil efficiencies as defined in (4.25)

Coil efficiency (T mm/A)	*r=25 mm w=10 mm*	*r=25 mm w=30 mm*	*r=25 mm w=50 mm*
Cos theta	6.53×10^{-4}	6.49×10^{-4}	6.48×10^{-4}
Intersecting circles (best case)	6.77×10^{-4}	6.99×10^{-4}	6.98×10^{-4}
Sector coil (0°-60°)	6.93×10^{-4}	6.93×10^{-4}	6.93×10^{-4}
Sector coil (0°-24°),(36°-60°)	6.65×10^{-4}	6.46×10^{-4}	6.37×10^{-4}
Sector coil (0°-36°),(44°-64°)	6.73×10^{-4}	6.67×10^{-4}	6.64×10^{-4}
Sector coil (0°-48°),(60°-72°)	6.63×10^{-4}	6.63×10^{-4}	6.63×10^{-4}
Sector coil ~(0°-43.2°),(53.2°-67.3°)	6.71×10^{-4}	6.67×10^{-4}	6.65×10^{-4}
Sector coil ~(0°-33.3°),(37.1°-53.1°),(63.4°-71.8°)	6.65×10^{-4}	6.61×10^{-4}	6.59×10^{-4}
Best efficiency	7.33×10^{-4}	7.20×10^{-4}	7.13×10^{-4}
Tevatron dipole	6.89×10^{-4}		
LHC dipole (removing iron and grading)	6.57×10^{-4}		
MDPTC1 (removing iron and grading)	6.61×10^{-4}		

Is it possible to have a better layout ? A recent work [9] shows that the answer is affirmative. Considering an arbitrary path around a round aperture of radius *r*, and imposing a perfect field quality, one can select the best layouts providing the highest field for a given coil surface and current density. The shapes of the most efficient layouts are intriguing, see Fig. 4.12. As observed in the paper, for smaller coil width the optimal solution is not far from an intersecting ellipses. For large coil widths,

the optimal shape becomes more complex. As for intersecting ellipses or cos theta layout, the configurations shown in Fig. 4.12 are not windable with a rectangular or a trapezoidal cable. However, the plot is extremely interesting since it sets the reference for the maximum efficiency, that ranges from 7.1 and 7.4×10^{-4} T mm/A, i.e. about 10% more than the sector coils (see Table 4.2). We add in the Table also the efficiencies computed for the Tevatron dipole, and for the LHC dipole [10] and the MDPCT1 dipole model [11] (in both cases removing the grading, i.e. placing the same current density in all layers, and the iron; both cases will be treated in Chapter 9). Results range from 6.6 to 6.9×10^{-4} T mm/A: this shows how the sector coil is close to what can be achieved in real dipole coils.

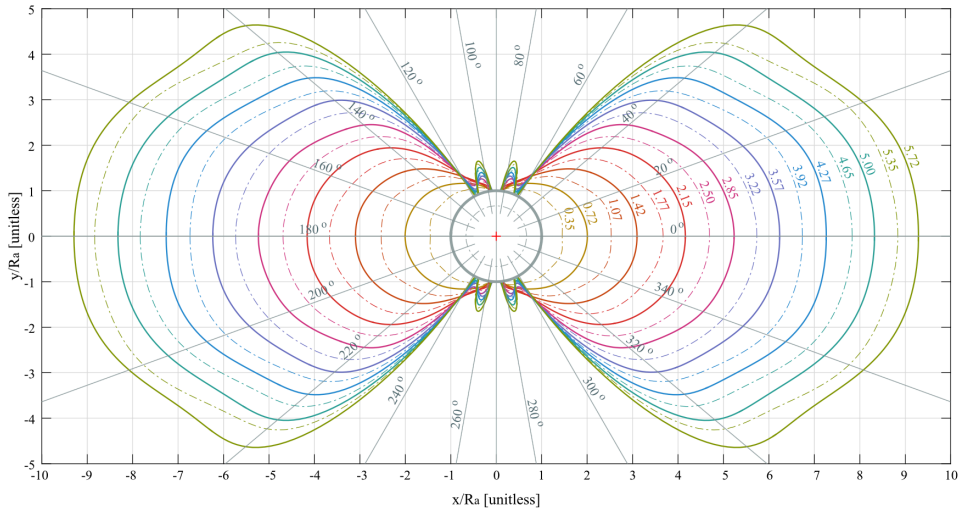


Fig. 4.12: Shapes of the most efficient coil lay-out, from [9]

4.5 Efficiencies of block and double helix layouts

Tevatron and RHIC dipole coils have a width of 10-15 mm, and a ratio between the coil width and the aperture radius of the order of 0.2. Dipoles for higher energy accelerators have a smaller aperture, and possibily also higher field. In the LHC dipole, with a 56 mm aperture diameter and 31 mm coil width based on double layer coil, the ratio w/r is around 1. In this case the second layer has a lower angle, since it is equivalent to place the precious conductor in the midplane at the edge of the coil, or on the top of a closer layer (see Fig 4.13 left): the first configuration is mechanically more stable, and the efficiency is not compromised. The angles of the outer layers provide also a way to steer field quality for the low order multipoles, as done in the Tevatron coil, and they are decoupled to the high order multipoles. The same approach has been used for the four layers dipole MDPCT1 (Fig. 4.13, right).

When the coil width becomes very large with respect to the aperture radius, the natural configuration of the coil tends to a solution whose contour looks like a rectangular coil rather than a sector coil (see Fig. 4.14). This trend is also visible in the optimal layout as shown in Fig. 4.12. This is the configuration that can be achieved in a block coil [13]. Note that compared to sector layouts, block coils are based on a radically different concept:

- the cable is wound perpendicular to the midplane; therefore it is rectangular and in the straight part of the magnet the coil is a flat racetrack;
- there is no arc structure and therefore one needs to have in internal structure to support the upper deck;
- The coil ends are flared to free the space needed to the beam tube.

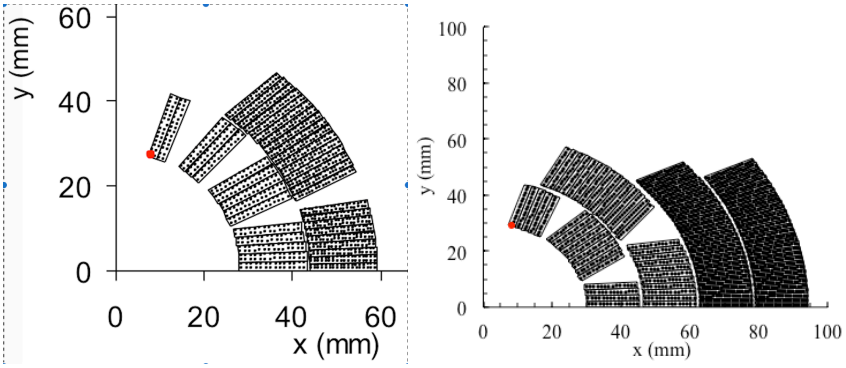


Fig. 4.13: Main LHC dipole cross-section (left), and MDPTC1 dipole cross-section (right).

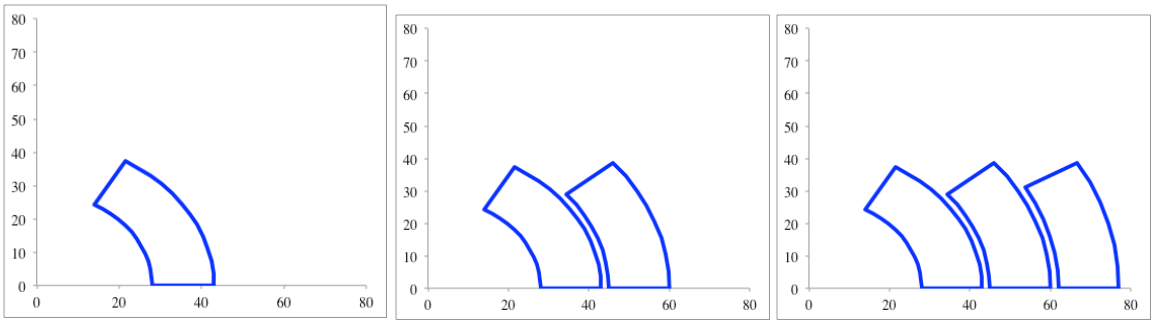


Fig. 4.14: Dipole coils based on sector with additional layers and decreasing angles

The case of the HD2 dipole [13] is shown in Fig. 4.15. The layout is extremely simple, requiring a cable width of the order of the magnet aperture (22 mm cable width for an aperture radius of 17.5 mm). The racetrack coils are flared in the ends to leave space for the beam tube. This innovative design was also generalized to four decks in the case of Fresca2 (see Fig. 4.16), where the cable width is 20 mm, on two layers per deck, for the 50 mm free aperture [14]. An advantage of this layout is that the amount of conductor can be gradually increased (up to a certain extent, given by field quality constraints), whereas in the sector coil it is quantized by the number of layers. A design based on three layers, i.e. relying on a total of three coils per dipole, has been used in RMM, a dipole with 50 mm aperture aiming reaching a bore field above 16 T [15], see Fig. 4.17.

The field quality is guaranteed by having the upper deck of the coil closer to the magnet vertical axis. Good field quality can be achieved without wedges, as in HD2 or in FrescaII layouts: the three free parameters are the cable height, the indent of the upper deck, and the width of the decks. An essential point is that there is no arch structure and therefore a support of the upper deck is needed to avoid its collapse on the aperture. This is taking some space in the aperture: for the design shown below, the HD2 has a free aperture of 17.5 mm radius, and FrescaII has 50 mm radius. The efficiency γ_c of the coil is 5-10% lower than in sector coils, but the efficiency approaches the sector coil for larger and larger ratios coil width/aperture width: it is 5.9×10^{-4} T mm/A for HD2, 6.1×10^{-4} T mm/A for FrescaII, and 6.4×10^{-4} T mm/A for RMM.

We conclude with section with an exercise to estimate the efficiency of the double helix layout, that is lower than the sector coil by at least the factor cos α.

Example 4.1: The D2 corrector [15] in the HL-LHC project is based on double helix design, with an angle α of 30 degrees. It requires 8800 m of 0.825 mm strand, for a double aperture magnet. The aperture diameter is 105 mm, and the diameter of the insulated strand is 1 mm. Ten wires are wound in a slot with 2.1 mm width and 5.1 mm height. The magnet operates at 392 A giving 2.62 T over 1.92 m of magnetic length. In case of ironless single coil configuration, the field at 392 A is 1.7 T. Compute the coil efficiency of this layout.

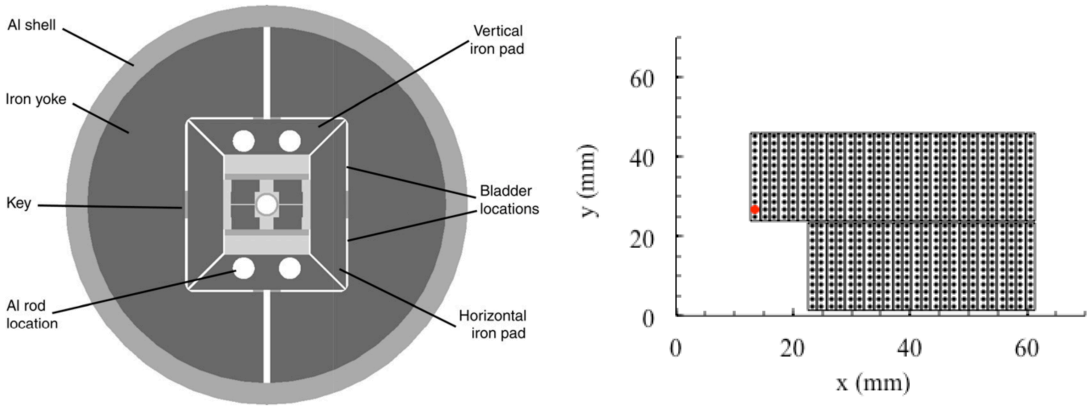


Fig. 4.15: Cross-section of HD2 dipole (left) and of its coil (right)

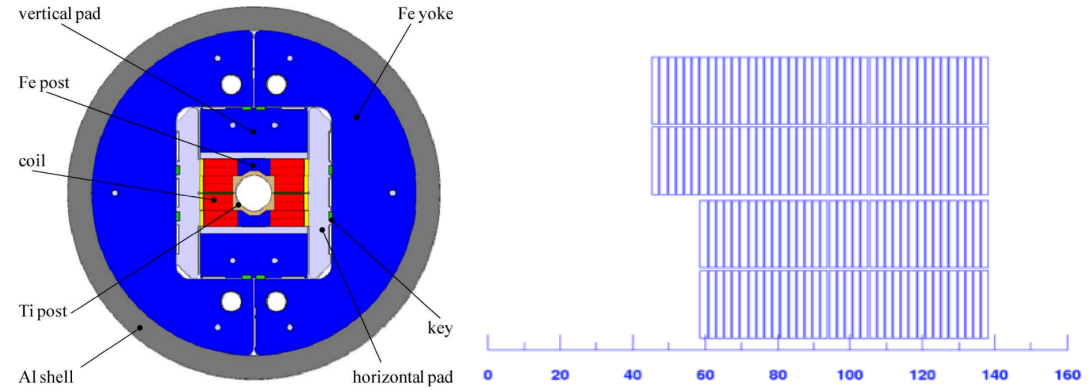


Fig. 4.16: Cross-section of FrescaII dipole (left) and of its coil (right).

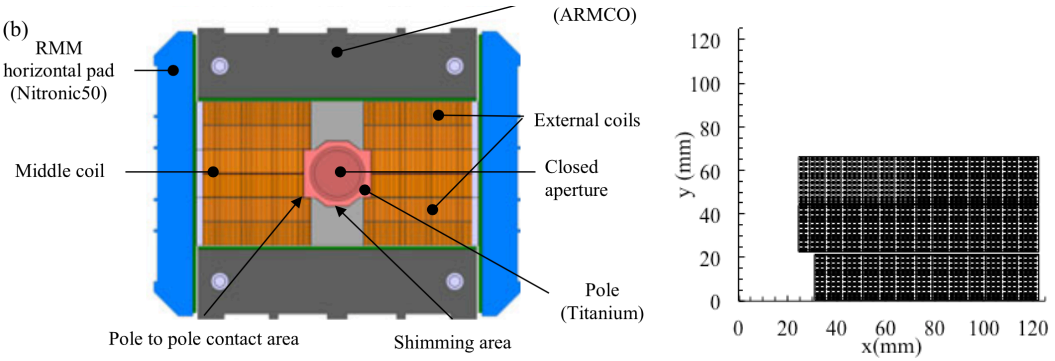


Fig. 4.17: Cross-section of RMM dipole (left) and of its coil (right).

The total volume of the insulated coil is $4.4\times10^6/10\times5.1\times2.1=9.42\times10^6$ mm^3. Since there are two apertures and the magnetic length is 1.92 m, the cross-section area is 2454 mm^2. Using the definition of equivalent coil width one finds

$$w_{eq} = r\left(\sqrt{1+\frac{3A}{2\pi r^2}}-1\right) = 52.5\left(\sqrt{1+\frac{3\times 2453}{2\pi\left(52.5\right)^2}}-1\right) = 10.2 \text{ mm} \tag{4.27}$$

Since the current density is 4920/5.1/2.1=459 A/mm^2, the efficiency is $1.7/10.16/459=3.64\times10^{-4}$ T A/mm, i.e. 55% of the efficiency of a sector winding. Note that the geometric factor of loss due to the angle of 30° is cosα=0.87.

4.6 Field quality optimization for quadrupoles based on sector coils

We now consider the case of a quadrupole magnet, as discussed in [4]. In this case the four sectors are arranged around the aperture, with current signs alternating. The integration of the second order contribution gives (see 4.10)

$$\frac{B_2}{R_{ref}} = -\frac{2\mu_0 j}{\pi}\sin(2\alpha)\ln\left(1+\frac{w}{r}\right) \tag{4.28}$$

and considering a 30° degrees sector coil one finds

$$G \equiv \frac{B_2}{R_{ref}} = -\frac{\sqrt{3}\mu_0 j}{\pi}\ln\left(1+\frac{w}{r}\right) \tag{4.29}$$

and using the more comfortable units

$$G = -\frac{\sqrt{3}\times 4\pi\times 10^{-7}}{\pi}10^6\, j\left[\mathrm{A/mm^2}\right]\ln\left(1+\frac{w}{r}\right) = 0.69\, j\left[\mathrm{A/mm^2}\right]\ln\left(1+\frac{w}{r}\right). \tag{4.30}$$

For instance, in the LHC main quadrupole one has a 440 A/mm^2 current density, and a coil width of 31 mm similar, and an aperture radius of 28 mm. Therefore one finds

$$G = 0.69\times 440\times\ln\left(1+\frac{31}{28}\right) \approx 0.69\times 440\times 0.75 = 220\ \mathrm{T/m}. \tag{4.31}$$

The equations for the field quality optimization are very similar to the dipole ones, provided that (i) angles are divided by a factor two and (ii) multipole orders are multiplied by a factor 2. Therefore a coil layout with one wedge can cancel both B_6 and B_{10} if that the wedge angles satisfy the equations

$$\begin{aligned}\sin(6\alpha_3)-\sin(6\alpha_2)+\sin(6\alpha_1)&=0\\ \sin(10\alpha_3)-\sin(10\alpha_2)+\sin(10\alpha_1)&=0\end{aligned}. \tag{4.32}$$

The integer solutions of interest for quadrupole layout, canceling b_6 and b_{10}, are

- (0°-12°), (18°-30°);
- (0°-18°), (22°-32°);
- (0°-24°), (30°-36°).

Field quality optimization is much easier in quadrupoles, since the the non-zero orders are more spaced (a quadrupole has only two multipoles to cancel up to order 10, i.e. b_6 and b_{10}, whereas a dipole has b_3 b_5 b_7 and b_9), and even the cross-section of two layers quadrupoles can be easily interpreted according to the solutions given above.

For instance, the LHC quadrupole has an inner layer based on the (0°-24°), (30°-36°), and the outer layer on a 30 degrees sector, where the wedge is needed just for placing the conductors back to a perpendicular positions (see Fig. 4.18, left). The 30 degrees layout does not cancel B_{10}, but since the cable width is large, the outer layer is so far from the aperture that its influence on B_{10} is negligible.

As a second case, let us consider the Tevatron main quadrupole: here the inner layer follows the (0°-12°), (18°-30°) layout, and the outer layer is a 30° wedgeless sector coil, as also in this case the influence of the outer layer on B_{10} is negligible (see Fig. 4.18, right).

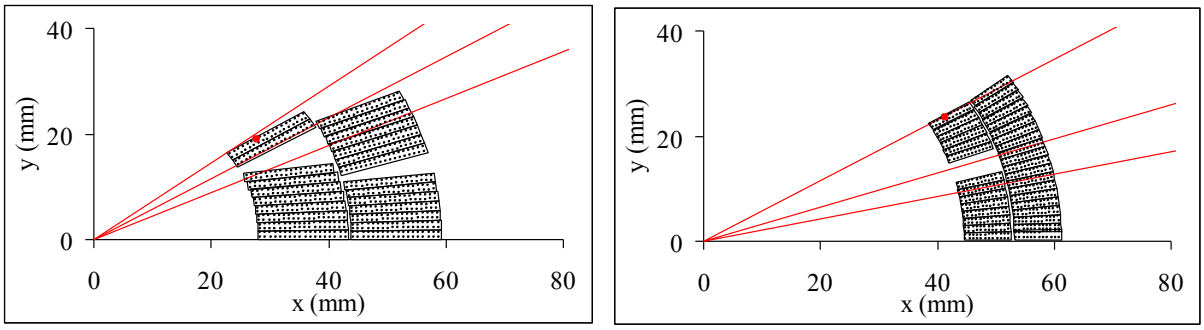


Fig. 4.18: Cross-section of LHC quadrupole coil, with 24°, 30° and 36° angles indicated in red (right), and of Tevatron quadrupole (left) and of its coil, with 18°, 24° and 30° angles indicated in red (right).

4.7 Guidelines for field quality optimization

Field quality optimization is a bit more complex than what shown in the previous sections. On the top of the geometric contribution discussed in this chapter, at low currents the magnetization of the superconductor can perturb the field harmonics, especially if the energy increase of the accelerator is large and therefore the main magnet are used at injection with very low currents; the phenomena of magnetization will be discussed in Chapter 6. The iron also plays a relevant role, and will be treated in Chapter 9; its influence on field and harmonics can be derived for some shapes in the case of non saturated domains, via the image current method. However, when the iron becomes saturated, finite element methods should be used to compute its impact on the main field and on the harmonics. One of the most used codes is ROXIE (see [16] for the first reference to the initial phase of this fundamental software, and [17] for the dedicated web site), a code dedicated to accelerator magnets developed at CERN by S. Russenschuck and collaborators, including also optimization routines, quench protection estimates. OPERA [18], a commercial code, is also widely used in the community for a vast range of applications. The perturbation induced by iron saturation on harmonics can be minimized via a proper shaping of iron, as discussed in Chapter 9.6. Both persistent currents and iron saturation are second order effect with respect to the geometric contribution. Therefore, the problem of field quality can and should be sliced, and the geometric part for a dipole should be the first and the more complex slice to swallow. No point in playing with iron shape if the coil geometry is not well optimized.

We have seen that for the geometrical contribution there are five parameters to be optimized (b_3 to b_{11}). The safest method is to first look for a coil layout that provides a good field quality (all multipoles within a few units), forgetting about iron saturation, superconductor magnetization, and just using the Biot-Savart contribution. Each block has in principle three free parameters: the angular position of the lower cable ϕ, the number of conductors, and the inclination of the block α (see Fig. 4.19). However, the third parameter has a very limited range, and the second is quantized by the finite thickness of the cable. Therefore with two wedges (three blocks) one can have not enough flexibility to optimize field quality. Three wedges (four blocks) can provide the needed flexibility, and four wedges (five blocks) provide an additional couple of free parameter that can turn out to be very useful in finding more efficient and flexible solutions.

If you have two layers, also in this case three wedges (five blocks) could be enough, usually split bewteen two in the inner layer and one in the outer. Four wedges provide a much larger flexibility, and is the solution adopted for the SSC prototypes and for the final design of the LHC dipoles. However, LHC dipole prototype using five blocks were also manufactured.

When you have two layers, start minimizing the higher orders with the inner layer, and then tune the outer layer to set to zero also b_3 and b_5. In general, a layout with well optimized high orders can be fine tuned to achieve any required large variations of low orders without spoiling the optimization. On

the other hand, if you start with a solution with a wrong high order (i.e. 50 units of b_9), it is very difficult to find a solution in its neighbourhood with zero units of b_9, and you will have to start your optimization from scratch.

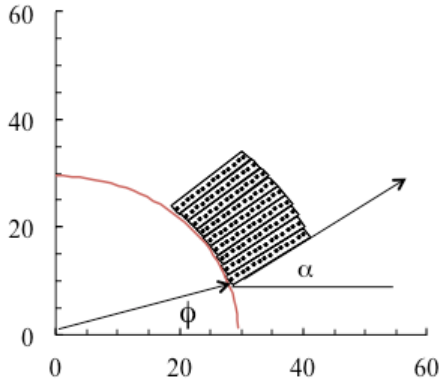


Fig. 4.19: The two angles ϕ and α giving the position of a coil block according to ROXIE convention.

Once you have found a good single aperture ironless cross-section given the aperture and the cable size, increase the complexity adding the second aperture (if the magnet is two in one) and the iron. The multipoles will move a bit from the target, and to compute how much you will need to run a finite element code with all the detail of the geometry. Once you have the offset that you need to correct, go back to your Biot-Savart model, make a sensitivity matrix of the multipoles on the wedge position, and try to solve the inverse problem. Since the computation of the Biot-Savart contribution takes a fraction of second, whereas a full finite element code can take minutes, this approach can be much faster and also gives you some analytical insight.

Fully automated methods to find the optimal field quality exist, and they are based on a optimizer associated to the algorithm of evaluation. The main difficulty of these methods is related to the existence of local minima for a parametric space that has several dimensions. The slicing technique allows avoiding taking the problem in the full complexity or, in the language of a mathematician, to diagonalize the problem. On January 20, 1961, John Fitzgerald Kennedy said in his inaugural address "Ask not what your country can do for you – ask what you can do for your country." Talking about field quality optimization, I would suggest: "Ask not what your computer can think for you, ask what you can think for your computer."

Acknowledgements

I wish to thank L. Rossi, who in 2006 suggested me to explore the analytical approaches to magnetic design using the model of the sector coils, and for the all the work carried out together about dipoles and quadrupoles magnetic design.

References

1. M. N. Wilson, "Superconduting magnets" Claredon Press (1987)
2. K.-H. Mess, P. Schmuser, S. Wolff, "Superconducting accelerator magnets" World Scientific, Singapore (1996)
3. L. Rossi, E. Todesco, "Electromagnetic design of superconducting dipoles based on sector coils", Phys. Rev. STAB 10 (2007) 112401
4. L. Rossi, E. Todesco, "Electromagnetic design of superconducting quadrupoles", Phys. Rev. STAB 9 (2006) 102401

5. D. I. Meyer and R. Flasck, “A new configuration for a dipole magnet for use in high energy physics applications” Nucl. Instrum. Meth. 80 (1970) 339-341
6. M. Anerella, et al., “The RHIC magnet system” Nucl. Instrum. Meth. A 499 (2003) 280-315
7. R. Palmer and A. V. Tollestrup, “Superconducting magnet technology for accelerators” Ann. Rev. Nucl. Part. Sci. 34 (1984) 247
8. L. Rossi, E. Todesco, “Electromagnetic efficiency of block design in superconducting dipoles” IEEE Trans. Appl. Supercond. 19 (2009) 1186-1190
9. J. v. Nugteren, et al “Idealized coil cross-sections with minimized conductor area for high field dipoles” IEEE Trans. Appl. Supercond. 28 (2018) 4000205
10. R. Perin, in *Encyclopedia of Applied Superconductivity* (IOP, London, 1998), pp. 919–950
11. A. Zlobin, et al. “Design concept and parameters of a 15 T Nb3Sn dipole demonstrator for a 100 TeV hadron collider” International Particle Accelerator Conference (2015) 3365-3367
12. G. L. Sabbi, et al., “Design of HD2: a 15 tesla Nb3Sn dipole with a 35 mm bore” IEEE Trans. Appl. Supercond. 15 (2005) 1128-1131
13. A. Milanese, et al., “Design of the EuCARD high field model dipole Fresca2” IEEE Trans. Appl. 22 (2012) 4002604
14. S. Izquierdo Bermudez, et al. “Design of ERMC and RMM, the base of the Nb3Sn 16 T magnet development at CERN” IEEE Trans. Appl. Supercond. 27 (2017) 4002004
15. G. Kirby, et al., “Hi-Lumi LHC twin aperture orbit correctors magnet system optimisation” IEEE Trans. Appl. Supercond. 27 (2017) 4002805
16. S. Russenschuck, T. Tortschanoff, “Mathematical optimization of supeorconducting accelerator magnets” IEEE. Trans. Magn. 30 (1994) 3419-3422
17. www.cern.ch/roxie
18. Vector Fields, "OPERA-2d & 3-d User Guide ", Vector Fields Limited, England, 1999

Ezio.Todesco@cern.ch

Chapter 8

Short sample fields and sensitivity analysis

Plan of the chapter
The maximum field that can be achieved in a superconducting magnet is given by the combination of current density and field in the coil that matches the superconductor critical surface: this is called the short sample field. In this chapter we will discuss how to give analytical expression for the short sample field as a function of the main features of the design. We will first introduce a linearization of the critical surface for Nb-Ti and Nb_3Sn, discussing the validity limits, and giving typical values for the state of the art superconductors in 8.1. We will then introduce the concept of filling factor to measure the dilution of the superconductor in the coil in 8.2. A third relevant element for the design of superconducting magnets is the ratio between the peak field in the conductor and the bore field, which is discussed in 8.3. In section 8.4 we put together all these elements to derive an equation for the short sample field as a function of the magnet parameters. The parametric dependence of this equation and the implications on the design of dipoles are discussed in section 8.5. The "small coils" and "large coils" regimes are also defined and the equations are considerably simplified via the introduction of a factor *X*. Using this concept, in 8.6 a sensitivity analysis is carried out in a fully analytical way. We conclude this chapter by sketching in 8.7 a more precise fit for the Nb_3Sn critical surface, still alowing to derive explicit equations for the short sample field for the dipole, and treating in 8.8 the case of the quadrupole.

8.1 Critical surface linearization for Nb-Ti and Nb_3Sn

For the Nb-Ti high current density strands that are used for accelerator magnets, the critical surface can be approximated via a linear fit

$$j_{sc,c} = s(b - B) \,. \tag{8.1}$$

For a mature Nb-Ti conductor as that one of the LHC dipoles [1], and in the range 7.5 to 10.5 T at 1.9 K and 5 to 8 T at 4.5 K, the slope s is ~600 A/(T/mm^2): this means that one tesla less in the conductor increases the critical current density by 600 A/mm^2. The intercept b is ~9.5 T at 4.5 K and ~12.9 T at 1.9 K (see Fig. 8.1, left).

For Nb_3Sn the surface is less linear with a positive second derivative (see Fig. 8.1, right), making the critical current is larger than its linear fit both in lower and higher field regions; however, the linear fit has a validity range of few tesla as for the Nb-Ti. We show in Fig. 8.1, right, case of the HL-LHC strand used for the MQXF magnets (1280 A/mm^2 at 15 T, and 1480 A/mm^2 at 12 T, both at 4.22 K [2]): the linear fit in the 12 to 15 T range gives a slope of ~400 A/(T/mm^2) and a intercept of 17.8 T at 4.5 K and 19.4 T at 1.9 K.

Comparing Nb_3Sn to Nb-Ti, we can make the following remarks:

- Taking 1500 A/mm^2 as a reference value for the critical current density in the superconductor (we will explain later in the text the special meaning of this value), and 1.9 K as an operational temperature, we see that Nb_3Sn allows to reach 16 T, compared to 10.5 T for Nb-Ti (50% more);
- The slope of Nb_3Sn is 33% smaller than the slope of Nb-Ti. Ideally, a conductor should have the slowest possible slope, since the larger the slope, the larger the reduction of critical current for higher fields. HTS are characterized by a very small slope (i.e. critical current is nearly independent of the field);
- Taking again 1500 A/mm^2 as a reference value, going from 4.5 K to 1.9 K cooling allows increasing the maximum reachable field in Nb-Ti by 40% (from 7 T to 10 T). In Nb_3Sn the gain is only 15%, i.e. from 14 T to 16 T.

In the following sections we will show how the maximum reachable field in a dipole depends on the slope and on the intercept of the linear fit.

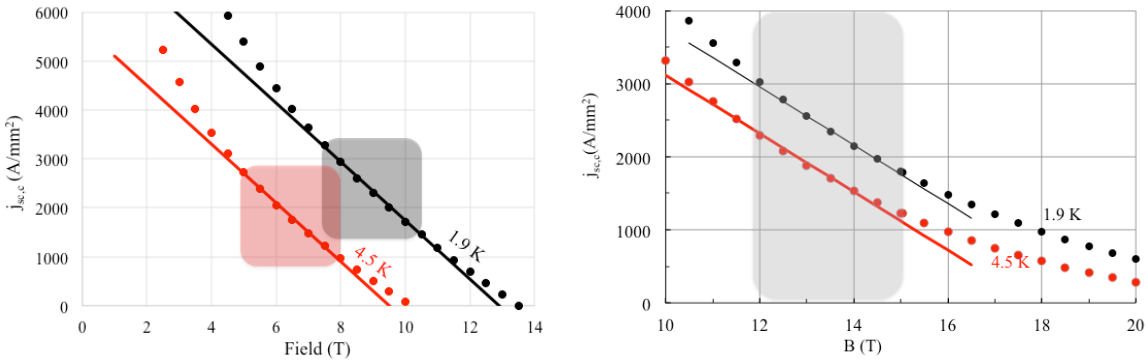


Fig. 8.1: Critical surface of Nb-Ti (left) and Nb_3Sn (right), indicated with round markers, and linear approximation valid in the shadowed ranges.

8.2 Filling factor

As explained in Chapter 5, 6 and 7, a bulk superconductor cannot be used to wind magnets; moreover the cable usually requires an insulation. Therefore the superconductor critical current density is diluted in the insulated cable via a filling factor κ, having the following components:

$$\kappa = \kappa_{w-c}\kappa_{c-i}\frac{1}{1+\nu}. \tag{8.2}$$

where κ_{w-c} is the volumetric fraction of the strand in the cable, κ_{c-i} is the volumetric fraction of the cable in the insulated cable, and ν is the copper to non copper ratio, i.e. the strand has a volume 1 of non-copper and a volume ν of copper. Note that ν=1 means same volume of copper and of non-copper; ν=2 means the the volume of copper is twice the volume of the non-copper, and so on. Let us analyse the threee components of Eq. (8.2)

- The voids in the cable are order of 15%, and therefore κ_{w-c}~0.85. Note that Nb-Ti cables can be compacted more than Nb_3Sn (see also Chapter 7).
- The volumetric fraction of insulation is order of 10% to 20%, and therefore κ_{w-c}~0.8 to 0.9. Note that the volumetric part of insulation in the insulated cable depends on the cable size: infact the insulation thickness is typically between 0.10 and 0.15 mm, and the larger the cable, the smaller the fraction of insulation in the cable, and the higher κ_{c-i}.

- The copper fraction should not be above 1; this requirement comes from stability (see Chapter 6) and protection (see Chapter 12 and 13). Note that it can range up to 2 (one third of superconductor and two thirds of copper) for main magnets. On the other hand, if much lower current density is used in high field regions (concept of grading, that will be described in Chapter 9), one can reduce the copper content towards 1. In correctors, where lower fields are targeted, and therefore the current density in the superconductor can be higher, the fraction of copper can become larger and larger (ν well above 2), up to guarantee self protection, thus simplyfing the whole system.

A list of copper fraction, and idlutions factors due to insulation and to voids in the cable are given in Table 8.1. Typically, 1/3 to1/4 of the insulated cable is composed of superconductor.

Table 8.1: Copper fraction, dilution factors due to cable and to insulation, and filling factor in some superconducting main accelerator magnets.

		ν	$\kappa_{w\text{-}c}$	$\kappa_{c\text{-}i}$	κ
LHC MB inner	Nb-Ti	1.650	0.87	0.87	0.286
LHC MB outer	Nb-Ti	1.950	0.86	0.83	0.243
Tevatron MB	Nb-Ti	1.854	0.82	0.81	0.232
RHIC MB	Nb-Ti	2.250	0.87	0.84	0.226
HERA MB	Nb-Ti	1.878	0.89	0.85	0.262
SSC MB inner	Nb-Ti	1.501	0.84	0.89	0.298
SSC MB outer	Nb-Ti	1.777	0.88	0.84	0.269
FRESCA inner	Nb-Ti	1.600	0.87	0.88	0.293
FRESCA outer	Nb-Ti	1.870	0.88	0.85	0.260
MSUT inner	Nb_3Sn	1.250	0.82	0.88	0.320
MSUT outer	Nb_3Sn	1.250	0.88	0.85	0.333
D20 inner	Nb_3Sn	0.430	0.80	0.84	0.471
D20 outer	Nb_3Sn	1.000	0.87	0.77	0.332
HD2	Nb_3Sn	1.000	0.82	0.86	0.349
FRESCA2	Nb_3Sn	1.250	0.81	0.80	0.289
FNAL HFDA02/03	Nb_3Sn	1.250	0.86	0.76	0.288
11 T	Nb_3Sn	1.100	0.79	0.86	0.323
MQXF	Nb_3Sn	1.200	0.82	0.83	0.308

When using an insulated cable, the overall critical current density that can be carried is

$$j_{sc,c}^{o} = \kappa j_{sc,c} \,. \tag{8.3}$$

wher the index *o* denotes that the current density is computed on the insulated conductor. In the literature the term engineering current density is also used: it refers to the critical current over the strand, i.e.

$$j_{sc,c}^{eng} = \frac{1}{1+\nu} j_{sc,c} \,. \tag{8.4}$$

Example 8.1: The MQXF bare cable [2] has a width of 18.363 mm, inner thickness of 1.530 mm and outer thickness of 1.658 mm, an insulation thickness is 0.145 mm. The cable is made of 40 strands with 0.85 mm diameter, and copper fraction is 1.2. (i) Compute the volumetric fraction of the strand in the cable $\kappa_{w\text{-}c}$, the volumetric fraction of the cable in the insulated cable$\kappa_{c\text{-}I}$, and the filling ratio κ. (ii) What is the increase in the filling factor if the insulation thickness is reduced to 0.100 mm ? (iii) And if the copper fraction is reduced to 1.1 ?

(i) The volumetric fraction of the strand in the cable is

$$\kappa_{w-c} = \frac{40 \times \pi \times (0.85/2)^2}{18.363 \times (1.658 + 1.530)/2} = \frac{22.698}{29.271} = 0.775 \,. \tag{8.5}$$

The volumetric fraction of cable in the insulated cable is

$$\kappa_{c-i} = \frac{18.363 \times (1.658 + 1.530)/2}{(18.363 + 0.145 \times 2) \times (1.658 + 1.530 + 0.145 \times 4)/2} = \frac{29.271}{35.142} = 0.833 . \tag{8.6}$$

The filling factor is

$$\kappa = \kappa_{w-c} \kappa_{c-i} \frac{1}{1+\nu} = 0.775 \times 0.833 \times \frac{1}{1+1.2} = 0.294 , \tag{8.7}$$

that is, 29.4% of the insulated cable cross-section is made of superconductor.

(ii) If the insulation thickness is reduced to 0.100 mm, one will have

$$\kappa_{c-i} = \frac{18.363 \times (1.658 + 1.530)/2}{(18.363 + 0.1 \times 2) \times (1.658 + 1.530 + 0.1 \times 4)/2} = \frac{29.271}{33.302} = 0.879 . \tag{8.8}$$

and therefore the filling factor increases to 31%:

$$\kappa = \kappa_{w-c} \kappa_{c-i} \frac{1}{1+\nu} = 0.775 \times 0.879 \times \frac{1}{1+1.2} = 0.310 . \tag{8.9}$$

(iii) If the copper fraction is reduced from 1.2 to 1.1, the filling factor is 30.8%:

$$\kappa = \kappa_{w-c} \kappa_{c-i} \frac{1}{1+\nu} = 0.775 \times 0.833 \times \frac{1}{1+1.1} = 0.308 . \tag{8.10}$$

8.3 A digression about the ratio between peak field and bore field

The last element that we are missing to have a complete analytical estimate of the short sample field in a sector dipole is the ratio between the peak field in the conductor and the bore field

$$\lambda \equiv \frac{B_p}{B} . \tag{8.11}$$

A superconducting magnet is limited by the peak field in the conductor: the fact that this field is higher than the bore field reduces the magnet performances. For this reason this discussion is placed in this chapter and not in Chapter 4, dealing with electromagnets and not specially focusing on superconductors.

For a sector coil one can compute the ratio between the peak field and the bore field; this is done for layouts without wedges and with some wedges to better optimize field quality, as discussed in Chapter 4; results are shown in Fig. 8.2.

The pattern is very similar for the different configurations:

- The ratio λ tends to 1 for larger coil widths: in a thick coil the loss due to this factor becomes smaller and smaller.
- The ratio λ diverges for smaller coil widths; in a thin coil the loss due to this factor becomes larger and larger.
- What counts is not the coil width, but the ratio between coil width and aperture radius.
- For a coil width equal to the aperture radius, the ratio l is order of 1.04, i.e. the peak field in the conductor is ~4% larger than the bore field.

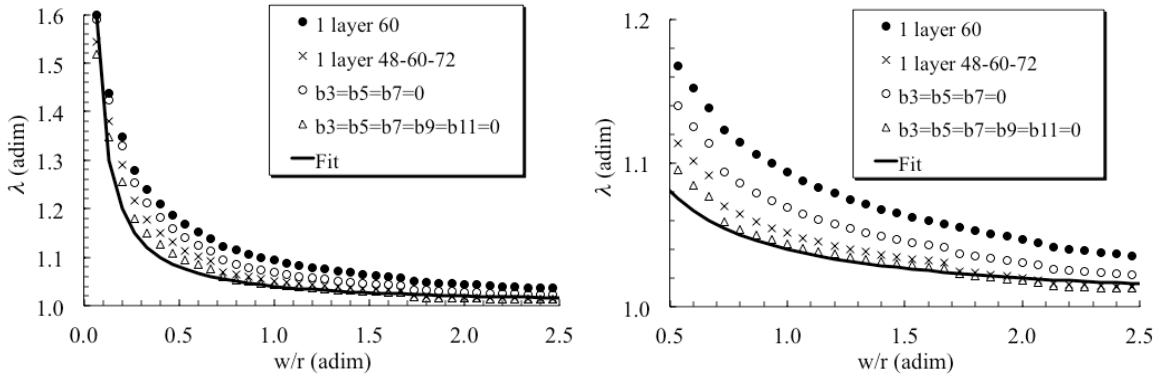


Fig. 8.2: Ratio λ between peak field in the coil and bore field versus ratio between sector width and aperture radius, for different wedges configurations.

We propose a fit for the ratio λ, for the layouts with the most optimized field quality (see Fig. 8.2). The equation used is

$$\lambda \sim 1 + a\frac{r}{w_{eq}} \tag{8.12}$$

with a~0.04. Obviously, the 60° secor coil has a worse λ, since the field quality is less optimized and therefore the larger field dishomogeneties increase the peak field on the coil. The comparision of this fit to the values computed in actual coil layouts with numerical codes is given in Fig. 8.3.

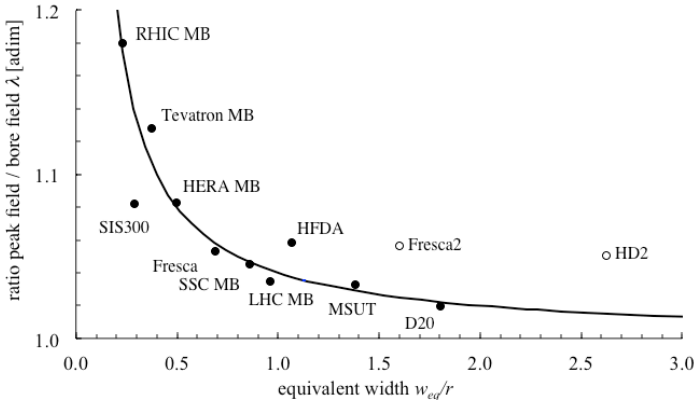


Fig. 8.3: Ratio λ between peak field in the coil and bore field, versus ratio between equivalent coil width and aperture radius, for different magnets, and proposed hyperbolic fit [7].

One can make the following remarks:

- The hyperbolic trend seen in the sector coils is confirmed by the actual layouts, i.e. the case of small ratios between coil width and aperture radius one has a much larger ratio between peak field and bore field (see for instance the RHIC coil [3]), whereas for very large col width the peak field approaches more and more the bore field (see for instance the LHC dipoles [1];
- Block coils as FrescaII [4] or HD2 [5] (see Chapter 4) present a less favorable λ with respect to cosθ layouts;
- A special optimization can give better results than the fit, as in the case of SIS300 layout [6]; here the last blocks were made with fewer conductors and with a large angular separation to

the next blocks. This special optimization allows reducing the ratio λ from an expected 1.16 to 1.08. We will see in the next sections what is the implication of this optimization on the magnet performances.

Example 8.2: What is the ratio between peak field and bore field that one can expect for the HL-LHC D1 [8], having a 15 mm width coil in a 150 mm aperture diameter? And what can be expected for the MDPCT1 dipole [9], having 4 layers with 15 mm wifth cable and 50 mm aperture diameter ?

In the HL-LHC dipole case one can estimate

$$\lambda \sim 1 + 0.04\frac{75}{15} = 1.20 \tag{8.13}$$

Infact, the bore field is 5.6 T and the peak field is 6.58, thus giving λ=1.175. Note that the actual layout has 6.26 peak field in the coil for a 5.60 T bore field, giving λ=1.12. Once more, for small ratios between coil width and aperture there is a clear trend to go to much higher λ, but our fit based on sector coils can be considerably improved.

In the MDPCT1 dipole case one finds

$$\lambda \sim 1 + 0.04\frac{25}{60} = 1.02\,. \tag{8.14}$$

The actual computations give 15.5 T peak field for a 15 T bore field, thus λ=1.04.

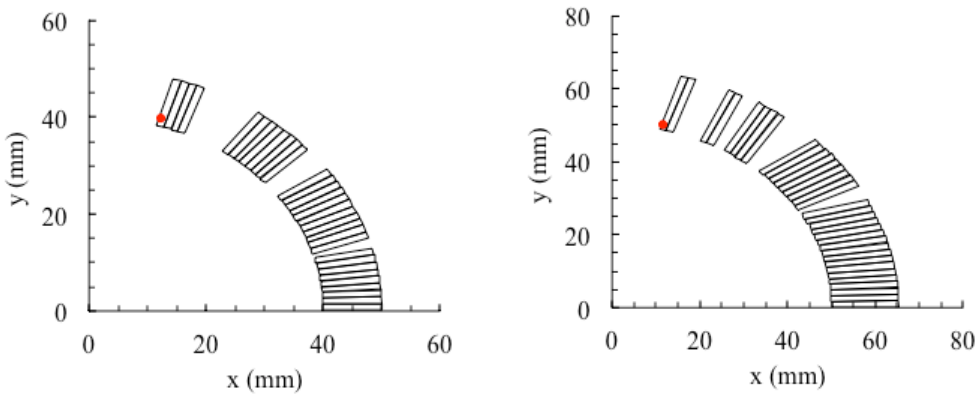


Fig. 8.4: RHIC (left) and SIS300 (right) coil layouts, with peak field location

8.4 Short sample field in dipoles

The short sample field is defined as the maximum bore field that can be obtained in a superconducting magnet, i.e. when the current density in the superconductor creates a peak field in the coil that correspond to the superconductor critical surface (see Fig. 8.5). The bore field is proportional to the overall current density according to the equations derived in Chapter 4

$$B = \gamma_c j w_{eq}\,. \tag{8.15}$$

where for an ironeless sector coil, with optimized field quality, we found $\gamma_c \sim 0.66\times10^{-4}$ T mm/A, i.e. each 10 mm of coil width with 400 A/mm^2 overall current density give ~2.6 T:

$$\gamma_c j w_{eq} = 0.00066 \times 400 \times 10 = 2.6 \text{ T}\,. \tag{8.16}$$

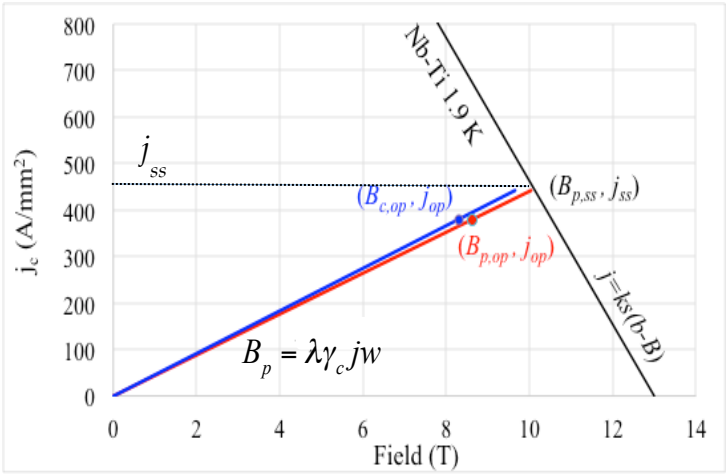


Fig. 8.5: Short sample field and operational field for the LHC dipole, linear approximation both for critical surface and for magnet loadline.

The peak field in the conductor is given by λ times the bore field

$$B_p = \lambda \gamma_c j w_{eq} . \tag{8.17}$$

According to (8.1) and (8.3) the linearization of the critical surface can be parametrized as

$$j = \kappa s (b - B_p) \tag{8.18}$$

and the intersection between (8.17) and (8.18) gives the short sample conditions (see Fig. 8.5)

$$j_{ss} = \kappa s (b - B_{p,ss}) \qquad B_{p,ss} = \lambda \gamma_c j_{ss} w_{eq} . \tag{8.19}$$

Replacing the right hand side in the left hand side one obtains

$$j_{ss} = \kappa s (b - \lambda \gamma_c j_{ss} w_{eq}) \qquad j_{ss}(1 + \lambda \kappa s \gamma_c w_{eq}) = \kappa s b \tag{8.20}$$

one can solve for the short sample current density

$$j_{ss} = b \frac{\kappa s}{1 + \lambda k s \gamma_c w_{eq}} = b \frac{\kappa s}{1 + \left(1 + a \frac{r}{w_{eq}}\right) k s \gamma_c w_{eq}} . \tag{8.21}$$

The short sample peak field and bore field are then given by

$$B_{p,ss} = b \left(1 + a \frac{r}{w_{eq}}\right) \frac{\kappa s \gamma_c w_{eq}}{1 + \left(1 + a \frac{r}{w_{eq}}\right) k s \gamma_c w_{eq}} \tag{8.22}$$

$$B_{ss} = b \frac{\kappa s \gamma_c w_{eq}}{1 + \left(1 + a \frac{r}{w_{eq}}\right) k s \gamma_c w_{eq}} . \tag{8.23}$$

We recall the meaning of the different variables:

Magnet parameters:

- r (mm) is the aperture radius of the magnet (remember that the magnet aperture usually refers to aperture diameter);
- w_{eq} (mm) is the equivalent coil width, as defined in Chapter 4 (the width of a 60º sector coil with the same aperture radius and the same surface);

Cable parameters:

- κ (adim) is the filling factor, i.e. the volumetric fraction of the superconductor in the insulated cable, see section 8.3 and Eq. (8.2);

Superconductor properties:

- b (T) is the intercept of the linearization of the critical surface as given in Eq. (8.1);
- s (A mm^2/T): is the slope of the linearization of the critical surface as given in Eq. (8.1);

Constants for the designs:

- The constant γ_c is the coil efficiency as defined in Chapter 4, and is the ratio between field and overall current density times the equivalent coil width (4.X); for ironless, single aperture, cos theta coils $\gamma_c \sim 6.6\times10^{-4}$ (T mm/A); the case of iron and double aperture is treated in the next chapter;
- The constant a~0.04 (adim) is the parameter of the fit for the ratio peak field bore field as given in section 8.2.

A full analysis of the dependence of short sample values on the magnet parameters will be done in the next section.

Example 8.3: Estimate the short sample field for a Nb-Ti magnet with 10 mm coil width at 4.5 K, 80 mm aperture and 0.25 filling factor.

One can compute the short sample field as

$$B_{ss} = b\frac{\kappa s\gamma_c w_{eq}}{1+\left(1+a\frac{r}{w_{eq}}\right)ks\gamma_c w_{eq}} = 9.5\frac{0.25\times600\times0.00066\times10}{1+\left(1+0.04\frac{40}{10}\right)\times0.25\times600\times0.00066\times10} = \tag{8.24}$$

$$= 9.5\frac{0.99}{1+1.16\times0.99} = 4.4 \text{ T}$$

Example 8.4: The LHC dipole operates with Nb-Ti cable at 1.9 K, and has two layers of 15 mm width cable, compute the short sample field and the short sample current ignoring (i) the contribution of the iron and (ii) the grading between the two coils. We also assume that (i) the equivalent width of the coil is 30 mm and (ii) the filling factor is 0.25. The aperture radius is 28 mm.

One can compute the short sample field as

$$B_{ss} = b\frac{\kappa s\gamma_c w_{eq}}{1+\left(1+a\frac{r}{w_{eq}}\right)ks\gamma_c w_{eq}} = \tag{8.25}$$

$$= 12.5\frac{0.25\times600\times0.00066\times30}{1+1.04\times0.25\times600\times0.00066\times30} = 12.5\frac{3.0}{1+1.04\times3.0} = 9.2 \text{ T}$$

Note that the actual short sample field of the LHC dipole is 9.7 T, thanks to the contribution of the iron, to the double aperure, and to the grading. This extremely simplified model already gives an estimate accurate within 10%.

8.5 Parametric dependence of short sample field

As clearly shown in the examples 8.3 and 8.4, the short sample estimates according to Eqs. (8.10)-(8.12) involve the computation of a adimensional factor X, given by

$$X \equiv \kappa s \gamma_c w_{eq} \tag{8.26}$$

that can be used to considerably simplify the computations. In Example 8.3 (RHIC dipole), this factor is 1.0, and in the example 8.4 (LHC dipole) is 3.0. The factor X is nothing else that the ratio between the slope of the critical surface and the slope of the loadline:

$$X \equiv \frac{\kappa s}{\frac{1}{\gamma_c w_{eq}}}. \tag{8.27}$$

that is, X=1 corresponds to a condition of perpendicularity between the loadline and the critical surface (see Fig. 8.6, left).

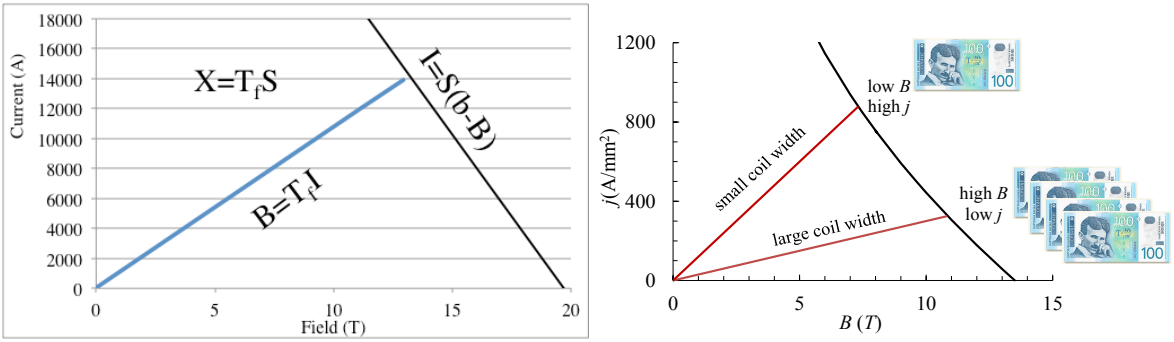


Fig. 8.6: geometrical interpretation of the factor X (left), and small coil width and large coil width regimes (right).

Using this definition, the complex definition of the short sample field (8.23) becomes as simple as as

$$B_{ss} = b\frac{X}{1+\lambda X}. \tag{8.28}$$

First observation is that for large coil width, the magnet can reach as short sample the intercept b of the critical surface:

$$w_{eq} \to \infty \qquad X \to \infty \qquad B_{ss} \to \frac{b}{\lambda} \to b. \tag{8.29}$$

In fact, one can define two regimes of the magnet design: a "thin coil" X<1 and a "thick coil" X>1. We use quotes since how much a coil is thin depends (see 8.24) not only on the coil width, but also on the filling factor and on the slope of the critical surface. A "small coil" is a very effective magnet, where the relatively low field allow profiting of high current densites (see Fig. 8.6, right); however, we will see that there are other design considerations (namely protection and mechanics, Chapters 10-13) that

do not allow to go above 1000 A/mm^2 in overall current density. On the other hand, a "large coil" is a less effective magnet, using much more conductor.

The second observation is about how the asymptotic limit of the maximum field tolerable by the superconductor (at zero current density) is reached. Eq. (8.25) clearly shows that this limit is reached for large coil widths as $w/1+w$. This means that larger and larger coil width produce smaller and smaller increases of the field, i.e. the layout becomes less and less effective. The dependence on the coil width, for a case of a magnet with 56 mm aperture (28 mm radius, corresponding to LHC main dipoles) made with Nb-Ti is shown in Fig. 8.7, left. The selection of the 30 mm coil width for LHC dipoles is based on the principle of the "largest affordable" coil width; one can see that 10 further mm of coil width would have provided only 0.5 T additional short sample field. Note that the quantity of superconductor is proportional to the coil cross-sectional surface, whose dependence on the coil width is more than linear

$$A = \frac{2\pi}{3}\left(w_{eq}^2 + 2rw_{eq}\right). \tag{8.30}$$

as shown in Fig. 8.7, right; doubling the quantity of conductor in the LHC dipole would have allowed to increase the short sample field by only 10% (one additional tesla).

The third observation is about the dependence on the filling factor, that is shown to be not more than one tesla for a large variation (from 33% to 25%), see Fig. 8.8. Finally, we point out that the short sample field has a very minor dependence on the aperture radius, that is present only in the expression of the ratio bore field peak field. We will make these observation more quantitative in the next section, where we will discuss a sensitivity analysis for the magnet design.

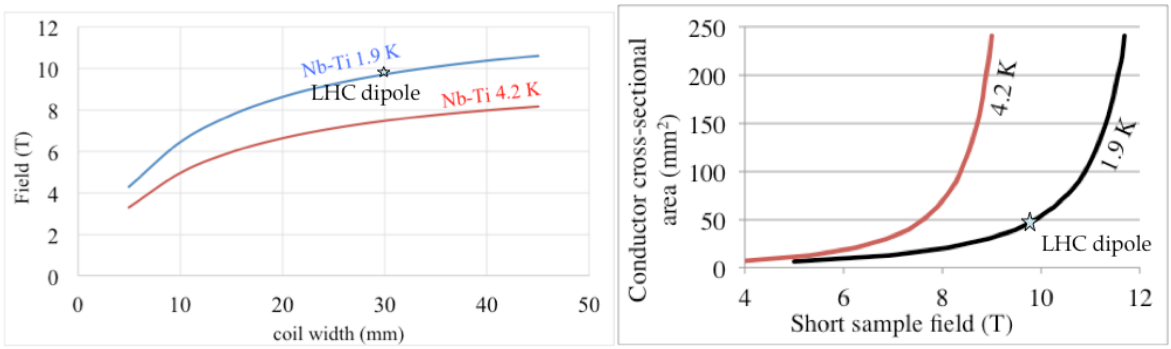


Fig. 8.7: Short sample field versus coil width for Nb-Ti at 1.9 K and at 4.5 K (left), and quantity of conductor versus short sample field (right) for 28 mm aperture radius coil, with filling factor κ=0.28.

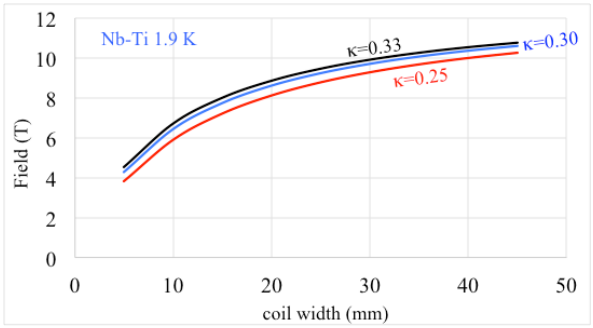


Fig. 8.8: Short sample field versus coil width in a Nb-Ti at 1.9 K and at 4.5 K, 28 mm aperture radius coil, with filling factor κ=0.28 (left) and for different filling factors

8.6 Sensitivity analysis using the factor *X*

Let us compute analytically what is the dependence of the short sample field in the different magnet parameters. This analysis is based on the linearization of the problem, i.e. it is valid only for small variations around the magnet design. However, it proves to be a powerful tool to understand where to address the efforts of magnetic design optimization.

The variation of the short sample field induced by a variation of the factor X is quite simple:

$$\Delta B_{ss} = b\frac{\Delta X}{1+\lambda X} - b\frac{\lambda \Delta X}{\left(1+\lambda X\right)^2} = b\frac{X}{1+\lambda X}\frac{\Delta X}{X}\left(1-\frac{\lambda X}{1+\lambda X}\right). \tag{8.31}$$

Therefore one has

$$\frac{\Delta B_{ss}}{B_{ss}} = \frac{\Delta X}{X}\left(\frac{1}{1+\lambda X}\right). \tag{8.32}$$

i.e., any relative increase of the factor X gives an increase in the short sample field that is reduced by $1/(1+\lambda X)$. Let us make few examples:

- A 10% increase of coil width gives a 10%/(1+λX) increase of short sample field;
- A 10% increase in the coil efficiency gives a 10%/(1+ λX) increase of short sample field;
- A 10% increase in the filling factor gives a 10%/(1+ λX) increase of short sample field;

Therefore, for small coil width ($X\sim 0$) any relative increase of X is fully affecting the short sample field, whereas for large coil width the short sample increase is much less (see Example 8.5 at the end of this section).

One can carry out the same computation for the dependence of the short sample field on λ; the result is radically different

$$\Delta B_{ss} = -b\frac{X^2}{\left(1+\lambda X\right)^2}\Delta\lambda = -\frac{bX}{1+\lambda X}\frac{X}{1+\lambda X}\Delta\lambda\,. \tag{8.33}$$

Therefore one has

$$\frac{\Delta B_{ss}}{B_{ss}} = -\frac{X}{1+\lambda X}\Delta\lambda\,, \tag{8.34}$$

i.e., any absolute decrease of the ratio peak field/bore field gives a relative increase in the short sample that is reduced by $\sim X/1+X$. Therefore, the situation is the opposite of the previous case: reducing λ has a much larger impact on the large coils short sample field than on the thin coils (see example 8.6 at the end of the section).

We conclude by discussing the dependence on the critical current. In principle there are two parameteric dependences: on the intercept b and on the slope s. Any relative increase of the intercept b is fully transmitted to the short sample field (see Fig. 8.9, left):

$$\frac{\Delta B_{ss}}{B_{ss}} = \frac{\Delta b}{b}\,. \tag{8.35}$$

However, when the current density improves, this usually involves only an increase of the slope (see 8.9, right), and not of the intercept.

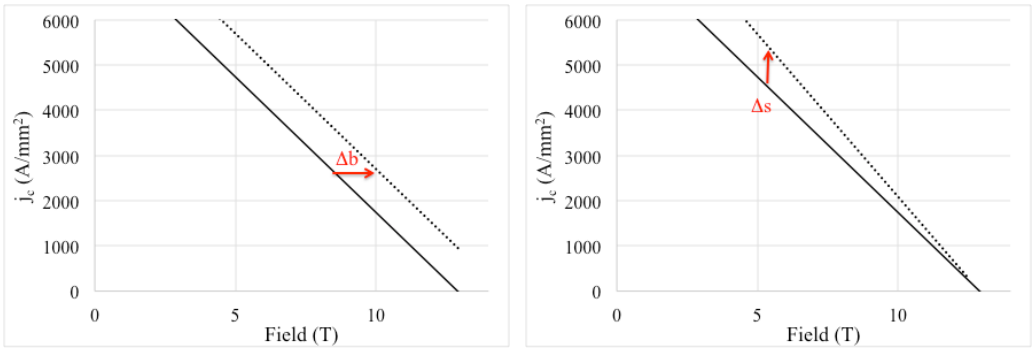


Fig. 8.9: Two way of increasing the critical current: increase of the critical field (left) and increase of the slope (right).

An increase of the slope is weighted as any other component of the factor X, i.e. according to

$$\frac{\Delta B_{ss}}{B_{ss}} = \frac{\Delta s}{s}\frac{1}{1+\lambda X}. \tag{8.36}$$

See also the example 8.7.

Example 8.5: The first layout of the LHC main dipole was using a 16.7 mm width bare cable [10]. Compute the increase in short sample field with respect to the final design that makes use of a 15.1 mm width bare cable [1].

Since we have seen in Example 8.3 that the LHC dipole has X=2.0, the 13% increase of the coil width (16.7/15.1=1.11) will give a 11%/3=3.5% increase of the short sample field.

Example 8.6: The LHC dipole has a ratio peak field/bore field of 1.035. What is the gain in the short sample field is this ratio is lowered to 1.02 ?

Since the factor X is 2.0, a decrease of 1.5% in the peak field/bore field will give according to (Eq. 8.34) 2/3 of 1.5% = 1.0% increase in the short sample field.

Example 8.7: The Nb_3Sn 11 T dipole has a 28 mm coil width and a 60 mm aperture radius. Compute the factor X, and estimate the gain in the short sample field if the critical current at 12 T, 4.22 K is increased from 2500 A/mm^2 to 3000 A/mm^2, but keeping the same critical field at zero current density (case shown in Fig. 8.9, right).

The factor X is estimated as

$$X = \kappa s \gamma_c w_{eq} = 0.28 \times 400 \times 0.00066 \times 28 = 2.1 \tag{8.37}$$

and therefore a 20% gain in the slope of the critical surface will increase the short sample field of 20%/3.1=6.5%.

8.7 A nonlinear fit of Nb_3Sn critical surface

In the case of Nb_3Sn, one can derive equations for the short sample that have a validity on a wider range of short sample fields. The critical surface is very close to an hyperbolic fit (see Fig. 8.10)

$$j = s\left(\frac{b}{B}+1\right). \tag{8.38}$$

This fit is very accurate since it corresponds to a linear dependence of the pinning force on the field:

$$F_p \equiv jB = sb + sB. \tag{8.39}$$

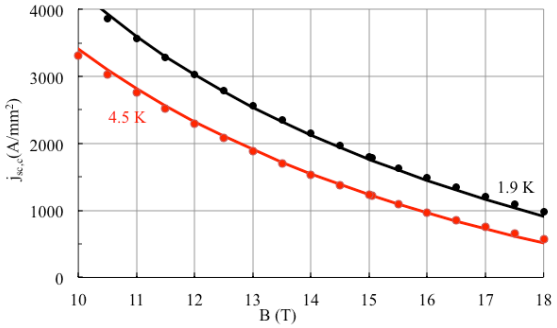


Fig. 8.10: Critical surface of Nb_3Sn (right), indicated with round markers, and hyperbolic approximation (solid lines).

Values for the HL-LHC strand are s=3100 A/mm^2 and b=21 T at 4.2 K, and s=3300 A/mm^2 and b=21.0 T at 4.5 K, and s=3300 A/mm^2 and b=23.0 T at 4.2 K, and s=3300 A/mm^2 and b=21 T at 4.5 K. The interception of this fit with the loadline

$$B_p = \lambda \gamma_c j w_{eq} \tag{8.39}$$

gives a second order equation that can be analytically solved. Using the factor X, one obtains

$$B_{ss} = \frac{X}{2}\left(\sqrt{\frac{4b}{\lambda X}+1}-1\right) \tag{8.40}$$

and can verify that also in this case the limit for large coils gives

$$B_{ss} = \frac{X}{2}\left(\sqrt{\frac{4b}{\lambda X}+1}-1\right) \xrightarrow{X\to\infty} \frac{X}{2}\left(\frac{2b}{\lambda X}+1-1\right) \xrightarrow{X\to\infty} \frac{b}{\lambda}. \tag{8.41}$$

In Fig. 8.11, right we show the short sample field for a Nb_3Sn dipole, with 25 mm aperture diameter, and filling factor 0.25 and the comparison to Nb-Ti case at 1.9 K. One can make the following remarks:

- At 1.9 K, Nb_3Sn allows to increase the short sample field of Nb-Ti by ~50%;
- The saturation observed for large coil width is slower in Nb_3Sn;
- Doubling the quantity of conductor above 14 T short sample field allows to increase it by 10% (see Fig. 8.11, right).

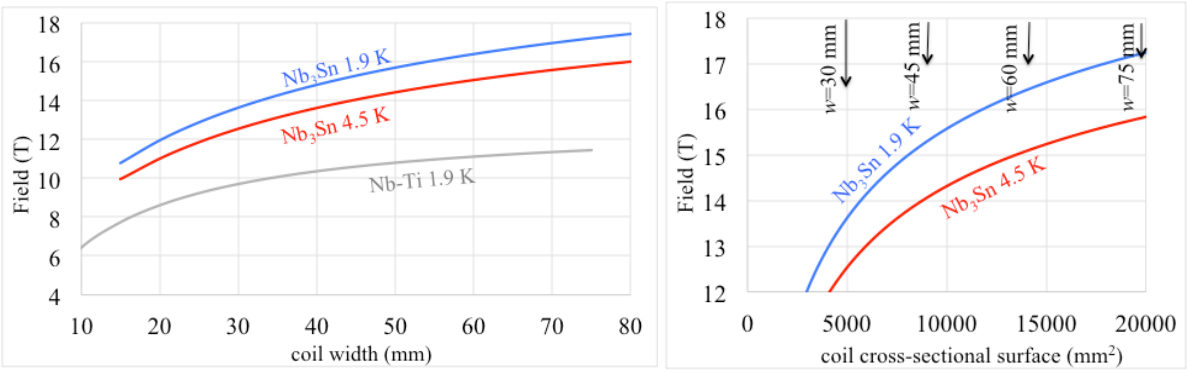


Fig. 8.11: Short sample field versus coil width for Nb_3Sn strand used in HL-LHC and Nb-Ti (right), and versus cross-sectional area of the surface (right).

8.8 The case of quadrupoles

In a quadrupole [11], the relation between gradient and coil width is given by

$$G = \gamma_c j \ln\left(1 + \frac{w_{eq}}{r}\right) \tag{8.42}$$

with γ_c~6.63×10^{-4} (T /m)/(A/mm^2), see Chapter 4. Similar to the dipole case (see section 8.2), one can define the ratio between peak field in the conductor and ideal peak field in the coil, that is the gradient times aperture

$$\lambda \equiv \frac{B_p}{Gr}. \tag{8.43}$$

Computing this ratio for a sector coil, one finds a new feature, not present in the case of the dipole; for large coil width, the ideal case (l=1) is not reached; typically it is very hard to get λ smaller than 1.10. This means that 10% of the superconductor performances are wasted on the peak field of the coil, without producing the expected gradient. The sector coil also shows a linear increase of λ with the coil width, after having reached a minimum (see Fig. 8.12), that can be fit according to

$$\lambda \approx a_{-1}\frac{r}{w_{eq}} + 1 + a_1\frac{w_{eq}}{r} \tag{8.44}$$

with a_1=0.11 a_{-1}=0.04. One can cure this linear divergence by having the sector coils at larger distances from the aperture with apertures smaller than 30°.The comparison between the fit and the values numerically computed for actual layouts is given in Fig. 8.12: for w/r>0.5 the field overshoot in the coil is between 1.10 and 1.15.

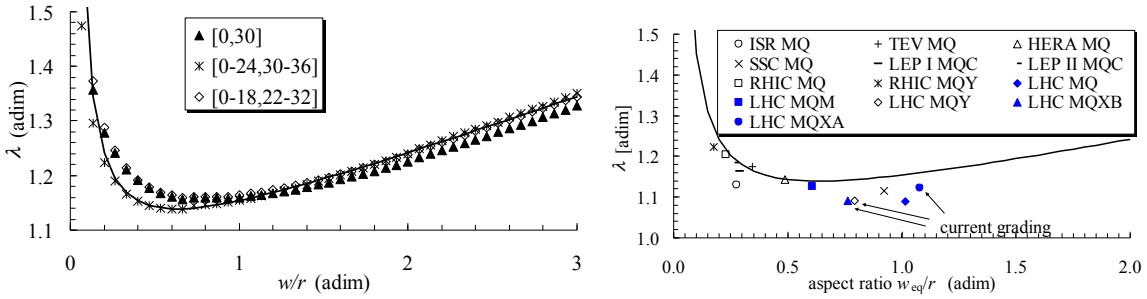


Fig. 8.12: Short sample field versus coil width in a Nb-Ti at 1.9 K and at 4.5 K, 28 mm aperture radius coil, with filling factor κ=0.28 (left) and for different filling factors

Going back to our equations, since

$$B_p = \lambda r G = \lambda r \gamma_c j \ln\left(1 + \frac{w_{eq}}{r}\right) \tag{8.45}$$

and the critical surface is given by

$$j = \kappa s\left(b - B\right) \tag{8.46}$$

one finds

$$B_{p,ss} = b\frac{\lambda r\kappa s\gamma_c \ln\left(1+\frac{w_{eq}}{r}\right)}{1+\lambda r\kappa s\gamma_c \ln\left(1+\frac{w_{eq}}{r}\right)} \tag{8.47}$$

and

$$G_{ss} = \frac{B_{p,ss}}{\lambda r} = b\frac{\kappa s\gamma_c \ln\left(1+\frac{w_{eq}}{r}\right)}{1+\lambda r\kappa s\gamma_c j \ln\left(1+\frac{w_{eq}}{r}\right)} \tag{8.48}$$

and using the factor X, defined for the quadrupole as

$$X \equiv \kappa r s\gamma_c \ln\left(1+\frac{w_{eq}}{r}\right) \tag{8.49}$$

one has

$$G_{ss} = \frac{b}{r}\frac{X}{1+\lambda X} \qquad B_{p,ss} = b\frac{\lambda X}{1+\lambda X} \tag{8.50}$$

Example 8.6: The LHC main quadrupole is based on Nb-Ti conductor, has a 28 mm aperture radius, and two layers of 15.4 mm width cable. Knowing that it operates at a nominal gradient of 220 T/m, compute the loadline fraction and the conductor peak field at nominal gradient.

The factor X is given by

$$X = \kappa r s\gamma_c \ln\left(1+\frac{w_{eq}}{r}\right) = 0.24\times 28\times 600\times 0.00066\times \ln(1+31/28) = 1.94\,, \tag{8.51}$$

Note that using the mixed units (A/mm^2 for the current density and T/m for the gradient), one has to be very careful to use the coherent units. In this case the slope of the critical surface is A/(T mm^2), the constant is in γ_c~6.63×10^{-4} (T /m)/(A/mm^2), and the aperture radius has to be expressed in mm. The ratio between conductor peak field and gradient times aperture can be estimated as

$$\lambda \approx a_{-1}\frac{r}{w_{eq}} + 1 + a_1\frac{w_{eq}}{r} = 0.04\frac{28}{31} + 1 + 0.11\frac{31}{28} = 1.15 \tag{8.52}$$

amd the short sample gradient is

$$G_{ss} = \frac{b}{r}\frac{X}{1+\lambda X} = \frac{12.9}{0.028}\frac{1.94}{1+1.15\times 1.94} = 460\times 0.60 = 276 \text{ T/m} \tag{8.53}$$

and the loadline fraction is 220/276=0.80, very close to the actual value of 0.803 given in [1], pg. 187 table 7.13. The peak field in the conductor is

$$B_{p,ss} = \lambda G r = 1.15\times 220\times 0.028 = 7.1 \text{ T} \tag{8.54}$$

Example 8.7: The HL-LHC triplet quadrupole is based on Nb_3Sn conductor, has a 150 mm aperture, and two layers of 18.4 mm width cable. Knowing that it operates at a nominal gradient of 132 T/m, compute the loadline fraction and the conductor peak field at nominal gradient.

The factor X is given by

$$X = \kappa r s \gamma_c \ln\left(1+\frac{w_{eq}}{r}\right) = 0.28\times 75\times 400\times 0.00066\times \ln(1+37/75) = 2.2\,, \tag{8.51}$$

The ratio between conductor peak field and gradient times aperture can be estimated as

$$\lambda \approx a_{-1}\frac{r}{w_{eq}} + 1 + a_1\frac{w_{eq}}{r} = 0.04\frac{75}{37} + 1 + 0.11\frac{37}{75} = 1.14 \tag{8.52}$$

amd the short sample gradient is

$$G_{ss} = \frac{b}{r}\frac{X}{1+\lambda X} = \frac{19.4}{0.075}\frac{2.2}{1+1.14\times 2.2} = 259\times 0.63 = 163 \text{ T/m} \tag{8.53}$$

and the loadline fraction is 132/163=0.81, versus the value of 0.78 given in [2]. The peak field in the conductor is

$$B_{p,ss} = \lambda G r = 1.14\times 132\times 0.075 = 11.3 \text{ T} \tag{8.54}$$

as given in [2].

References

1. O. Bruning, et al. (eds), "The LHC design report Vol I" CERN-2004-003
2. P. Ferracin, et al., "Development of MQXF: the Nb3Sn low-b quadrupole for the Hilumi LHC", IEEE Trans. Appl. Supercond. 26 (2016) 4000207.
3. M. Anerella, et al., "The RHIC magnet system" Nucl. Instrum. Meth. A 499 (2003) 280-315
4. A. Milanese, et al., "Design of the EuCARD high field model dipole Fresca2" IEEE Trans. Appl. 22 (2012) 4002604
5. G. L. Sabbi, et al., "Design of HD2: a 15 tesla Nb_3Sn dipole with a 35 mm bore" IEEE Trans. Appl. Supercond. 15 (2005) 1128-1131
6. P. Fabbricatore, et al., "Development of a curved fast ramped dipole for FAIR SIS300", IEEE Trans. Appl. Supercond. 18 (2008) 232.
7. L. Rossi, E. Todesco, "Electromagnetic design of superconducting dipoles based on sector coils", Phys. Rev. STAB 10 (2007) 112401
8. T. Nakamoto, et al., "Model magnet development of D1 beam separation dipoles for the HL-LHC upgrade, IEEE Trans. Appl. Supercond. 25 (2015) 4000505.
9. A. Zlobin, et al. "Design concept and parameters of a 15 T Nb3Sn dipole demonstrator for a 100 TeV hadron collider" International Particle Accelerator Conference (2015) 3365-3367
10. J. Ahlback, et al., "Electromagnetic and mechanical design of a 56 mm aperture model dipole for the LHC", IEEE Trans. Magn. 30 (1994) 1746.
11. L. Rossi, E. Todesco, "Electromagnetic design of superconducting quadrupoles", Phys. Rev. STAB 9 (2006) 102401

Version 0.1

Ezio.Todesco@cern.ch

Chapter 9

The role of ferromagnetic iron, graded coils, and two-in-one dipoles

Plan of the chapter
Even though the field in a superconducting magnet for accelerators is mainly given by the current line contributions, ferromagnetic iron plays a significant role in the design. In 9.1 we outline the basics of ferromagnetic materials, and the concepts of magnetization and saturation; the requirements on the thickness of the iron to cancel the fringe field outside a magnet are discussed in Section 9.2. The contribution of iron in the case of non saturated ferromagnetic materials can be estimated for simple geometries via the current image method; simple equations for the field enhancement due to circular iron around a sector coil are derived in Section 9.3. The impact of this enhancement on the short sample field is analysed in Section 9.4; the contribution to field harmonics and the case of quadrupoles is discussed in Section 9.5. The effect of iron saturation on field quality, and the strategies to minimize it are outlined in Section 9.6. The case of double aperture magnets, as required in the LHC arcs or in the LHC recombination dipoles close to the insertions, is discussed in Section 9.7: approximated equations are given to estimate the impact of one aperture on the other one. In Section 9.8 we discuss the concept of graded coils that allows to have more efficient magnets, using higher current densities in the lower field regions. Finally, in Section 9.9 we discuss two case studies to apply the analytical approaches developed in Chapters 8 and 9.

9.1 Ferromagnetic materials, magnetization and saturation

A loop of current I encircling a surface S has a magnetic momentum defined as

$$m = IS \tag{9.1}$$

and its dimensions in the international system are A m^2. As outlined in Appendix B.8, electrons have an intrinsic magnetic momentum that can be well approximated (within 0.1%) by the Bohr magneton [1], defined as the momentum given by a charge e rotating at the speed of light on a circular orbit whose radius is the Compton radius of the electron

$$m_B \equiv \frac{ec}{2\pi r_C}\pi r_C^2 = \frac{ecr_C}{2} \qquad\qquad r_C \equiv \frac{h}{2\pi mc}, \tag{9.2}$$

and its value is order of 10^{-23} A m^2:

$$m_B \equiv \frac{ecr_C}{2} = \frac{eh}{4\pi m} = \frac{1.60\times10^{-19}\times6.63\times10^{-34}}{4\pi\times9.11\times10^{-31}} = 9.3\times10^{-24}\ \text{A m}^2. \tag{9.3}$$

A density of magnetic momentum in the matter gives rise to a magnetization

$$M \equiv \frac{dm}{dV} \tag{9.4}$$

whose units are A/m. The magnetization is related to the external magnetic field H (in A/m) via the susceptivity χ

$$M = \chi H \tag{9.5}$$

and gives rise to an additional magnetic field in the matter according to

$$B_{matter} = \mu_0 M. \tag{9.6}$$

The total magnetic field is given by

$$B = \mu_0 \left(H + M \right) = \mu_0 \left(1 + \chi \right) H \equiv \mu_0 \mu_r H \equiv \mu H \tag{9.7}$$

where we defined the permeability constant in the matter μ and the relative permeability constant μ_r

$$\mu \equiv \left(1 + \chi \right) \qquad \mu_r \equiv \frac{\mu}{\mu_0} \tag{9.8}$$

According to the value of the susceptivity, and one distinguishes between diamagnetic materials, paramagnetic materials and ferromagnetic materials [2].

In the case of ferromagnetic materisl, the relation between the external field and the magnetization is shown in Fig. 9.1, right, and is characterized by an initial linear regime, and a saturation, corresponding to the condition where all magnetic momenta are aligned. The order of magnitude of the magnetic field in the saturation condition can be estimated by computing the volume of an iron atom

$$\upsilon = \frac{m_{Fe}}{\rho_{Fe}} = \frac{55.8 \times 1.67 \times 10^{-27}}{7847} = 1.2 \times 10^{-29} \, \mathrm{m}^3 \tag{9.9}$$

and therefore assuming that each atom has one Bohr magneton to align, the saturation field can estimated as order of a tesla

$$B_{sat} \approx \mu_0 \frac{m_b}{\upsilon} = 4 \times \pi \times 10^{-7} \frac{9.3 \times 10^{-24}}{1.2 \times 10^{-29}} = 1.0 \ \mathrm{T}. \tag{9.10}$$

Most ferromagnetic materials have a saturation field of 1.8-2 T, and some special iron alloys can reach up to 2.2 T.

Magnetization is also characterized by hysteresis, i.e. the presence of a residual magnetic moment when the external field is ramped up and then brought back to zero (see Fig. 9.1, right). This introduces a complexity in the operation of magnets at very low current, since in this domain the behaviour is not linear, and is path dependent. The superconductor itself is another source of residual magnetization, and even in absence of iron superconducting magnets behave in a nonlinear way in the region of very low current. This sets limits on the lowest operational current of main magnets in an accelerator, and therefore on the energy swing of the collider; Tevatron, RHIC, HERA and LHC operate/d with an energy swing between 6 and 25.

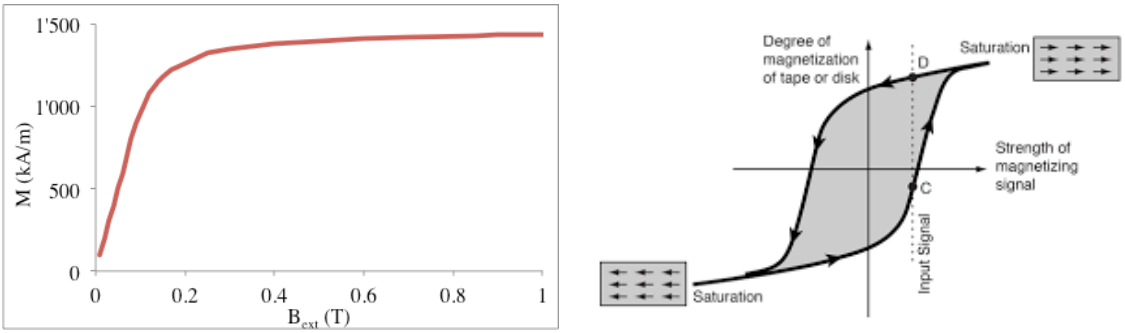


Fig. 9.1: Magnetization as a function of the external magnetic field (left); hysteresis loop (right)

9.2 The role of ferromagnetic iron in superconducting magnets: shielding conditions

Ferromagnetic iron is used in the design of superconducting magnets for threee different scopes:

- It provides a return flux to the magnetic field; this allows to avoid having large (>100 mT) magnetic fields in the neighbourhoods of the magnet, as discussed in this section.
- It enhances the magnetic field produced by the coils; this aspect will be discussed in the next section.
- In some cases, it partially or totally contributes to the mechanical structure, as discussed in chapters 10 and 11.

The requirements on the magnetic field in the space surrounding the accelerator magnets mainly stems from the presence of instrumentation. This is not a hard requirement (for instance this limit is not specified for the LHC tunnel), and is of the order of 1-100 mT; it is 10 mT for the HL-LHC [3], and 100 mT for the FCC [4]. To totally avoid any leak of field outside the magnet, one has to have a iron thickness large enough to allow the return flux going through the iron, with the conditions that the field in the iron does not exceed the saturation field B_{sat} (see Fig. 9.2).

$$rB = t_{Iron} B_{sat} \qquad t_{Iron} = \frac{rB}{B_{sat}} \tag{9.11}$$

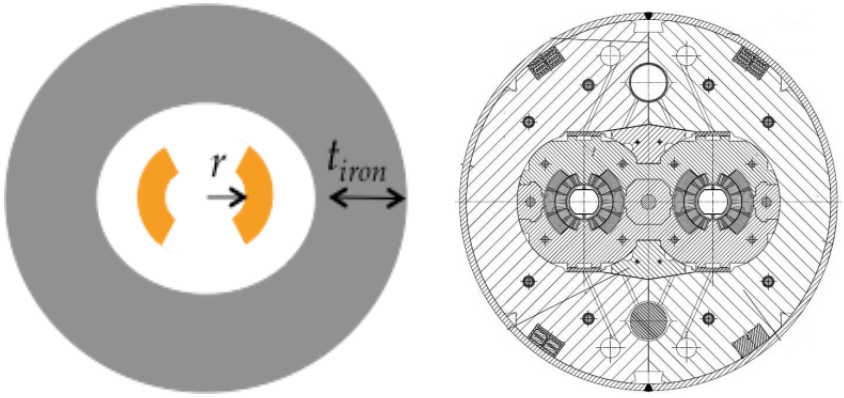


Fig. 9.2: Shielding condition for a dipole (left) and LHC dipole cross-section (right).

In a quadrupole, the condition reads as

$$\int_0^r Gr\,dr = \frac{Gr^2}{2} = t_{Iron} B_{sat} \qquad t_{Iron} = \frac{Gr^2}{2B_{sat}} \tag{9.12}$$

Example 9.1: The LHC main dipole [5,6] has a 8.3 T field and 56 mm aperture. Estimate the thickness of the iron needed in the midplane to have a complete shielding, and compare to the design value, knowing that interbeam distance is d_i=192 mm, and collar thickness is w_c=40 mm, and the magnet diameter is D=570 mm (see Fig. 9.2, right).

Using 2 T as saturation field, the iron thickness can be estimated as

$$t_I = \frac{rB}{B_{sat}} = \frac{28 \times 8.3}{2} = 116 \text{ mm} \tag{9.13}$$

that is not far from the actual thickness used in the design

$$t_I = \frac{D - d_i}{2} - r - w - w_c = \frac{570 - 192}{2} - 28 - 31 - 40 = 90 \text{ mm}. \tag{9.14}$$

Example 9.2: The triplet quadrupole for the HL-LHC [7] has 150 mm aperture, and 132 T/m gradient. Compute the thickness of the iron needed for a total shielding.

$$t_I = \frac{r^2 G}{2B_{sat}} = \frac{0.075^2 \times 132}{2 \times 2} = 0.185 \text{ m} = 185 \text{ mm} \tag{9.15}$$

Note that for quadrupoles the use of T/m for the gradient and mm for distances can be extremely tricky. Here we used meters to apply equation (9.12), and then converted in mm.

9.3 Field contribution due to ferromagnetic iron in the linear regime via the current images method

In the case of a sector coil, the presence of the iron at a distance R_I gives rise to an additional contribution to the magnetic field in the aperture that can be estimated via the current images method. As shown in Fig. 9.3, right, the iron creates for each current line in the coil at distance ρ, an image current line at distance ρ' carrying the current I'

$$\rho' = \frac{R_I^2}{\rho} \qquad I' = \frac{\mu_r - 1}{\mu_r - 1} I \qquad I' = I \text{ for } \mu_r >> 1 \tag{9.16}$$

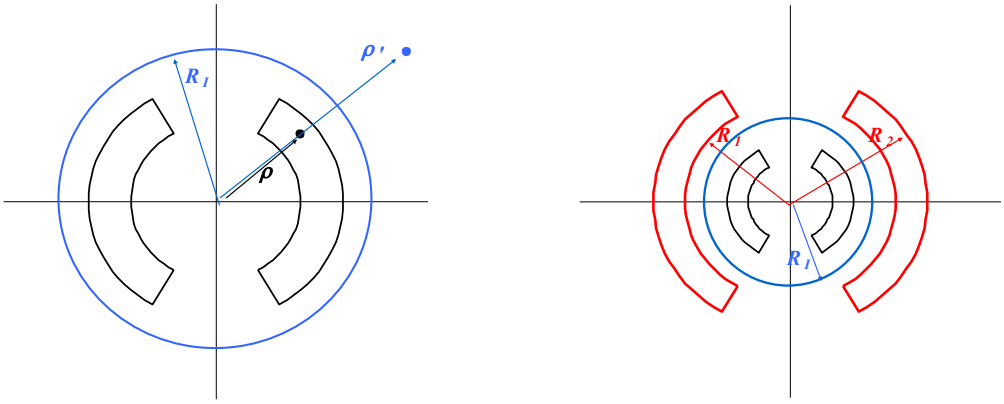


Fig. 9.3: Current image produced by non saturated iron (left) and virtual coil (left).

We can estimate the additional contribution to the field given by the image coil (see Fig. 9.3, right) as

$$\Delta B = \gamma_c j' \left(R_2 - R_1 \right). \tag{9.17}$$

and therefore one has

$$\frac{\Delta B}{B} = \frac{j' \left(R_2 - R_1 \right)}{jw} = \frac{j' R_I^2}{jw} \left(\frac{1}{r} - \frac{1}{r+w} \right) = \frac{j' R_I^2}{j} \frac{1}{r\left(r+w\right)}. \tag{9.18}$$

The current density in the virtual coil is lower than the current density in the coil since the area of the vitrual coil is larger, and the total current shall be preserved:

$$j' \left(R_2^2 - R_1^2 \right) = j \left[(r+w)^2 - r^2 \right] \tag{9.19}$$

$$j' \left(\frac{1}{r^2} - \frac{1}{\left(r+w\right)^2} \right) R_I^4 = j \left[(r+w)^2 - r^2 \right] \tag{9.20}$$

$$j' \frac{1}{r^2 \left(r+w\right)^2} R_I^4 = j \qquad \frac{j' R_I^2}{j} = \frac{r^2 \left(r+w\right)^2}{R_I^2} \tag{9.21}$$

and therefore replacing (9.21) in (9.18) one lands on a very simple expression

$$\frac{\Delta B}{B} = \frac{j' R_I^2}{j} \frac{1}{r\left(r+w\right)} = \frac{r^2 \left(r+w\right)^2}{R_I^2 r \left(r+w\right)} = \frac{r\left(r+w\right)}{R_I^2} \tag{9.22}$$

One can make the following remarks:

- Since the iron is placed at distance that is always larger than the outer radius of the coil, $R_I > r+w$ and therefore the iron contribution can at most double the contribution of the coil;
- The highest contribution happens for $R_I = r+w$ (i.e. there is no non-ferromagnetic material between the coil and the iron), and for thin coil widths $w \Rightarrow 0$;
- For very large coil widths, the iron contribution becomes less and less important.

In Fig. 9.4 we show the increase of field due to the iron for few accelerator dipole magnets, as a function of the ratio coil width/aperture radius, and of the ratio iron radius / aperture radius.

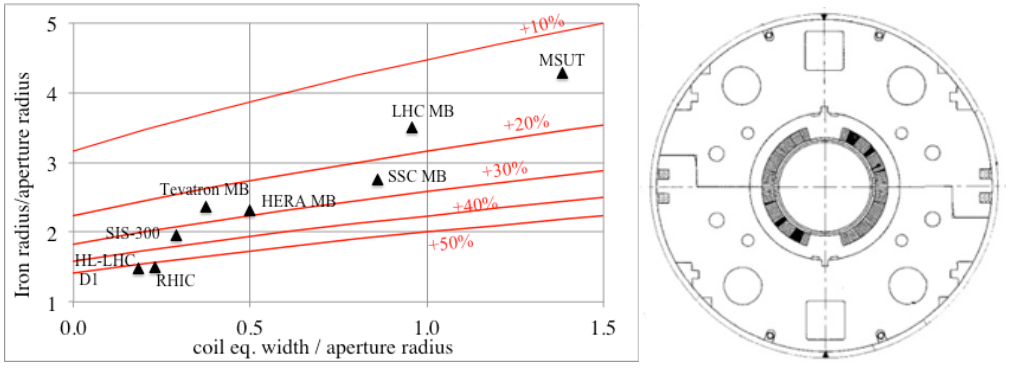


Fig. 9.4: Field increase due to iron (lines in red, and values expressed in %) in the space of iron radius versus coil width, both normalized to the aperture radius (left); RHIC dipole cross-section (right)

Example 9.3: The LHC dipole has 28 mm aperture radius, and 31 mm coil width. The iron is placed at 98 mm from the aperture centre. Estimate the increase of the field due to iron.

One finds that

$$\frac{\Delta B}{B} = \frac{r\left(r+w\right)}{R_I^2} = \frac{28\left(28+31\right)}{98^2} = 0.17\,. \tag{9.23}$$

Using the equivalent coil width (27 mm instead of 31 mm, see section 9.8), one obtains 0.16. The numerical value obtained via finite element codes in the twin aperture case is 16%, but this very good agreement is misleading. This is to be compared to the numerical model of the single aperture, that provides 0.21; equation (9.22) is quite accurate for non graded cases, but is less precise in the case of grading (see Section 9.8). For instance in the SSC dipole case (50 mm aperture), the analytical approach gives 24% increase due to the iron, and the numerical model provides 31%. A generalization of (9.22) to include grading can be done, but it goes beyond the aim of these notes.

Example 9.4: The RHIC dipole [8] has 80 mm aperture, and a single layer coil with a 10-mm-width cable, see Fig. 9.4, right. There is a 10-mm-width plastic spacer bewteen the coil and the iron. Estimate the iron contribution to the main field.

One finds that

$$\frac{\Delta B}{B} = \frac{r\left(r+w\right)}{R_I^2} = \frac{40\left(40+10\right)}{60^2} = 0.56\,. \tag{9.24}$$

The numerical value is 55%.

9.4 Field contribution of the iron to the short sample field

Here we show how to extend the equations given in Chapter 4 and 8 for the shor sample field to a magnet design that includes iron; defining the iron contribution as in (9.22):

$$\Delta_I \equiv \frac{\Delta B}{B} = \frac{r\left(r+w\right)}{R_I^2} \tag{9.25}$$

we have to replace in the equations of Chapter 8

$$\gamma_c \rightarrow \gamma_c\left(1+\Delta_I\right); \tag{9.26}$$

therefore the equation for the field as a fuction of coil width and current density becomes

$$B = \gamma_c\left(1+\Delta_I\right) jw\,, \tag{9.27}$$

the factor X becomes

$$X \equiv ks\gamma_c\left(1+\Delta_I\right) w_{eq}\,, \tag{9.28}$$

and the short sample conditions are given by

$$B_{ss} = b\frac{X}{1+\lambda X} = b\frac{ks\gamma_c\left(1+\Delta_I\right) w_{eq}}{1+\lambda ks\gamma_c\left(1+\Delta_I\right) w_{eq}}\,. \tag{9.29}$$

In case of a small impact of the iron, one can use the sensitivity analysis discussed in Chapter 8.6:

$$\frac{\Delta B_{ss}}{B_{ss}} = \frac{\Delta X}{X}\frac{1}{1+\lambda X} = \Delta_I \frac{1}{1+\lambda X} \tag{9.30}$$

therefore the relative field increase due to iron Δ_I gives a relative increase in the short sample field that is reduced by the factor $1/(1+\lambda X)$.

Example 9.5: Estimate the short sample field increase due to the iron for the LHC main dipole and for the RHIC main dipole using the linear approximation (9.30).

For the LHC main dipole, Δ_i=0.17 and X=3.0, λ=1.04:

$$\frac{\Delta B_{ss}}{B_{ss}} = 0.17\frac{1}{1+1.04\times 3.0} = 0.041\,. \tag{9.31}$$

For the RHIC main dipole, the coil width is 10 mm, filling factor is 0.226 (see table 8.1) and therefore the factor X in absence of iron is

$$X \equiv ks\gamma_c w_{eq} = 0.226\times 600\times 0.00066\times 10 = 0.90\,. \tag{9.32}$$

The ratio peak field/bore field is

$$\lambda = 1+0.04\times\frac{40}{10} = 1.16\,, \tag{9.33}$$

the field increase due to iron is Δ_i=0.56 as shown in example 9.4 and therefore the short sample increase is about half of it:

$$\frac{\Delta B_{ss}}{B_{ss}} = 0.56\frac{1}{1+1.16\times 0.90} = 0.27\,. \tag{9.34}$$

9.5 Field contribution of the iron to field harmonics, and the quadrupole case

Another very simple formula can be derived for the iron contribution to the non normalized multipoles:

$$\frac{\Delta B_n}{B_n} = \left[\frac{r\left(r+w\right)}{R_I^2}\right]^n\,. \tag{9.35}$$

This equation shows that since $r(r+w)/R_I^2$ is smaller than one, contribution to high order multipoles becomes less and less important. Therefore, when optimizing field quality, one has in general to account for the iron only for b_3 and neglect it for higher orders.

In the case of a quadrupole, the gradient increase due to iron is given by

$$\frac{\Delta G}{G} = \left[\frac{r\left(r+w\right)}{R_I^2}\right]^2\,. \tag{9.36}$$

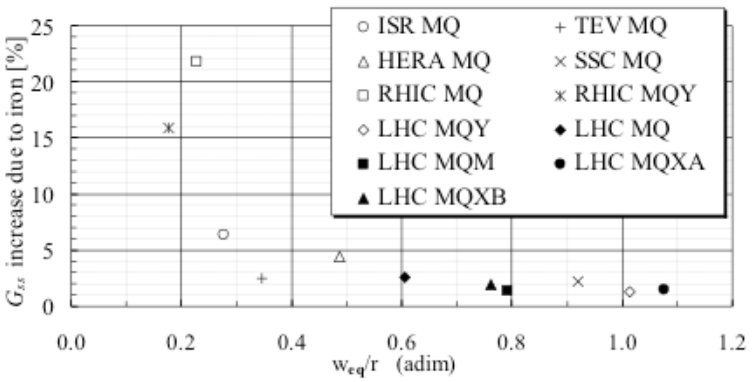


Fig. 9.5: Gradient increase due to iron versus ratio between coil width and magnet aperture radius.

Example 9.6: Estimate the gradient increase due to iron for LHC main quadrupole, that has a 56 mm aperture, a double layer coil with same cable as the LHC dipole outer layer, and 30 mm thick collars; and compare to the data given in Fig. 9.5.

The distance of the iron from the aperture is 28+31+30=89 mm; therefore one has

$$\frac{\Delta G}{G} = \left[\frac{r\left(r+w\right)}{R_I^2}\right]^2 = \left[\frac{28\left(28+31\right)}{89^2}\right]^2 = 0.21^2 = 0.044 \tag{9.37}$$

i.e., 4.4%, compared to 3.5% obtained with numerical tools.

9.6 Impact of iron saturation on field quality

The saturation of iron induces a non-linear dependence of the field on the current; the transfer function, defined as the bore field divided by the current, at low fields is increased with respect to the current lines contribution by the factor $(1+\Delta_I)$. Going to higher fields, the iron saturates and its relative contribution gradually vanishes: the trasnfer function decreases from its low field value, including the $(1+\Delta_I)$ enchancement, to its high field value where the iron contribution is not any more proportional to the current but it is just an constant. The effect is shown in Fig. 9.6 for the main LHC dipole, that at nominal current of 11850 A has a transfer function that is 0.7% smaller than at low currents. Going to very high currents (well above the magnet short sample), the reduction of the transfer function tends asymptotically to the contribution of the iron, i.e. to 17%. In Fig. 9.6 the low field region has a branching of the transfer function due to the hysteresis of the magnetization of the superconductor (usually named persistent currents, see Chapter 5). We also show a rather extreme case of saturation, i.e. the case of HL-LHC D1 [9]. This is a magnet with 5.6 T bore field, with the iron placed close to a single layer coil. In this case the saturation is order of 10%, i.e. 15 times larger than in the LHC dipole.

The saturation curves of the transfer functions can be precisely modelled via finite element models using commercial codes as OPERA [10] or ANSYS [11]. The software package ROXIE [12,13] developed at CERN since the 90's to design superconducting magnets for accelerators makes use of a boundary elements-finite elements methods; they allows meshing only the regions where the nonlinear materials (iron) are present, thus reducing the computational time. Saturation of the transfer function has to be included in the software that sets the currents in the magnets during the accelerator operation. This can be done via tables generated on models or measurements, or using fitting functions as defined in the LHC field model (FiDeL [14]). The functions used for fitting the S-shaped saturation are based on the arcotangent, or on the integral of a Gaussian (erf function).

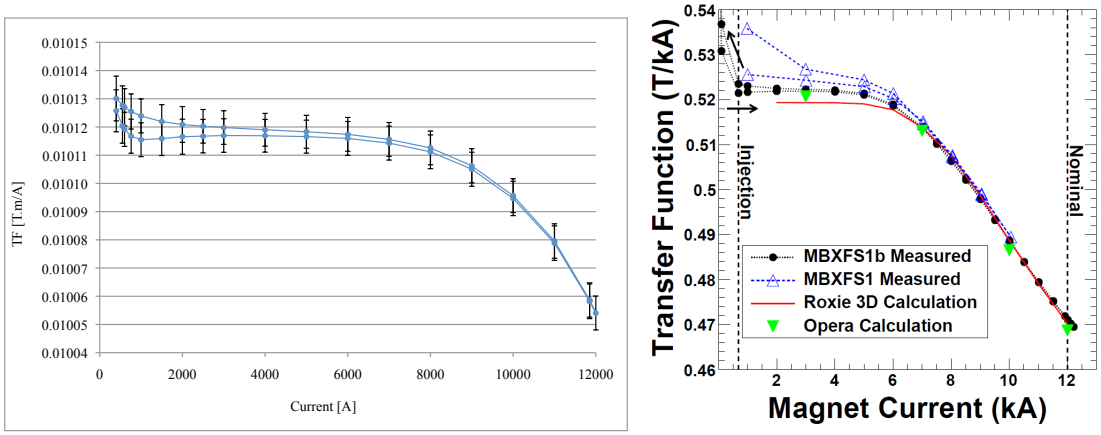


Fig. 9.6: Saturation of transfer function for LHC main dipole (left) and for HL-LHC D1 (right).

Iron saturation has a strong impact on the low order mutipoles, in particular on b_3 and b_5. The physical reason lays in the non-uniform dependence of the field on the angle around the aperture, the field being more intense at 90° and 270°, and being lower and with reversed sign in the midplane at 0° and 180°. Due to the dependence of b_3 and b_5 on these angles, saturation generates a significant variation, that spoils the careful compensation provided by the arrangement of the blocks as described in Chapter 4. To correct this effect, the iron can be shaped with holes. In Fig. 9.7 we show the dependence of b_3 and b_5 on the current for the case of the HL-LHC separation dipole D1. One can see that the presence of holes in the iron allows to reduce the swing in b_3 due to saturation from 40 to 10 units.

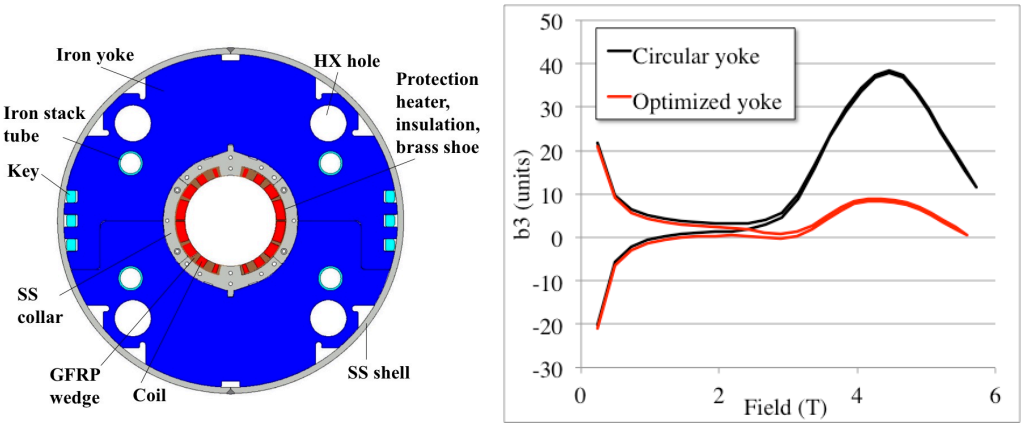


Fig. 9.7: Cross-section of the HL-LHC D1 (right) and dependence of b_3 on current before and after optinization of the iron shape [15].

9.7 Two-in-one dipoles

Colliders making use of the same type of particles, as LHC, RHIC or SSC, requires two separate apertures having the dipolar field in the opposite direction. For the LHC the solution of having single dipoles providing opposite fields in the two side-by-side apertures has been used, with the main advantage of having more compact magnets and cheaper costs. The presence of the other aperture increases the field; this is obviously a second order effect with respect to the main contribution given by the coils around the aperture, as the other aperture is far and moreover the two coils of the other aperture give contributions with opposite sign. A sketch of the geometry of the LHC dipoles is given in Fig. 9.8, and this is the base of our analytical estimate of this design feature.

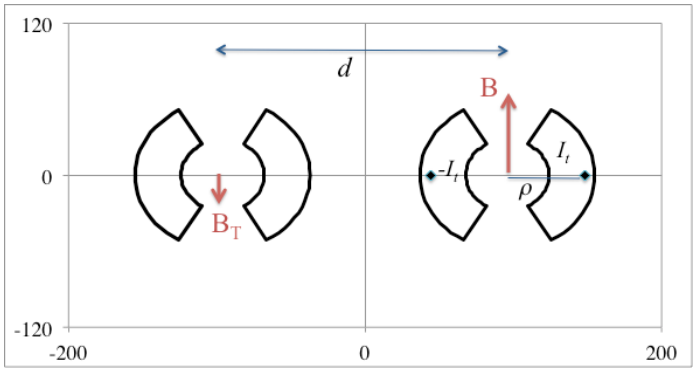


Fig. 9.8: Schematics of the field variation in a double aperture dipole induced in the left aperture by the right aperture, geometry of the LHC main dipoles, distances in mm.

Having a dipole with aperture radius r, coil width w and current density j, the field and the total current flowing the the coils are given by

$$B = \gamma_c j w \qquad I_t = \frac{2\pi j}{3}\left[\left(r+w\right)^2 - r^2\right]. \tag{9.38}$$

To have a first order estimate of the impact of one aperture on the other one, we can replace the sector coil with two current lines at distance ρ providing the same field:

$$B = \frac{\mu_0 I_t}{2\pi\rho} - \frac{\mu_0\left(-I_t\right)}{2\pi\rho} = \frac{\mu_0 I_t}{\pi\rho} \tag{9.39}$$

therefore the distance ρ of the equivalent current line is

$$\rho = \frac{\mu_0 I_t}{\pi B} \tag{9.40}$$

The effect of the other aperture can be estimated via the Biot-Savart contribution of two current lines:

$$\Delta B_T = \frac{\mu_0 I_t}{2\pi}\left(\frac{1}{d-\rho} - \frac{1}{d+\rho}\right) = \frac{\mu_0 I_t}{\pi}\frac{\rho}{d^2-\rho^2} = B\frac{\rho^2}{d^2-\rho^2} \tag{9.41}$$

and one obtains the very simple expression

$$\Delta_T \equiv \frac{\Delta B_T}{B} = \frac{\rho^2}{d^2-\rho^2}. \tag{9.42}$$

In the interaction region, an accelerator based on twin aperture magnets requires separation and recombination dipoles to bring the beams in collisions (see section 2.4). The latter ones are double aperture dipoles, with the field in the same direction. In this case the field contribution of the other aperture is negative, as shown in example 9.8.

Example 9.7: Estimate the field increase due to the twin aperture design for the LHC main dipole, knowing that has 40 turns with 11850 A operational current, 8.3 T central field, and 192 interbeam distance.

The distance of the equivalent current line is estimated as

$$\rho = \frac{\mu_0 I_t}{\pi B} = \frac{4\pi\times10^{-7}\times40\times2\times11850}{\pi\times8.3} = 0.046\ \text{m} = 46\ \text{mm} \tag{9.43}$$

and the relative field increase in the other aperture is

$$\Delta_T = \frac{\rho^2}{d^2 - \rho^2} = \frac{45^2}{192^2 - 45^2} = 0.058 = 5.8\% \tag{9.44}$$

that is order of 0.5 T. Note that since the LHC dipole has a factor X of ~3, the presence of the other aperture increases to short sample field of order of 1.5%.

Example 9.8: The LHC D2 recombination dipole [16] has 105 mm aperture, 31 turns with 12000 A, and an interbeam distance of 188 mm. Estimate the reduction of the field induced by the other aperture.

The distance of the equivalent current line is estimated as

$$\rho = \frac{\mu_0 I_t}{\pi B} = \frac{4\pi \times 10^{-7} \times 31 \times 2 \times 12000}{\pi \times 4.5} = 0.066 \text{ m} = 66 \text{ mm} \tag{9.45}$$

and the relative field increase in the other aperture is

$$\Delta_T = -\frac{\rho^2}{d^2 - \rho^2} = -\frac{66^2}{188^2 - 66^2} = -0.12 = -12\% \tag{9.46}$$

that is, order of 0.6 T.

9.8 Graded coils, and material grading

Graded coils make use of higher current density in the regions where the field is lower, resulting in a more efficient use of conductor. In most of the cases, the higher current density is obtained via using a cable with a lower cross-section surface in the low field regions, and having it in series with the cable used in the high field region. This adds a complexity in the coil fabrication/assembly, since it requires a splice between different superconducting cables.

In the case of inner layer with coil width w_i and current density j_i, and outer layer with coil width w_o and larger current density $j_o > j_i$, the equation for the bore field are given by

$$B = \gamma_c \left(j_i w_i + j_o w_o \right). \tag{9.47}$$

with γ_c=0.00066 T mm/A, and the grading is defined as

$$g \equiv \frac{j_o}{j_i} > 1. \tag{9.48}$$

The efficiency of the grading can be estimated in two different ways:

- Given the short sample field, the reduction of the mass of the conductor;
- Given the mass of the conductor, the increase in the short sample field.

In the case with $w_i = w_o = w$ the mass of conductor of the graded coil is proportional to

$$M_g \propto \left(r + 2w \right)^2 - r^2 \tag{9.49}$$

and if we use no grading the same field is obtained by increasing the outer coil width by a factor g, and the conductor mass is proportional to

$$M_{ng} \propto \left(r + w + gw\right)^2 - r^2 ; \tag{9.50}$$

therefore the saving in the conductor mass for the same short sample field is

$$\frac{M_g}{M_{ng}} = \frac{\left(r + 2w\right)^2 - r^2}{\left(r + w + gw\right)^2 - r^2} . \tag{9.51}$$

On the other hand, grading allows increasing the achieved field by

$$\frac{\Delta B_{g-ng}}{B_{ng}} = \frac{\gamma_c \left(wj + wgj\right)}{\gamma_c (wj + wj)} - 1 = \frac{1+g}{2} - 1 = \frac{g-1}{2} \tag{9.52}$$

and therefore the short sample field sensitivity on the grading is

$$\frac{\Delta B_{ss,g-ng}}{B_{ss,ng}} = \frac{\Delta B_{g-ng}}{B_{ng}} \frac{1}{1+\lambda X} = \frac{g-1}{2} \frac{1}{1+\lambda X} . \tag{9.53}$$

The example of the LHC dipole is shown in Fig. 9.9. The inner coil operates at 380 A/mm^2. The outer layer makes use of a cable with 19% smaller surface, is in series with the inner layer, resulting in a current density 23% larger than in the inner dipole, i.e. 470 A/mm^2. With these current densities, the peak field in the inner layer is 8.7 T, and the magnet operates in the inner layer at 86% of the short sample conditions. In the outer layer the peak field is 6.5 T, and therefore the outer layer is at 76% of the short sample (note that in principle the grading could be increased to reach the same margin on the loadline on both layers). The saving of conductor for a fixed short sample field is

$$\frac{M_g}{M_{ng}} = \frac{\left(r + 2w\right)^2 - r^2}{\left(r + w + gw\right)^2 - r^2} = \frac{\left(28 + 2\times 15.4\right)^2 - 28^2}{\left(28 + 15.4 + 1.23\times 15.4\right)^2 - 28^2} = 0.88 \tag{9.54}$$

i.e., thanks to grading 12% less conductor has been used. The increase of short sample field induced by the grading is

$$\frac{B_{ss,g}}{B_{ss,ng}} = \frac{g-1}{2} \frac{1}{1+\lambda X} = \frac{1.23-1}{2} \frac{1}{1+1.04\times 3.6} = 0.024 . \tag{9.55}$$

In Table 9.1 we summarize the grading parameters for some superconducting dipoles: the Nb-Ti main dipoles of the SSC [17] and of the LHC [7,8], and the MSUT [18] and D20 [19] Nb_3Sn short model dipoles. In quadrupoles, grading has been also used in the same layer (inner grading), see the LHC MQY [20] and the LHC MQXA [21] as examples in Fig. 9.10, where the second layer has different conductors.

Table 9.1: Main parameters of grading in four superconducting dipoles.

Magnet	g (adim)	M_{ng}/M_g (adim)	$B_{g\text{-}ng}/B$ (%)	$B_{ss,g\text{-}ng}/Bss$ (%)	w_i (mm)	Mid th. i (mm)	w_o (mm)	Mid th. o (mm)	r (mm)	X (adim)
LHC MB	1.23	0.86	11	2.5	15.400	2.140	15.400	1.740	28	3.6
SSC MB	1.30	0.83	15	3.5	12.803	1.614	11.845	1.343	25	3.3
MSUT	1.65	0.70	33	6.3	22.060	2.512	16.670	2.013	25	4.2
D20	1.81	0.62	40	3.8	29.380	1.609	23.740	1.103	25	9.5

A second type of grading consists in using the same current density, using different materials according to the peak field in each part of the coil: it can be defined as material grading in hybrid

magnets. The advantage is to use cheaper materials in regions where the field is lower (for instance, Nb-Ti in the outer layers where the field is lower than 8 T, and Nb_3Sn where the field is larger than 8 T); the complexity is given by mechanical aspects related to having coils made with different materials (plus the splice). In Fig. 9.11 we show an hypothetical grading applied to the MQXF Nb_3Sn quadrupole. Since in the outer layer midplane block the peak field in operational conditions is 7 T, one could use Nb-Ti in this region, and since the cost of Nb-Ti is 6 times larger than Nb_3Sn used, and the outer layer midplane block has 16 conductors out of a total of 50, one would save about 27% in the conductor cost.

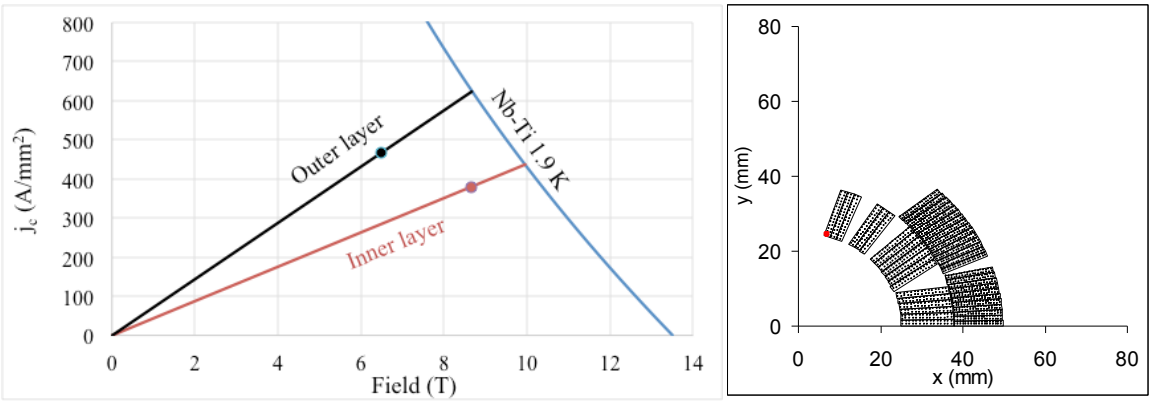


Fig. 9.9: Grading in the LHC dipole: loadline, critical surface and operational points (left) and coil cross-section, one quarter shown (right).

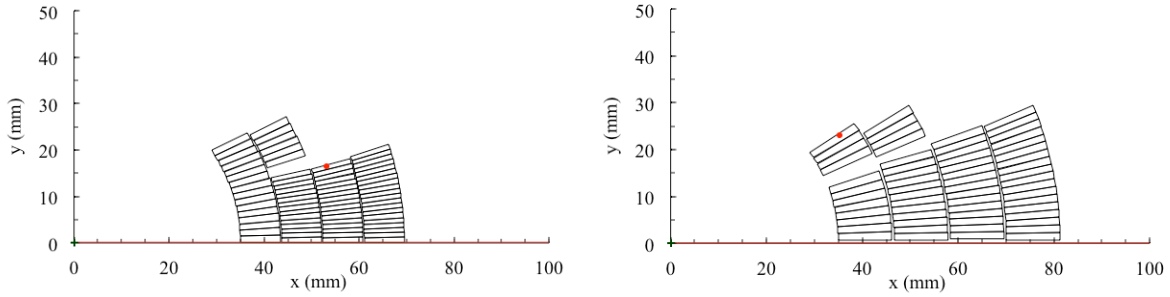


Fig. 9.10: Grading in the LHC MQY quadrupole (left) and in the LHC MQXA quadrupole (right)

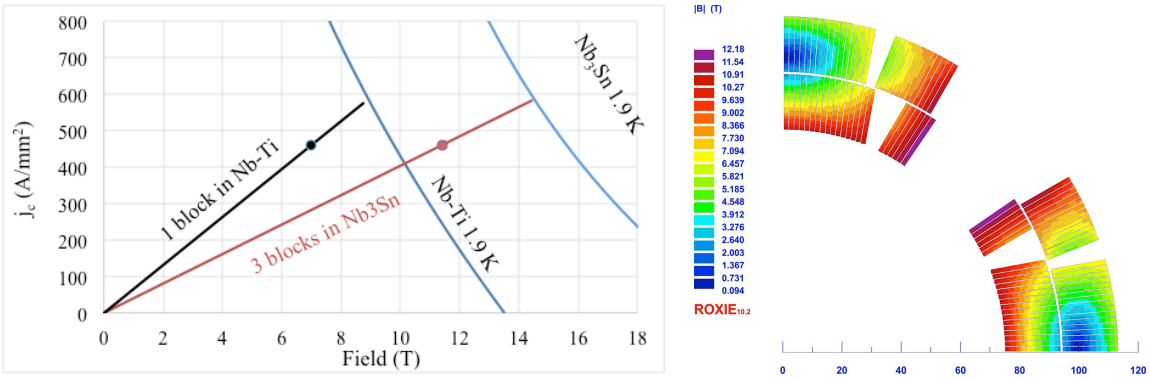


Fig. 9.11: Material grading: a case study using the MQXF quadrupole

9.8 Two case studies

In this last section we apply the several elements described in chapter 8 and 9 to provide a full analytical model of two cases, including the concepts of equivalent coil width, graded coils, iron and two-in-one contribution.

We start with the D2 recombination dipole [16], a double aperture magnet with 105 mm aperture diameter, using 31 turns of a cable whose insulated width is 15.4 mm and the cross-sectional surface is 26.8 mm^2. Operational current is 12000 A, and the distance of the iron to the aperture can be approximated by 110 mm, knowing that it is squarish rather than circular. The aim of the exercise is to estimate the central field, via the different contributions of coil, iron and double aperture.

Opertional overall current density is 12000/26.8=450 A/mm^2. Equivalent coil width is about 15% smaller than the cable width

$$w_{eq} = r\left[\sqrt{\frac{3A}{2\pi r^2}+1}-1\right] = 52.5\left[\sqrt{\frac{3\times 4\times 31\times 26.8}{2\pi 52.5^2}+1}-1\right] = 13.4 \text{ mm}. \tag{9.56}$$

Such a coil in a standing alone mode gives

$$B = \gamma_c wj = 0.0066\times 13.4\times 450 = 4.0 \text{ T}. \tag{9.57}$$

The double aperture contribution can be estimated via the position of the equivalent current line

$$\rho = \frac{\mu_0 I_t}{\pi B} = \frac{4\pi\times 10^{-7}\times 31\times 2\times 12000}{\pi\times 4.0} = 0.075 \text{ m} = 75 \text{ mm}. \tag{9.58}$$

Note that in the estimate of example 9.8 we obtained 66 mm, since we started from the input of the dipole field of 4.5 T. Using this equivalent current line, the double aperture reduction is

$$\Delta_T = -\frac{\rho^2}{d^2-\rho^2} - \frac{75^2}{188^2-75^2} = -0.19 = -19\% \tag{9.59}$$

and it is largely compensated by the field increase given by the iron:

$$\Delta_I = -\frac{r(r+w)}{R_I^2} = \frac{52.5(52.5+15.4)}{100^2} = 0.36. \tag{9.60}$$

The total field is

$$B = \left(1+\Delta_I\right)\left(1+\Delta_T\right)\gamma_c wj = 1.36\times 0.81\times 4.0 = 4.4 \text{ T} \tag{9.61}$$

versus a numerical value of 4.5 T.

The second case we consider is the main LHC dipole. It makes use in the inner layer of 15 turns of a cable whose insulated width is 15.4 mm and the cross-sectional surface is 32.96 mm^2. In the outer layer 25 turns of a thinner cable are used, with the same width and a cross-sectional surface of 26.80 mm^2. Operational current is 11850 A, and the distance of the iron to the aperture is 98 mm (see Fig. 9.8). Also in this case, we estimate (i) the central field, via the different contributions of coil, iron and double aperture, and (ii) the loadline fraction.

Opertional overall current density is 11850/32.96=360 A/mm^2 in the inner layer and 11850/26.80=442 A/mm^2 in the outer layer. The equivalent coil width of the inner layer is

$$w_{eq,i} = r\left[\sqrt{\frac{3A_i}{2\pi r^2}+1}-1\right] = 28\left[\sqrt{\frac{3\times 4\times 15\times 32.96}{2\pi\times 28^2}+1}-1\right] = 13.4 \text{ mm} \tag{9.62}$$

about 10% smaller than the cable width. For the outer layer the coil width is

$$\begin{aligned} w_{eq,o} &= (r+w_{eq.i})\left[\sqrt{\frac{3A_o}{2\pi(r+w_{eq.i})^2}+1}-1\right] = \\ &(28+13.4)\left[\sqrt{\frac{3\times 4\times 25\times 26.80}{2\pi\times(28+13.4)^2}+1}-1\right] = 13.3 \text{ mm} \end{aligned} \tag{9.63}$$

Therefore, the field given by the current line contribution is

$$B = \gamma_c\left(j_i w_{eq,i} + j_o w_{eq,o}\right) = 0.00066\left(360\times 13.4 + 442\times 13.3\right) = 3.2+3.9 = 7.1 \text{ T} \tag{9.64}$$

The iron contribution estimated using the equivalent coil width of 26.7 mm is 16% (see Eq. 9.23), i.e. 1.1 T; the two-in-one contribution is 5.8 % (see Eq. 9.44), i.e. 0.4 T. The total field can be estimated as

$$B = \left(1+\Delta_I\right)\left(1+\Delta_t\right)\gamma_c\left(j_i w_{eq,i} + j_o w_{eq,o}\right) = 1.16\times 1.058\times 7.1 = 8.7 \text{ T}. \tag{9.65}$$

Including iron and two-in-one effects, the factor X can be estimated as

$$X \equiv \left(1+\Delta_I\right)\left(1+\Delta_t\right)\kappa s\gamma_c w_{eq} = 1.16\times 0.28\times 1.058\times 600\times 0.00066\times 26.7 = 3.67. \tag{9.66}$$

Therefore, the short sample conditions are given by

$$B_{ss} = b\frac{X}{1+\lambda X} = 12.9\frac{3.67}{1+1.04\times 3.67} = 10.1 \text{ T} \tag{9.67}$$

and the loadline margin is 8.7/10.1=86%. Note that using numerical methods one finds 9.7 T for the short sample, 8.33 T for the field at 11850 A, and 0.86 of loadline fraction.

References

1. M. Born, “Atomic physics” Blackie & Son (London, 1937) 48
2. E. M. Purcell, et al., “Electricity and magnetism (Berkeley Physics Course, Vol. 2)” MacGraw-Hill (New York, 1973)
3. E. Todesco, et al., "A first baseline for the magnets in the high luminosity LHC insertion regions" IEEE Trans. Appl. Supercond. 24 (2014) 4003305
4. A. Abada, et al., “FCC-hh: the hadron collider” Eur. Phys. J. Spec. Top. 228 (2019) 755-1107
5. R. Perin, in Encyclopedia of Applied Superconductivity (IOP, London, 1998), pp. 919–950.
6. L. Rossi, et al., “The LHC main dipoles and quadrupoles toward series production”, IEEE Trans. Appl. Supercond. 13 (2003) 1221-8.
7. P. Ferracin, et al., “Development of MQXF: the Nb3Sn low-b quadrupole for the Hilumi LHC”, IEEE Trans. Appl. Supercond. 26 (2016) 4000207
8. M. Anerella, et al., “The RHIC magnet system” Nucl. Instrum. Meth. A 499 (2003) 280-315
9. T. Nakamoto, et al., “Model magnet development of D1 beam separation dipoles for the HL-LHC upgrade, IEEE Trans. Appl. Supercond. 25 (2015) 4000505
10. Vector Fields, "OPERA-2d & 3-d User Guide ", Vector Fields Limited, England, 1999

11. ANSYS® www.ansys.com
12. S. Russenschuck, T. Tortschanoff, “Mathematical optimization of superconducting accelerator magnets” IEEE. Trans. Magn. 30 (1994) 3419-3422
13. S. Russenschuck, “Field computation for acceelrator magnets” Wiley-VCH (2010)
14. N. Sammut, L. Bottura and J. Micallef, “Mathematical formulation to predict the harmonics of the superconducting Large Hadron Collider magnets” Phys. Rev. STAB 9 (2006) 012402
15. K. Suzuki, private communication.
16. S. Farinon, et al., "The design of superconducting separation dipoles D2 for the High Luminosity Upgrade of LHC" IEEE Trans. Appl. Supercond. 26 (2016) 4001504
17. RR. Hanft, et al., IEEE Trans. Nucl. Sci. 30, 3381 (1983).
18. H. ten Kate, et al., "Development of an experimental 10 T Nb3Sn dipole magnet for the CERN LHC" IEEE Trans. Magnetics 27 (1991) 1996-1999
19. D. Dell'Orco, et al. "Design of the Nb3Sn dipole D20" IEEE Trans. Appl. Supercond. 3 (1993) 82-86
20. R. Benjegerdes, et al., "Operational characteristics, parameters, and history of a (13 T) Nb3Sn dipole" Particle Accelerator Conference (1999) 3233-3235
21. G. Kirby, et al., "Design and fabrication of a 1 m model of the 70 mm bore twin aperture superconducting quadrupole for the LHC insertions" IEEE Trans. Appl. Supercond. 10 (2000) 61-64
22. Y. Ajima, et al., “The MQXA quadrupoles for the LHC low-beta insertions” Nucl. Instrum. Meths. A 550 (2005) 499-513.

Version 0.2

Ezio.Todesco@cern.ch

Appendix A

A digression on mathematical methods in beam optics, and how chaos entered beam dynamics

Plan of the chapter
This digression is devoted to the discrete formalism used to solve the equations relative to the transverse dynamics in a particle accelerator. We first start (section A.1) with the concept of Hamiltonian equations and with the idea of representing of the motion in the phase space, given by position and momentum. We then discuss (section A.2) how solutions of differential equations can be treated with propagators, and we introduce the concept of mapping, where the trajectory is sampled at discrete time intervals. In section A.3 we recall the use of phase advance and beta functions, already discussed in Chapter 1 and 2, to write a solution of the Hills equations; we also show that in this formalism the differential equation for the beta function of a sequence of magnets is far from being trivial, and explicit solutions cannot be written. Then, in section A.4 the formalism of mappings is applied to derive conditions on the beta function and on the phase advance. The particle motion in dipoles and quadrupoles has analytic solutions that allow computing the transfer maps: these calculations are carried out in section A.5. In section A.6 the values of beta functions, phase advance and quadrupole gradients are given for a FODO cell made of a sequence of focusing and defocusing quadrupoles separated by dipoles. The same approach is used in section A.7 to derive the beta functions in the areas around the experiments. There is an analogy between beam dynamics and optics: in fact, the linear part of the beam dynamics is also called beam optics: in section A.8 we will explore the limits of this analogy by comparing the beam envelope to the optics of light rays in two cases: the FODO cell and the final focus system close to the interaction regions. When imperfections of the magnets are taken into account, nonlinear motion appears: in Section A.9 we give an historical view on how nonlinearities entered classical mechanics, up to the birth of the concept of chaotic motion. Finally, in Section A.10 we show the sensitivity of the particle motion on the betatron frequencies using a simple model (Hénon map) of a FODO cell with a sextupole; this justifies the needed precision in the settings and in the operation of the quadrupole magnets.

A.1 Hamiltonian approach and the concept of phase space

The Hamiltonian and the discrete formalisms are two essential ingredients to solve the motion equations in beam dynamics, and are also used in many other fields of physics. The aim of these first five sections is to compute the equations of motion for a single particle in a lattice of alternating gradient mangets, as first proposed in [1,2,3], using the same mathematical tools developed in these references.

In this first section we will introduce the Hamiltonian formalism and the concept of phase space. According to the Newton approach, the equation of motion of a particle in a force field are given by

$$F = ma\,. \tag{A.1}$$

Considering a one dimensional motion along x, and for the sake of simplicity a force that is a function of the position only, we can write it as a second order differential equation

$$F(x(t)) = m\frac{d^2x(t)}{dt^2} \tag{A.2}$$

that can be solved when coupled to initial conditions, i.e. position and velocity in a given instant t=0

$$\begin{aligned} x(0) &= x_o \\ \frac{dx}{dt}(0) &= v_0. \end{aligned} \tag{A.3}$$

The Hamilton approach replaces the second order differential equation (A.2) with two first order differential equations. One defines the momentum as the mass times the velocity

$$p = m\frac{dx}{dt} \tag{A.4}$$

and therefore the equation of motions are

$$\begin{aligned} \frac{d}{dt}x &= \frac{p}{m} \\ \frac{d}{dt}p &= F(x). \end{aligned} \tag{A.5}$$

These equations can be derived by a function that is called Hamiltonian, that in our simplified case is

$$H(p,x) = \frac{p^2}{2m} + V(x) \tag{A.6}$$

where the potential is defined as the integral of the force

$$V(x) \equiv -\int^{x} F(s)\,ds\,. \tag{A.7}$$

Having the Hamiltonian of the system, the motion equations are given by

$$\begin{aligned} \frac{d}{dt}x &= \frac{\partial}{\partial p}H(p,x) = \frac{p}{m} \\ \frac{d}{dt}p &= -\frac{\partial}{\partial x}H(p,x) = -\frac{d}{dx}V = F. \end{aligned} \tag{A.8}$$

Note that (i) under some conditions that we do not clarify here, the Hamiltonian is the energy of the system and (ii) the Hamiltonian is usually derived via the Lagrangian function, but this rigorous derivation is skipped here; see a general course of mechanics (for instance, the Berkeley Physics Course [4]) if you are interested in more details.

To make the approach more clear, let us consider the simple case of the harmonic oscillator, i.e. a particle of mass m with a force proportional to the displacement x:

$$F(x) = -kx \qquad\qquad V(x) = \frac{k}{2}x^2\,. \tag{A.9}$$

The equations in the Hamiltonian formalism are

$$\frac{d}{dt}x = \frac{p}{m}$$
$$\frac{d}{dt}p = -\frac{d}{dx}V = -kx \tag{A.10}$$

and the solution as a function of the initial conditions x_0, p_0 is

$$x(t) = x_0 \cos\sqrt{k}t + \frac{p_0}{\sqrt{k}}\sin\sqrt{k}t$$
$$p(t) = -\sqrt{k}x_0 \sin\sqrt{k}t + p_0 \cos\sqrt{k}t. \tag{A.11}$$

The phase space is defined as the two-dimensional space (x,p). To each point of the phase space one can associate one and only one trajectory, since it corresponds to one set of initial conditions. The trajectories in phase space are an asynchronous view of the motion. For a pendulum, the trajectories are ellispses whose ratio between x and p is the frequency $\sqrt{k}$. (see Fig. A.1).

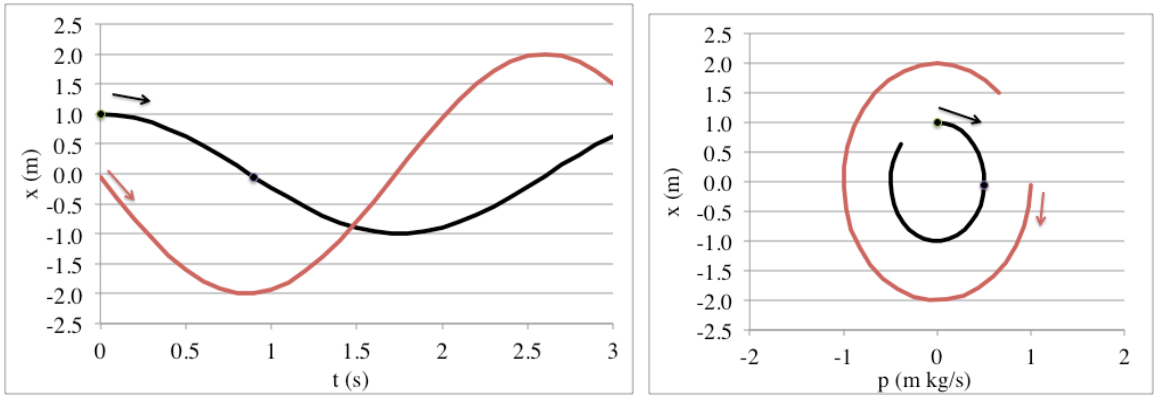


Fig. A.1: Two solutions of the harmonic oscillator, for two different initial conditions (left) and the same solutions as trajectories in the phase space (right).

A.2 Continuous and discrete formalism: propagators and transfer matrices

The solution of the equation of motions as a function of the initial conditions can be written as

$$x(t) = G_x(x_0, p_0; t)$$
$$p(t) = G_p(x_0, p_0; t) \tag{A.12}$$

where the initial conditions x_0, p_0 are propagated to position and momentum in any time; the functions G_x, G_p are called propagatory of the solution of the system (A.5) of differential equations. In the linear case (i.e. if the force is proportional to x, as in a spring or in an harmonic oscillator), the propagator is a function linear in the initial conditions x_0, p_0 , and therefore can be expressed as a transfer matrix, rewriting (A. 11) as

$$\begin{pmatrix} x(t) \\ p(t) \end{pmatrix} = \begin{pmatrix} \cos\sqrt{k}t & \frac{1}{\sqrt{k}}\sin\sqrt{k}t \\ -\sqrt{k}\sin\sqrt{k}t & \cos\sqrt{k}t \end{pmatrix} \begin{pmatrix} x_0 \\ p_0 \end{pmatrix}. \tag{A.13}$$

A.3 Expressing Hills equations with phase advance and beta function

The motion of a particle in a sequence of magnets can be described by the Hills equation

$$\frac{d^2x}{ds^2} + K_1(s)x = 0 \tag{A.14}$$

where s denotes the coordinate along the accelerator, of total length L and therefore with the periodic condition $s+L=s$, and x denotes either the vertical or the horizontal coordinate in the trasnverse plane. Moreover the function K_1 is given by

$$\begin{aligned} K_1(s) &= \frac{G}{B\rho} \qquad \text{when } s \text{ in quadrupoles} \\ K_1(s) &= 0 \qquad \text{when } s \text{ in dipoles.} \end{aligned} \tag{A.15}$$

In this section we show that using the continuous formalism based on differential equations, even in simple cases one does not manage to solve analytically the equations of motion.

Let us start by writing the solution as an oscillator where both frequency and amplitude vary with s

$$x(s) = \sqrt{\beta(s)}\sin(\psi(s)). \tag{A.16}$$

This awkward formulation (with the square root of beta instead of a simple amplitude coefficient) has the advantage that the phase advance can be expressed in terms of the integral of the inverse of the beta function. In order to prove this, we compute the first and the second derivative

$$x'(s) = \frac{1}{2}\frac{\beta'(s)}{\sqrt{\beta(s)}}\sin(\psi(s)) + \sqrt{\beta(s)}\psi'(s)\cos(\psi(s)). \tag{A.17}$$

$$\begin{aligned} x''(s) = &\left(\frac{1}{2}\frac{\beta''(s)}{\sqrt{\beta(s)}} - \frac{1}{4}\frac{\beta'^2(s)}{\sqrt{\beta^3(s)}} - \sqrt{\beta(s)}\psi'^2(s)\right)\sin(\psi(s)) + \\ &\left(\frac{\beta'(s)\psi'(s)}{\sqrt{\beta(s)}} + \sqrt{\beta(s)}\psi'^2(s)\right)\cos(\psi(s)). \end{aligned} \tag{A.18}$$

Replacing in (A.14) and setting to zero the coefficient of the cosine term, one finds that

$$0 = \frac{\beta'(s)\psi'(s)}{\sqrt{\beta(s)}} + \sqrt{\beta(s)}\psi'^2(s) = \frac{\beta'(s)\psi'(s) + \beta(s)\psi'^2(s)}{\sqrt{\beta(s)}} = \frac{\left(\beta(s)\psi'(s)\right)'}{\sqrt{\beta(s)}} \tag{A.19}$$

and therefore

$$\beta(s)\psi'(s) = \text{const}. \tag{A.20}$$

Setting the constant to 1, one obtains

$$\psi'(s) = \frac{1}{\beta(s)} \qquad \psi(s) = \int_{s_0}^{s} \frac{dt}{\beta(t)} \tag{A.21}$$

and therefore the second derivative can be rewritten as

$$x''(s) = \left(\frac{1}{2} \frac{\beta''(s)}{\sqrt{\beta(s)}} - \frac{1}{4} \frac{\beta'^2(s)}{\sqrt{\beta^3(s)}} - \frac{1}{\sqrt{\beta^3(s)}} \right) \sin(\psi(s)) \tag{A.22}$$

and sobstituing in the Hills equation one finds the differential equation for beta as a function of the gradient *K*:

$$\frac{\beta''(s)\beta(s)}{2} - \frac{\beta'^2(s)}{4} - 1 + \sqrt{\beta^3(s)} K_1(s) = 0. \tag{A.23}$$

Here we are facing a pretty complex second order nonlinear differential equation, that cannot be solved analytically. However, we will show that using the Hamiltonan approach and the transfer maps these equations can be solved via pencil and paper.

A.4 Going to the Hamiltonian formalism and transfer maps

Let us take an approach based on Hamiltonian formalism and on propagators; using the result (A.21) of the prevoius section, the first derivative can be written using only the beta functions

$$x'(s) = \frac{\beta'(s)}{2\sqrt{\beta(s)}} \sin(\psi(s)) + \frac{1}{\sqrt{\beta(s)}} \cos(\psi(s)) \tag{A.24}$$

and one can write the solution as a linear combination of sin and cos through the matrix *A(s)*

$$\begin{pmatrix} x(s) \\ x'(s) \end{pmatrix} = \begin{pmatrix} \sqrt{\beta(s)} & 0 \\ \dfrac{\beta'(s)}{2\sqrt{\beta(s)}} & \dfrac{1}{\sqrt{\beta(s)}} \end{pmatrix} \begin{pmatrix} \sin(\psi(s)) \\ \cos(\psi(s)) \end{pmatrix} \equiv A(s) \begin{pmatrix} \sin(\psi(s)) \\ \cos(\psi(s)) \end{pmatrix}. \tag{A.25}$$

The inverse of the matrix is

$$A^{-1}(s) = \begin{pmatrix} \dfrac{1}{\sqrt{\beta(s)}} & 0 \\ -\dfrac{\beta'(s)}{2\sqrt{\beta(s)}} & \sqrt{\beta(s)} \end{pmatrix} \tag{A.26}$$

and we define the normalized coordinates *X* and *X'* as the coordinates where the motion is a simple rotation, without the amplitude modulation:

$$\begin{pmatrix} X(s) \\ X'(s) \end{pmatrix} = \begin{pmatrix} \sin(\psi(s)) \\ \cos(\psi(s)) \end{pmatrix} = A^{-1}(s) \begin{pmatrix} x(s) \\ x'(s) \end{pmatrix}. \tag{A.27}$$

Now we can easily construct the solution of the motion from the coordinate s_0 to s_1 using the transformation to the normalized space where the motion is a simple rotation. We can first go in the normalized space, then propagate the solution, and then go back to the physical space

$$\begin{pmatrix} x(s_1) \\ x'(s_1) \end{pmatrix} = A(s_1) \begin{pmatrix} \cos(\psi(s_1 - s_0)) & \sin(\psi(s_1 - s_0)) \\ -\sin(\psi(s_1 - s_0)) & \cos(\psi(s_1 - s_0)) \end{pmatrix} A^{-1}(s_0) \begin{pmatrix} x(s_0) \\ x'(s_0) \end{pmatrix}. \tag{A.28}$$

For a periodic lattice of length $2L$, we now compute the solution over a period, i.e. having s_0=0 and s_1=$2L$. The periodicity implies $A(s_0)$=$A(s_1)$. Since s_0 and s_1 are the same position. To simplify the notation, we rename $\psi \equiv \psi(2L)$ the phase advance over the cell.

$$\begin{pmatrix} x(2L) \\ x'(2L) \end{pmatrix} = \begin{pmatrix} \sqrt{\beta} & 0 \\ \dfrac{\beta'}{2\sqrt{\beta}} & \dfrac{1}{\sqrt{\beta}} \end{pmatrix} \begin{pmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{pmatrix} \begin{pmatrix} \dfrac{1}{\sqrt{\beta}} & 0 \\ -\dfrac{\beta'}{2\sqrt{\beta}} & \sqrt{\beta} \end{pmatrix} \begin{pmatrix} x(0) \\ x'(0) \end{pmatrix} \equiv M \begin{pmatrix} x(0) \\ x'(0) \end{pmatrix}. \tag{A.29}$$

Let us compute the transfer matrix

$$M \equiv \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} = \begin{pmatrix} \sqrt{\beta} & 0 \\ \dfrac{\beta'}{2\sqrt{\beta}} & \dfrac{1}{\sqrt{\beta}} \end{pmatrix} \begin{pmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{pmatrix} \begin{pmatrix} \dfrac{1}{\sqrt{\beta}} & 0 \\ -\dfrac{\beta'}{2\sqrt{\beta}} & \sqrt{\beta} \end{pmatrix}$$

$$= \begin{pmatrix} \sqrt{\beta} & 0 \\ \dfrac{\beta'}{2\sqrt{\beta}} & \dfrac{1}{\sqrt{\beta}} \end{pmatrix} \begin{pmatrix} \dfrac{1}{\sqrt{\beta}}\cos\psi - \dfrac{\beta'}{2\sqrt{\beta}}\sin\psi & \sqrt{\beta}\sin\psi \\ -\dfrac{1}{\sqrt{\beta}}\sin\psi - \dfrac{\beta'}{2\sqrt{\beta}}\cos\psi & \sqrt{\beta}\cos\psi \end{pmatrix} = \begin{pmatrix} \cos\psi - \dfrac{\beta'}{2}\sin\psi & \beta\sin\psi \\ -\dfrac{4+\beta'^2}{4\beta}\sin\psi & \dfrac{\beta'}{2}\sin\psi + \cos\psi \end{pmatrix}. \tag{A.30}$$

We see that we can now compute the phase advance and the beta function from the coefficients of the transfer matrix

$$\cos\psi = \frac{M_{11} + M_{22}}{2} \tag{A.31}$$

$$\beta = \frac{M_{12}}{\sin\psi}. \tag{A.32}$$

Note that β is the beta function in the location s_0, or s_0+$2L$, and ψ is the phase advance over the cell. This proves that if we know the transfer matrix M, we can compute the solution of the equations of motion in the location s_0. In the next section we show how to compute M by composing the matrices of each magnet.

A.5 Transfer maps for drifts and quadrupoles, and thin lens approximation

For a single element where the function K_1 is constant, the Hill equation reduces to an harmonic oscillator and can be solved analytically. In a drift space, free of magnetic field, the solution is

$$\begin{aligned} x(s) &= x(0) + x'(0)s \\ x'(s) &= x'(0) \end{aligned} \tag{A.33}$$

and in the matrix formalism

$$\begin{pmatrix} x(s) \\ x'(s) \end{pmatrix} = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x(0) \\ x'(0) \end{pmatrix} \tag{A.34}$$

and therefore the transfer matrix of a drift of length L is

$$O = \begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix}. \tag{A.35}$$

For a quadrupole one has

$$\begin{aligned} x(s) &= x(0)\cos\sqrt{K_1}s + \frac{x'(0)}{\sqrt{K_1}}\sin\sqrt{K_1}s \\ x'(s) &= -x(0)\sqrt{K_1}\sin\sqrt{K_1}s + x'(0)\cos\sqrt{K_1}s \end{aligned} \tag{A.36}$$

and in the matrix formalism

$$\begin{pmatrix} x(s) \\ x'(s) \end{pmatrix} = \begin{pmatrix} \cos\sqrt{K_1}s & \frac{1}{\sqrt{K_1}}\sin\sqrt{K_1}s \\ -\sqrt{K_1}\sin\sqrt{K_1}s & \cos\sqrt{K_1}s \end{pmatrix} \begin{pmatrix} x(0) \\ x'(0) \end{pmatrix} \tag{A.37}$$

and for an element of length l_q the transfer matrix reads

$$F = \begin{pmatrix} \cos\sqrt{K_1}l_q & \frac{1}{\sqrt{K_1}}\sin\sqrt{K_1}l_q \\ -\sqrt{K_1}\sin\sqrt{K_1}l_q & \cos\sqrt{K_1}l_q \end{pmatrix}. \tag{A.38}$$

We know consider a thin quadrupole, in the approximation of negligible length and constant integrated gradient (see also Fig. A.2)

$$\begin{aligned} l_q &\to 0 \\ l_q K_1 &\to \frac{1}{f}; \end{aligned} \tag{A.39}$$

the transfer matrix will converge to a simple form

$$F = \begin{pmatrix} 1 & 0 \\ -f^{-1} & 1 \end{pmatrix}. \tag{A.40}$$

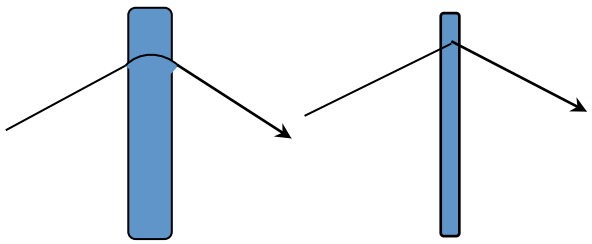


Fig. A.2: Thin lens approximation: the length of the magnet along the beam path goes to zero, preserving its integrated strength; the trajectory keeps its continuity but its derivative has a discontinuity (kick).

A.6 Putting all together: stability and beta functions of a FODO cell

Knowing the transfer matrix of simple elements, as a drift and a quadrupole in the thin lens approximation, we can now construct the transfer matrix of the cell, as a sequence of drifts of length L and of focusing/defocusing quadrupoles (see Fig. A.3).

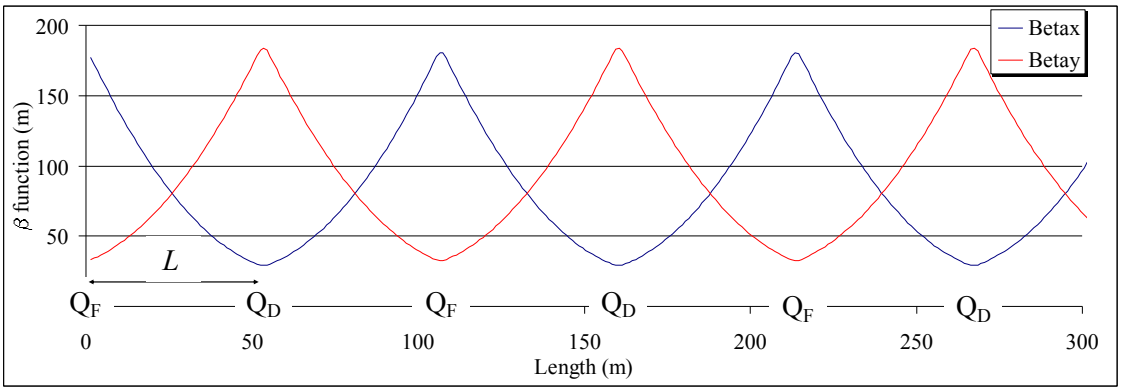


Fig. A.3: The sequence of FODO cell in the LHC lattice, characterized by a quadrupole spacing *L* and by alternate quadrupole focusing/defocusing of focal length *f*; the coordinate along the accelerator is shown in *x*, and beta functions in *y*.

The matrix M in the defocusing quadrupole will read

$$M_D = DOFO = \begin{pmatrix} 1 & 0 \\ f^{-1} & 1 \end{pmatrix}\begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -f^{-1} & 1 \end{pmatrix}\begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix}. \tag{A.41}$$

and with some algebra

$$M = \begin{pmatrix} 1 & L \\ f^{-1} & 1+Lf^{-1} \end{pmatrix}\begin{pmatrix} 1 & L \\ -f^{-1} & 1-Lf^{-1} \end{pmatrix} = \begin{pmatrix} 1-Lf^{-1} & 2L-L^2f^{-1} \\ -Lf^{-2} & 1-L^2f^{-2}+Lf^{-1} \end{pmatrix}. \tag{A.42}$$

we can finally compute using (A.31) the phase advance dependence on *L* and *f*

$$\cos\psi = \frac{2-L^2f^{-2}}{2} \qquad \rightarrow \qquad f = \frac{L}{\sqrt{2-2\cos\psi}} \tag{A.43}$$

and using (A.32) the beta function

$$\beta = \frac{2L-L^2f^{-1}}{\sin\psi}. \tag{A.44}$$

For a 90° phase advance cell, $\psi = \pi/2$ and

$$\cos\psi = 0 = \frac{2-L^2f^{-2}}{2} \Rightarrow \frac{1}{f} = \frac{\sqrt{2}}{L} \approx \frac{1.41}{L}. \qquad \beta_D = \frac{2L-L^2f^{-1}}{\sin\psi} = L(2-\sqrt{2}) \approx 0.6L$$

$$Gl_q = \frac{B\rho}{f} = \sqrt{2}\,\frac{B\rho}{L}. \tag{A.45}$$

Estimating the beta function in the focusing quadrupole, one has to compute the matrix propagating the solution from a focusing quadrupole to the next one

$$M_F = FODO = \begin{pmatrix} 1 & 0 \\ -f^{-1} & 1 \end{pmatrix}\begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ f^{-1} & 1 \end{pmatrix}\begin{pmatrix} 1 & L \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1+Lf^{-1} & 2L+L^2f^{-1} \\ -Lf^{-2} & 1-L^2f^{-2}-Lf^{-1} \end{pmatrix} \tag{A.46}$$

and the phase advance is obviously the same, whereas the beta function reads

$$\beta_F = \frac{2L+L^2f^{-1}}{\sin\psi} = L(2+\sqrt{2}) \approx 3.4L. \tag{A.47}$$

Example A.1: Compute the relations between beta functions in the quadrupoles, the quadrupole strength, and the cell length for a 60° and for a 45° phase advance.

For a 60° phase advance, since $\psi=\pi/3$, one has

$$\sin\psi = \frac{\sqrt{3}}{2} \qquad \cos\psi = \frac{1}{2}. \tag{A.48}$$

The focusing strength and the integrated gradient are (see A.43 and A.45)

$$f = L \qquad Gl_q = \frac{B\rho}{L}, \tag{A.49}$$

so they are 30% less than a 90°cell with the same quadrupole spacing. The beta functions are

$$\beta_D = \frac{2L+L^2f}{\sin\psi} = 2\sqrt{3}L \approx 3.5L \qquad \beta_F = \frac{2L-L^2f}{\sin\psi} = \frac{2}{3}\sqrt{3}L \tag{A.50}$$

are therefore the maximum beta is very similar to the 90° phase advance with the same quadrupole spacing L.

For a 45° phase advance, $\psi=\pi/4$ and one has

$$\sin\psi = \cos\psi = \frac{\sqrt{2}}{2}. \tag{A.51}$$

The focusing strength and the integrated gradient are

$$f = \frac{L}{\sqrt{2-\sqrt{2}}} \approx 1.3L \tag{A.52}$$

so it is half of a 90°cell with the same quadrupole spacing. The beta functions are

$$\beta_D = \frac{2L+L^2f^{-1}}{\sin\psi} = \sqrt{2}(2+\sqrt{2-\sqrt{2}})L \approx 3.9L \qquad \beta_F = \sqrt{2}(2-\sqrt{2-\sqrt{2}})L \tag{A.53}$$

i.e., they are 15% larger than the 90° cell with the same quadrupole spacing L.

Among the two options, the 60° phase advance is more efficient since it requires a similar magnet aperture as the 90° (similar beta function in the focusing quadrupoles) but requires 30% less integrated gradient in the quadrupoles.

A.7 Phase advances and beta functions in the interaction regions

In this section we will compute the beta functions in the experimental area, that is a region free of magnets, characterizied by a symmetry condition in the centre. The approximate layout, in the case of the LHC high luminosity insertion regions is shown in Fig. A.4, and is discussed in Chapter 2.

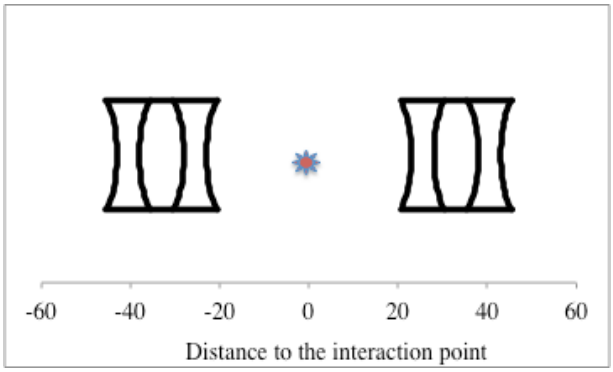


Fig. A.4: Schematic the triplet layout arund the interactions point in the LHC.

We will here prove Eq. (2.15), i.e. that beta functions have a parabolic dependence on the distance to the centre of the experiment, plus the offset given by the beta function in the interaction point β^*:

$$\beta(s) = \beta^* + \frac{s^2}{\beta^*}. \tag{A.54}$$

Firstly, let us estimate the propagation of the optical functions from the center of the experiment (taken as a reference system at *s=0*) to a distance *l* along *s*. The derivative of the beta function in the center will be zero by symmetry conditions. Therefore, using (A.30), the transfer matrix from the center of the experiment to the position *l* will read

$$M \equiv \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} = \begin{pmatrix} \sqrt{\beta(l)} & 0 \\ \dfrac{\beta'(l)}{\sqrt{\beta(l)}} & \dfrac{1}{\sqrt{\beta(l)}} \end{pmatrix} \begin{pmatrix} \cos\psi(l) & \sin\psi(l) \\ -\sin\psi(l) & \cos\psi(l) \end{pmatrix} \begin{pmatrix} \dfrac{1}{\sqrt{\beta^*}} & 0 \\ 0 & \sqrt{\beta^*} \end{pmatrix}. \tag{A.55}$$

The computation gives

$$M = \begin{pmatrix} \sqrt{\dfrac{\beta(l)}{\beta^*}}\cos\psi(l) & \sqrt{\beta(l)\beta^*}\sin\psi(l) \\ \cdots & \cdots \end{pmatrix} \tag{A.56}$$

where we only computed the first two terms of the matrix, since the third and the fourth are not necessary for our estimates. Since transfer matrix in the experiments is a drift

$$\begin{aligned} x(l) &= x(0) + x'(0)l \\ x'(l) &= x'(0) \end{aligned} \qquad M = \begin{pmatrix} 1 & l \\ 0 & 1 \end{pmatrix}, \tag{A.57}$$

comparing the two expressions one obtains

$$\begin{aligned} \sqrt{\frac{\beta(l)}{\beta^*}}\cos\psi(l) &= 1 \\ \sqrt{\beta^*\beta(l)}\sin\psi(l) &= l \end{aligned} \quad \text{that gives} \quad \begin{aligned} \cos\psi(l) &= \sqrt{\frac{\beta^*}{\beta(l)}} \\ \sin\psi(l) &= \frac{l}{\sqrt{\beta^*\beta(l)}} \end{aligned} \tag{A.58}$$

and therefore we recover the expression (A.54)

$$1 = \cos^2\psi + \sin^2\psi = \frac{\beta^*}{\beta(l)} + \frac{l^2}{\beta^*\beta(l)} = \frac{1}{\beta(l)}\left(\beta^* + \frac{l^2}{\beta^*}\right)$$

$$\beta(l) = \beta^* + \frac{l^2}{\beta^*} \ . \tag{A.59}$$

Note that with the usual methods of differential equations finding the above solution is not trivial. Even though the equation to be solved is as simple as

$$\frac{d^2x}{ds^2} = 0, \tag{A.60}$$

using the Ansatz

$$x(s) = \sqrt{\beta(s)}\cos(\psi(s)) \tag{A.61}$$

and estimating the first and second derivatives

$$\frac{dx}{ds} = \frac{\beta'(s)}{2\sqrt{\beta(s)}}\cos(\psi(s)) - \frac{1}{\sqrt{\beta(s)}}\sin(\psi(s)) \tag{A.62}$$

$$\frac{d^2x}{ds^2} = \left(\frac{\beta''(s)}{2\sqrt{\beta(s)}} - \frac{1+\left[\beta'(s)\right]^2/4}{\left[\beta(s)\right]^{3/2}}\right)\cos(\psi(s)) \tag{A.63}$$

one ends up with a non trivial differential equation

$$\frac{\beta(s)\beta''(s)}{2} = 1 + \frac{\left[\beta'(s)\right]^2}{4} \ . \tag{A.64}$$

One can check that using a parabolic solution

$$\beta(s) = a + bs^2 \tag{A.65}$$

and replacing in (A.64), one has

$$b(a + bs^2) = 1 + b^2s^2 \qquad ab = 1 \tag{A.66}$$

and therefore

$$\beta(s) = a + \frac{s^2}{a} \tag{A.67}$$

as found in (A.59). However, this computation is based on the hypothesis (A.65) whereas the map method gives directly the solution without the need of any further hypothesis.

A.8 The analogy of beam dynamics with optics: angels and demons

It may seem surprising that a sequence of focusing and defocusing lenses can be focusing. The statement should be better given as "a sequence of focusing and defocusing lenses can give a confined beam" (i.e. whose trajectories do not diverge). As an example, a sketch of the light rays of a focusing/defocusing length with focal distance L=50 m is given in Fig. A.5. The path of optics rays in a sequence of focusing and defocusing spaced by L=50 and focusing length $\pm L\sqrt{2}$ is shown in Fig. A.6, left. This shows that the sequence of opposite focusing lenses with adequate relation between spacing and focusing gives confined trajectories.

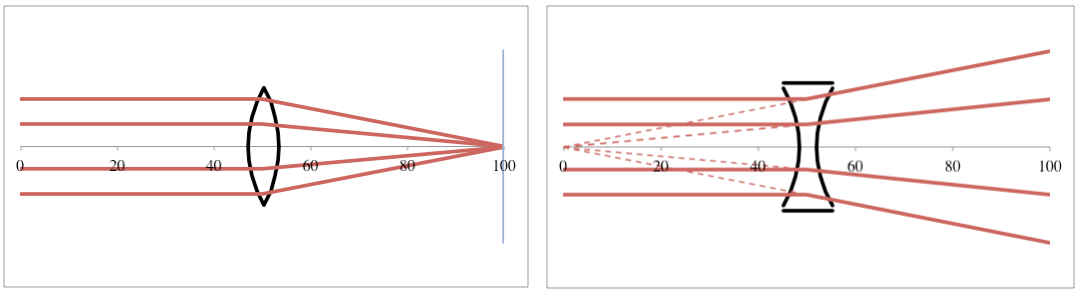


Fig. A.5: Focusing lens with focal distance L=50 m (left) and defocusing lens with focal length -50 m (right).

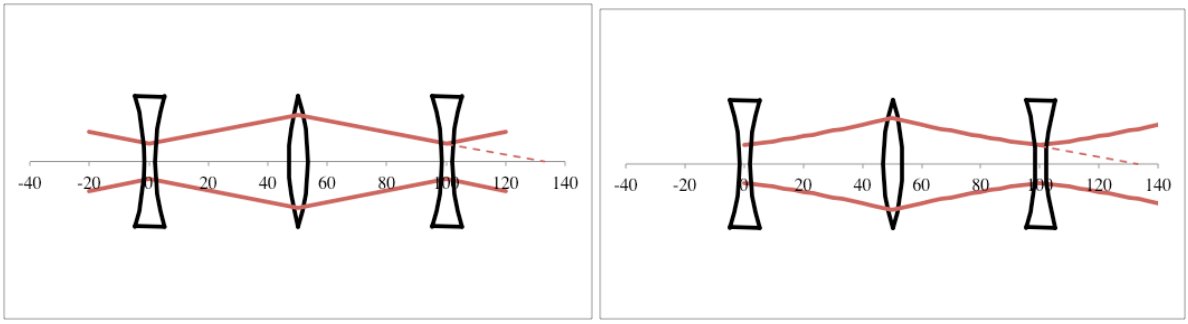


Fig. A.6: Path of rays in an optical system composed of focusing and defocusing thin lens, with focal length larger than distance between the lenses (left), and beta functions in a FODO cell (right).

The analogy with the beam envelope in a sequence of focusing and defocusing quadrupoles is rather striking and is shown in Fig. A.6, right. Here we plot the square root of the beta function (i.e. the size of the beam, see Eq. A.16) for the same case of quadrupole spacing of 50 m and focusing/defocusing strength of $\pm L\sqrt{2}$. Note that the beta function in a drift has a quadratic dependence, and therefore in our case

$$\beta(s) = k_0 + k_1 s + k_2 s^2 \tag{A.68}$$

with

$$\beta_D = \beta(0) = k_0 \tag{A.69}$$

and more complicated expressions for k_1 and k_2, that are not relevant for our discussion. The beam size is proportional to the square root of the beta function

$$x(s) \propto \sqrt{\beta_D + k_1 s + k_2 s^2} \tag{A.70}$$

and this is what is plot in Fig. A.6, right. Now, everybody knows that the square root of a parabola is not a straight line, but in our example it is quite close to the optical path of the rays shown on the left side. Therefore the optical analogy between the light rays and the beam size is quite accurate, but not exact. Strictly speaking, it is false, but it is very close to reality.

Every analogy has its angels and demons, and it should be used with extreme care. Analogies provide a fast and intuitive way of "understanding" things but, as shown for the case of the relativistic mass in Chapter 1, they can sometimes lead to wrong conclusions. It is very interesting to apply the optics analogy to the case of the beta functions in the interaction region. One can consider the triplet before the experiments as a focusing system, with a focusing distance equal to the distance of the triplet baricentre to the centre of the experiment (see Fig. A.7). Also in this case the beta function in the experiment, free of accelerator magnets, have a quadratic form. In the previous section we have show that it is given by

$$\beta(s) = \beta(0) + \frac{s^2}{\beta(0)} = \beta^* + \frac{s^2}{\beta^*}. \tag{A.71}$$

Also in this case, the analogy is quite exact, but not completely. In Fig. A.7 and A.8 we plot the square root of the beta function, proportional to the beam size: once more, it appears very similar to the light rays in a focusing lens, i.e., a the beam size reduces linearly with the distance to the centre of the experiment. The magnification shows that the analogy is wrong: the beam size (obviously) is not zero in the centre of the experiment, but has a finite size, proportional to the square root of β(0). Beware of analogies, use them with extreme caution, and whenever you have doubts, use mathematics to guide and support your intution.

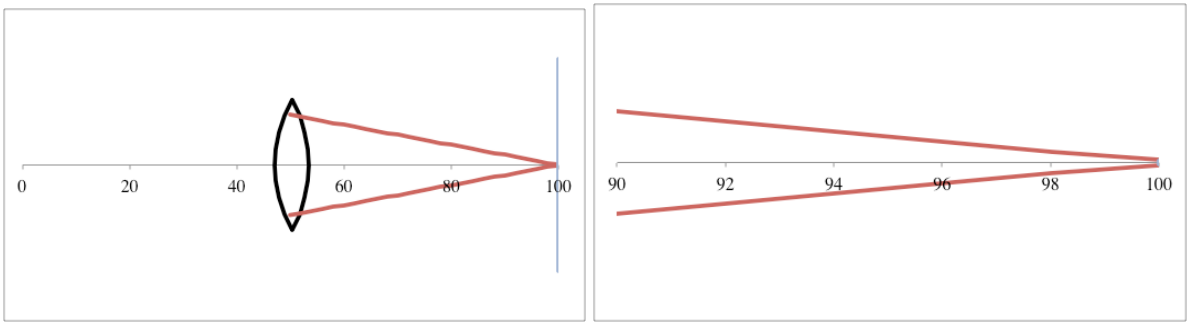


Fig. A.7: Beta functions in the experimental region for the LHC case (left) and magnification of the area around the interaction point (right).

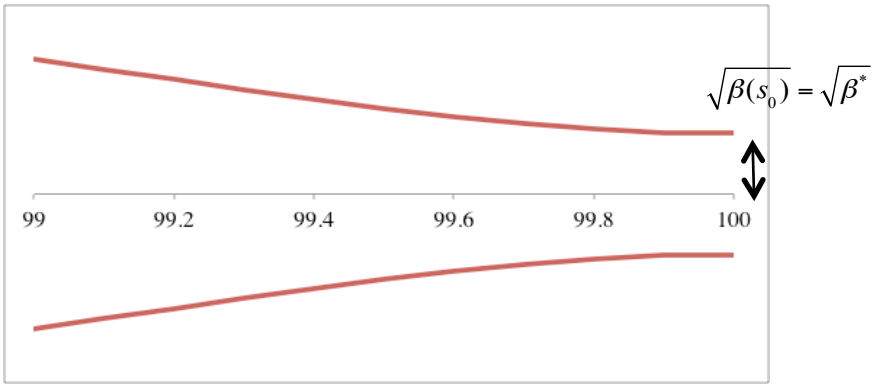


Fig. A.8: Futher magnification of the area around the interaction point (right).

A.9 From a prize for the king of Sweden to the butterfly effect: how chaos entered beam dynamics

In 1885, the well-known Swedish mathematician Gösta Mittag-Leffler issues a special prize for mathematics in the honor of Oscar II, king of Sweden and Norway [5]. Among the four questions given to the competitors, the first one is the problem of the stability of the solar system. It is well-known since Newton and Keplero that in the case of two bodies one has stable elliptic orbits, and unstable hyperbolic orbits. However, when a third planet is added, the problem does not have known analytic solutions; in principle the the perturbations induced by Jupiter on the orbit of the Earth around the Sun could sum up and finally drive the Earth out of the solar system. At that time, most of the astronomers believed that the mass of the other planets was so small with respect to the mass of the Sun that stability was practically guaranteed for infinite times. However, no mathematical proof of this statement was available and this was considered as one of the most relevant questions in the physcis of the 19th century.

The competitors of the contest have three years to give an answer to any one of the four questions. Henri Poincarè, one of the most acknowledged mathematicians of the time, sends a work where he finally gives a proof to the first question in the case of three bodies, with one of negligible mass (not affecting to orbits of the two more massive bodies) and having the motion restricted on a plane: the so-called restricted three-body problem.

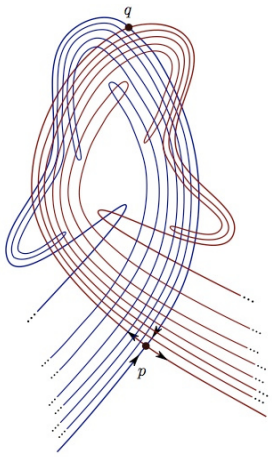


Fig. A.9: Henri Poincarè (French, 1854-1912) and orbits passing through an homoclinic point

Even though the first question is solved in a simplified version, the committee assigns the prize to Poincarè in January 1889; the assigment is however postponed as a less known astronomer claims to have found the same result. When in November 1889 Mittag-Leffler decides to reject the claim of the astronomer, Poincarè asks to put the publication on hold: he has found a mistake. What initially seems to the matematician as an imperfection of his argument, revelas to be a crucial aspect of the geometrical nature of the problem [5,6]. Poincarè discovers the existence of homoclinic points, where the surfaces of the motion invariants, where the trajectories are constrained, gets intersected with extremely complex patterns. If there are enough degrees of freedom, as it happens when a third body is considered, these points can give rise to non-confined orbits: therefore the stability of the solar system, given for granted by the whole community, is false under some initial conditions. Moreover, in homoclinic points trajectories that are very close can exponentially diverge during the evolution in time

$$d(t) \approx A\exp\left(\lambda t\right) \tag{A.72}$$

with $\lambda > 0$. The quantity λ was being introduced by Lyapounov (later known as Lyapounov exponent) in the same period [7]. When Laplace, nearly one hundred years before, was claiming to be able to forecast future and past, given the forces and the initial conditions, he was assuming an infinite

precision in the knowledge of the initial conditions. Since any measurement is affected by an error, in the neighbourhood of homoclinic points one cannot make any forecast after a given time.

It is interesting to see how Mittag-Leffler treats the embarassing case: the prize is confirmed to Poincaré, who is awarded with 2500 crowns. However, the work is retired (a copy survived and is today in the Gottag-Leffler Institute, Uppsala) and Poincarè has to pay for the publications fees, whose amount is order of 7000 crowns.

The result of Poincarè is somehow forgotten, and at the beginning of the 20th century the problem of physics is not any more the stability of the solar system, but the structure of the atoms and its relation with electromagnetism and the theory of quanta. In Russia, where the political establishment does not appreciate the philosophical ambiguities of quantum mechanics, the school of mathematical physicists continue to study the problem, with major advancements. In the 60's, Kolmogorov gives a first proof [8] of what will become the KAM theorem (named after Kolmogorov, Arnol'd and Moser). This theorem states that there are unstable orbits, but their abundance is extremely small when the perturbation is small; this result reconciles the intuition of the astronomers of 19th century with the nichilist result of Poincarè. The mechanism that drives the extremely small fraction of initial conditions from regions where the perturbations are very small towards unstable orbits is called Arlon'd diffusion. See Ref. [9] for a divulgative book on these beautiful mathematical results, full of consequences for the world of physics.

Fig. A.10: Andrey Kolmogorov (Russian, 1903,1987), Vladimir O. Arnold (Russian, 1937-2010) and Jürgen Moser (German, 1928-1999).

In the same years, computers start to be used to simulate many different phenomena. Among them, simplified versions of the equations ruling meteorology are studied in the US by Edward Lorenz [10]. When repeating the same computation with 3 digits rather than with 6 digits, Lorenz finds that the final result of the simulation is totally different, rather than changing on the third digit: he has re-discovered the sensitivity to initial conditions in a system (later called Lorenz attractor) where the error associated to the knowledge of the initial conditions propagates exponentially, as in the Poincarè homoclinic points. It is clear that for these systems one has to rethink about many paradigms of physics, and in particular about the possibility of carrying simulations to make weather forecast; he presents this result in a conference, using the seagull metaphor [11]: "… the movement of the wings of a seagull could affect the weather forecast …". However, the result is mostly ignored in the scientific community.

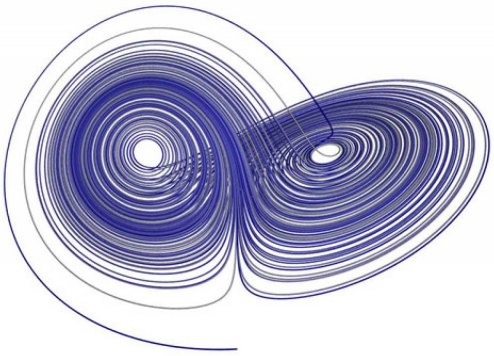


Fig. A.11: Edward Lorenz (1854-1912) and the Lorenz attractor: small variations in the initial conditions can bring the trajectory on one side or on the other one.

Ten years later, in the seventies, the chairman of a conference session suggests him to use a different metaphor as a title of his talk: "Does the flap of a buttefly's wings in Brazil set off a tornado in Texas?" This time the communication makes a strike, and in a few years the oxymor of "deterministic chaos" is applied to many fields of physics, from biology to economics, becoming a viral topic in our society [12].

At the end of the eighties, deterministic chaos also is a major concern for the accelerator physics teams that are working on the design of the Large Hadron Collider in Geneva and of the Superconducting Super collider in Texas: here, the concern are the non-linear components in the magnetic field (multipoles) that, even though very small (10^{-4} times the main field) could act as Jupiter perturbations on the Earth, and progressively drive the particles out of the beam - let us not forget that the ratio between Jupiter and Sun mass is 10^{-3}. Also in this case, the KAM theorem guarantees that for very small perturbations (i.e. particle in the beam core) the fraction of initial conditions able to escape from the stability region is extremely (exponentially) small.

I remember a long discussion in a workshop in Tuscany about the possible impact of Arnol'd diffusion on the beam losses of the LHC in the mid of the 90's, with a majority of colleagues convinced of its disruptive effect on the accelerator performances. The discussion was interrupted by the reassuring remarks of a very elegant elder man, who started saying with a slight Russian accent "Well, I do not agree … I think that Arnol'd diffusion is …", generating a unanimous hilarity. Not understanding what was going on, I asked my neighbourhood who was the colleague, and I was told "Arnol'd !".

Even though Arnol'd diffusion is not an issue for beam dynamics, nonlinearities (i.e. sextupole, octupole, and higher orders) play a relevant role in setting the dimension of phase space around the reference orbit where the overwhelming quantity of orbits are stable (so-called dynamic aperture). In particular, this size has a large sensitivity on the fractional part of frequency of the orbits. This is setting very stringent requirements on the quadrupole gradients, as it will be shown in the last section of this digression.

A.10 The conservative Hénon map and sensitivity to quadrupole gradients

Having discussed how nonlinearities entered the world of classial mechanics along a 100-year long timeline, what happens when we we add a nonlinear term to the solution of the linear equations of motion as described in sections 1 to 8 ? Let us consider an accelerator made of a one FODO cell, of total length $2L$, with the focusing quadrupole at s=0, the defocusing quadrupole at s=L, and with a sextupole term placed in the focusing quadrupole. Forgetting about the sextupole, the solution is as writen in (A.29)

$$\begin{pmatrix} x(2L) \\ x'(2L) \end{pmatrix} = \begin{pmatrix} \sqrt{\beta} & 0 \\ \dfrac{\beta'}{2\sqrt{\beta}} & \dfrac{1}{\sqrt{\beta}} \end{pmatrix} \begin{pmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{pmatrix} \begin{pmatrix} \dfrac{1}{\sqrt{\beta}} & 0 \\ -\dfrac{\beta'}{2\sqrt{\beta}} & \sqrt{\beta} \end{pmatrix} \begin{pmatrix} x(0) \\ x'(0) \end{pmatrix}. \tag{A.73}$$

In s=0 the sextupole acts as a kick

$$\frac{d^2x}{ds^2} + K_2(s)x^2 = 0 \tag{A.74}$$

and defining as 0+ as the position immediately after the setupole, and 0+ the position immediately before, one has

$$\begin{aligned} x(0_+) &= x(0_-) \\ x'(0_+) &= x'(0_-) + K_2\left[x(0_-)\right]^2 \end{aligned} \tag{A.75}$$

and therefore the full one turm map reads

$$\begin{pmatrix} x(2L) \\ x'(2L) \end{pmatrix} = \begin{pmatrix} \sqrt{\beta} & 0 \\ \dfrac{\beta'}{2\sqrt{\beta}} & \dfrac{1}{\sqrt{\beta}} \end{pmatrix} \begin{pmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{pmatrix} \begin{pmatrix} \dfrac{1}{\sqrt{\beta}} & 0 \\ -\dfrac{\beta'}{2\sqrt{\beta}} & \sqrt{\beta} \end{pmatrix} \begin{pmatrix} x(0) \\ x'(0) + K_2x^2(0) \end{pmatrix}. \tag{A.76}$$

We now move in the reference system (X,X') where the linear motion is a simple rotation:

$$\begin{pmatrix} X(0) \\ X'(0) \end{pmatrix} = \begin{pmatrix} \dfrac{1}{\sqrt{\beta}} & 0 \\ -\dfrac{\beta'}{2\sqrt{\beta}} & \sqrt{\beta} \end{pmatrix} \begin{pmatrix} x(0) \\ x'(0) \end{pmatrix} \tag{A.77}$$

finding

$$\begin{pmatrix} X(2L) \\ X'(2L) \end{pmatrix} = \begin{pmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{pmatrix} \begin{pmatrix} \dfrac{x(0)}{\sqrt{\beta}} \\ -\dfrac{\beta' x(0)}{2\sqrt{\beta}} + \sqrt{\beta}x'(0) + \sqrt{\beta}K_2x^2(0) \end{pmatrix} \tag{A.78}$$

that can be rewritten as

$$\begin{pmatrix} X(2L) \\ X'(2L) \end{pmatrix} = \begin{pmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{pmatrix} \begin{pmatrix} X(0) \\ X'(0) + \dfrac{K_2X^2(0)}{\sqrt{\beta}} \end{pmatrix}. \tag{A.79}$$

At the end of the 80's, the astronomer M. Hénon defines the map [13]

$$\begin{pmatrix} x \\ x' \end{pmatrix} = \begin{pmatrix} \cos\nu & \sin\nu \\ -\sin\nu & \cos\nu \end{pmatrix} \begin{pmatrix} x \\ x'+x^2 \end{pmatrix} \tag{A.80}$$

as one of the simplest conservative dynamical system showing complex dynamics. We define the orbit associated to an initial condition x_0, x'_0 be rewritten as the sequence of points in the phase space

$$\begin{aligned} x_{n+1} &= x_n \cos\nu + (x'_n + x_n^2)\sin\nu \\ x'_{n+1} &= -x_n \sin\nu + (x'_n + x_n^2)\cos\nu. \end{aligned} \tag{A.81}$$

The map is a rotation of an angle ν, plus a nonlinear quadratic term proportional to the x coordinate. This simple dynamical system corresponds to the motion of a particle in a accelerator with all linear elements (dipoles and quadrupoles), plus one thin sextupole. In the above formulation, this is the motion of particles constrained on x (i.e. vertical coordinate y=0 at all times). We will use this simple model to shown how the zone of stable orbits around the origin critically depends on the angle ν.

Five orbits, corresponding to initial conditions with larger and larger amplutides, are shown in Fig. A.12. For small amplitudes (grey orbit), the orbit lays on a circle since the nonlinear perturbation is small. For increasing amplitudes the trajectories become more complex. Since in the example the frequency is close to the resonance ¼=0.25, nonlinearities excite resonances and create trajectories that are broken in four islands (yellow and red orbits). For larger amplitudes, the resonance is bypassed and one goes back to closed orbit (blue orbit). For initial conditions with amplitudes larger than the blue orbit, the nonlinearity becomes dominant and the trajectory diverges. This is what in accelerator physics is usually called dynamic aperture, i.e. the border of the stable motion. The adjective dynamic is used to distinguish from the mechanical aperture, that is the aperture of the magnets. The beam has to stay well within the blue line, otherwise it is lost after some turns of the accelerator.

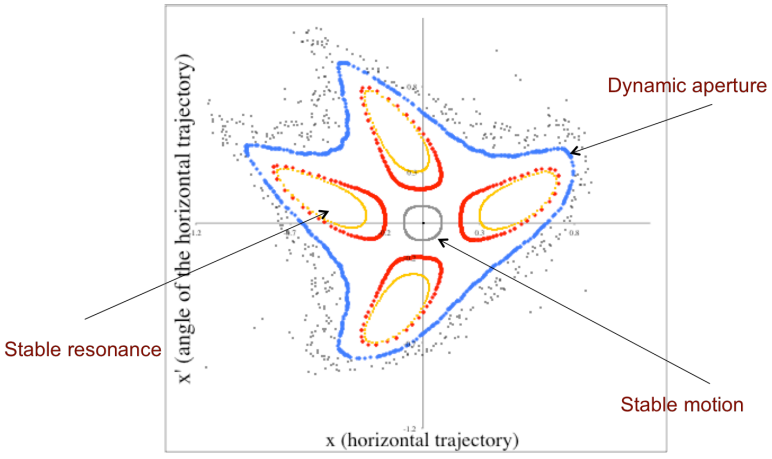


Fig. A.12: The phase space of the one-turn map of a particle accelerator with a linear part and a sextupolar kick (Hénon map) for ν=69.252.

The frequency of the map is related to the gradient of the quadrupoles. If we change the frequency of the map on the third digit we observe a radical change of the stability border, that shrinks by more than a factor two (see Fig. A.13).

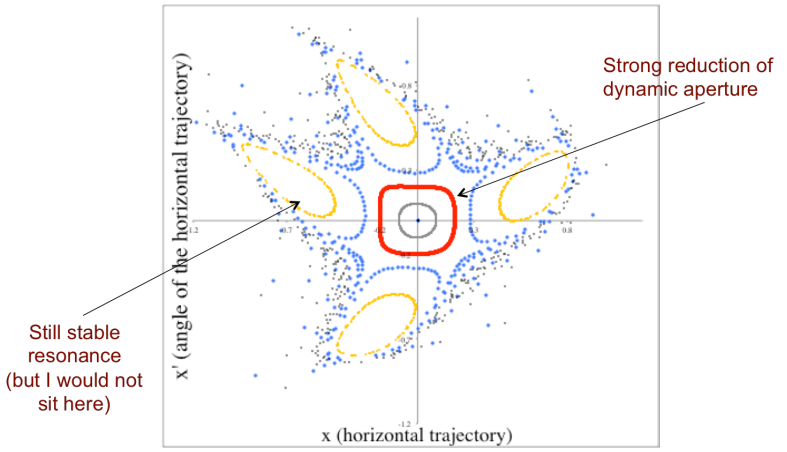


Fig. A.11: The phase space of the one-turn map of a particle accelerator with a linear part and a sextupolar kick (Hénon map) for ν=69.254.

The integer part of the frequency is nothing else than the number of cells times the phase advance: here we have used the numbers relative to the LHC, where one has 23.5 cells in each octant, giving a 90° degrees advance, for a total of 47 integer oscillations; other 20 oscillations are made in the insertion regions. Having to control the tune in 0.001 over 60 means reaching a precision in steering the quadrupole gradient, via the power converter, of 15 ppm.
In the previous example, with a two dimensional phase space, a particle starting within the red orbit of Fig. A.11 has no possibility of crossing this orbit and diverge to large amplitudes. Let us now consider the extension of the Hénon map to the 4 dimensional phase space, i.e. the particle motion in transverse space

$$\begin{aligned} x_{n+1} &= x_n \cos\nu_x + \left(x'_n + x_n^2 - y_n^2\right)\sin\nu_x \\ x'_{n+1} &= -x_n \sin\nu_x + \left(x'_n + x_n^2 - y_n^2\right)\cos\nu_x \\ y_{n+1} &= y_n \cos\nu_y + \left(y'_n - 2x_n y_n\right)\sin\nu_y \\ y_{n+1} &= -y_n \sin\nu_y + \left(y'_n - 2x_n y_n\right)\cos\nu_y \end{aligned} \tag{A.82}$$

In this mapping, as in the three body problem of Poincarè, there are enough degrees of freedom for having homoclinic points bringing you from the very inside of the beam core to outside. In practice, the concept of dynamic aperture strictly speaking is not applicable since there are initial conditions arbitrary close to the reference orbit that can escape via Arnol'd diffusion. As in the solar system, these initial conditions are extremely rare: this is why LHC works notwithstanding Arnol'd diffusion.

Acknowledgements

I wish to thank G. Turchetti for teaching me mappings and perturbative theory in nonlinear systems, and for the works done together on nonlinearities in accelerator physics during the 90's, with colleagues from University of Bologna and from CERN. I would not have written and proposed this Appendix without the enthusiasm of Glyn Kirby, who during the lectures of 2020 suggested to open this digression.

References

1. E. D. Courant, H. S. Snyder, "Theory of alternating gradient synchrotron" *Ann. Phys.* **3** (1958) 1-48.
2. E. Courant, M. S. Livingston, and H. Snyder, "The strong focusing synchrotron – a new high energy accelerator" *Phys. Rev.* **88** (1952) 1190.
3. N. C. Christofilos, unpublished manuscript (1950).
4. C. Kittel, et al., "Mechanics (Berkeley Physics Course, Vol. 1)" MacGraw-Hill (New York, 1973).
5. M. Rägstedt, "From order to chaos: the prize competition in honor of King Oscar II" www.mittag-leffler.se/library/henri-poincare .
6. J. Barrow-Green, "Oscar II's prize competition and the error in Poincaré's memoir on the three body problem" *Arch. Hist. Exact. Sci.* **48** (1994), 107-131.
7. A. Lyapounov, "The general problem of the stability of motion" CRC press (1908, in Russian, English edition in 1992).
8. A. N. Kolmogorov, "On the conservation of conditionally periodically motions under small perturbation of the Hamiltonian" *Dokl. Akad. Nauk.* SSR **98** (1954).
9. H. S. Dumas, "The KAM story " World Scientific (Singapore, 2014).
10. E. Lorenz, "Deterministic non periodic flow" *J. Atmos. Sci.* **20** (1963) 130-141.
11. R. C. Hilborn, "Sea gulls, butterflies and grasshoppers: a brief history of the butterfly effect in the nonlinear dynamics" *Am. J. Phys.* **72** (2004) 425.
12. J. Gleick "Chaos: making an new science" Viking books (1987).
13. M. Hénon, "Numerical study of quadratic area-preserving mappings" *Quart. Appl. Math.* **27** (1969) 291-312.

Version 0.3

Ezio.Todesco@cern.ch

Appendix B

A digression on electromagnetism, its divergences and lengths in atomic physics

Plan of the appendix
In this digression we first recall an argument that shows how the force induced on a charged particle by the magnetic field can be seen an electrostatic field in the reference frame of the particle induced by contraction of lengths given by special relativity. We then discuss in sections B.2 and B.3 the divergences that are intrinsic to the Maxwell equations for infinite charge density and for infinite current density; the first one allows defining the classical electron radius, a scale length that appears in Thomson scattering. In section B.4 we show how starting from the second half of 19th century the scale invariance for space and time is broken by setting an upper limit to speed and a lower limit to angular momentum. The combination of the two constants allows normalizing the strenght of the electric charge, giving rise to the fine structure constant. In Section B.5 we introduce the Bohr radius, showing how it is four order of magnitudes larger the classical electron radius. In Section B.6 we describe how the Compton length is the logarithmic average of the Bohr radius and of the classical electron radius, and how the three lengths are spaced by a factor equal to the fine structure constant. In section B.7 we show how the main equations of quantum electrodynamics can be written in terms of a scale constant (the Compton length) and a coupling constant (the fine structure constant), and therefore the historical names still in use today do not account for the physical meaning of these quantities. Finally, in section B.8 we introduce the intrisic properties of the electron, namely the angular momentum (spin) and the magnetic momentum; we show how the alignment alle the intrinsic momentum of electrons can provide fields of the order of 1 T, as happens in ferromagnetic materials.

B.1 Magnetic field as a side effect of special relativity

Maxwell equations can be written as

$$\begin{aligned} \nabla E &= \frac{\rho}{\varepsilon_0} & \nabla B &= 0 \\ \nabla \times E &= -\frac{\partial B}{\partial t} & \nabla \times B &= \mu_0 j + \mu_0 \varepsilon_0 \frac{\partial E}{\partial t} \end{aligned} \tag{B.1}$$

where we use the international system units; this is not the original form of Maxwell of 1865, but rather the expression reworked by Heaviside in 1884. In the CGS system, the form becomes cleaner, as the vacuum permittivity constants μ_0 and ε_0 disappear and the only constant in the equations is the speed of light

$$\nabla E = 4\pi\rho \qquad \nabla B = 0$$
$$\nabla \times E = -\frac{1}{c}\frac{\partial B}{\partial t} \qquad \nabla \times B = \frac{4\pi}{c} j + \frac{1}{c}\frac{\partial E}{\partial t} \tag{B.2}$$

In this formulation, a symmetry between magnetic field and electric field appears. The symmetry would be perfect if a magnetic charge density would exist, giving rise to a non-zero divergence of the magnetic field (magnetic monopoles) and to current densities of magnetic charges contributing to $\nabla \times E$. Notwithstanding the experimental efforts, no evidence of magnetic monopoles has been found so far [1].

Fig. B.1: James Clerk Maxwell 1831-1879, Scottish (left), and Oliver Heaviside 1850-1925, English (right)

A very interesting interpretation points in a totally different direction from any attempt of symmetrizing Maxwell equations: the magnetic field is the electrostatic field seen from a moving reference system [2]. This gives us also a quite powerful intuition on the enigmatic form of the electromagnetic force produced by the magnetic field on a particle moving with a speed v

$$\vec{F} = e\vec{B} \times \vec{v} \,; \tag{B.3}$$

the force has two awkward features: (i) it is perpendicular to the velocity and to the magnetic field and (ii) it is proportional to the particle speed. In the following text we will show this equivalence in the case of a magnetic field generated by a current parallel to the speed of the moving charge: this "strange" force becomes in the reference system of the particle nothing else than the electrostatic force, proportional to the square of the distance of the charges, and to the product of the charges.

Where does the charge arise, as the current line is electrically neutral? In the reference system of the particle, the relativistic effect of contraction of lengths acts differently on the flow of positive and negative particles constituting the current, and therefore the current line becomes charged (see Fig. B.2). This simple argument also justifies the orientation of the force, perpendicular both to the speed and to the magnetic field.

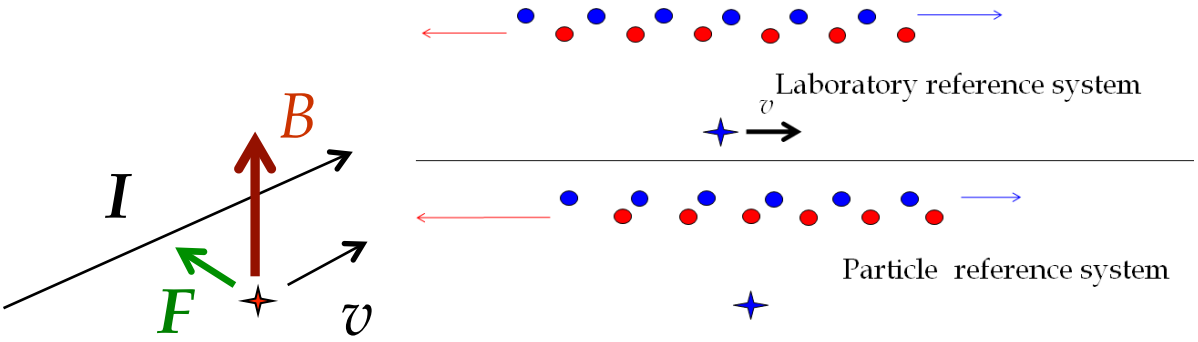


Fig. B.2: Electromagnetic force (left), distribution of the charges in the laboratory reference system (right, upper part) and in the particle system (right, lower part)

We consider a magnetic field given by an infinite straight current line at a distance r from the particle, and parallel to the direction of the particle speed v (see Fig. B.2). The magnetic field induced by the the current line is

$$B = \frac{\mu_0 I}{2\pi r} \tag{B.4}$$

and therefore the force in the laboratory system of the current line is

$$F = \frac{\mu_0 I e v}{2\pi r} . \tag{B.5}$$

Let us now consider the same configuration in the reference system of the particle, indicated by quantities with the prime. The velocity is now zero, and the electromagnetic force due to the magnetic field disappears. Due to the relativistic effect of the contraction of space, a density of linear charge $\lambda'(I,v)$ will emerge giving rise to an electrostatic force (see Fig. B.2)

$$F' = \frac{e\lambda'}{4\pi\varepsilon_0}\int_{-\infty}^{\infty}\frac{ds}{s^2+r^2} = \frac{e\lambda'}{4\pi\varepsilon_0 r}\int_{-\infty}^{\infty}\frac{dy}{y^2+1} = \frac{e\lambda'}{2\pi\varepsilon_0 r} . \tag{B.6}$$

This very simple computation shows that (i) the electrostatic force has the same direction of the vector product between magnetic field and distance from the particle to the current wire, and (ii) the force is proportional to the inverse of the distance. Moreover, since

$$c^2 = \frac{1}{\mu_0\varepsilon_0} \tag{B.7}$$

one can rewrite (B.6) as

$$F' = \frac{e\lambda'}{2\pi\varepsilon_0 r} = \frac{e\mu_0 c^2\lambda'}{2\pi r} . \tag{B.8}$$

The force in the laboratory reference system is related to the force in the system of the particle via

$$F' = \gamma_r F = \frac{F}{\sqrt{1-\frac{v^2}{c^2}}} \tag{B.9}$$

and therefore the linear charge density in the particle reference system is related to the current in the laboratory system by

$$\lambda'(I,v) = \frac{Iv}{c^2}\gamma_r \,. \tag{B.10}$$

This last expression can be derived by the special relativity contraction of lengths; we do not give it here for the sake of brevity, and we refer to [2] for the full computation.

The argument is on the one side totally trivial: as the Maxwell equations are invariant, and the Lorentz force is proportional to the particle velocity, a force due to magnetic field for a moving particle has to be seen as an electrostatic force in the reference system of the particle, where its velocity is zero. On the other side, the argument presents the magnetic field as an "apparent" field due to use of the "wrong" reference system, i.e., as a side effect of special relativity. Note that apparent forces are defined as forces arising in references systems not in uniform motion (for instance the centripet force for a particle on a circular orbit). In this case we put apparent between quotes since the reference system is in uniform motion. We would also add that since the fact that the magnetic field is a relativistic effect, explains why it is so difficult to make "large" magnetic fields, and to store "large" energies in magnetic fields.

B.2 Electrostatic divergence and electron classical radius

Electrostatics has a divergence for infinite charge densities, since the energy associated to its electric field is infinite. This means that a charged particle must have a finite size. Let us analyse this divergence; the energy associated electrostatic field is

$$U = \int \frac{\varepsilon_0 E^2}{2} dV \tag{B.11}$$

and the electric field is given by

$$E = \frac{e}{4\pi\varepsilon_0 r^2}\,; \tag{B.12}$$

therefore if we consider the electron as a sphere of dimension r_e, the energy associated to the electric field outside the sphere is given by

$$U = \int \frac{\varepsilon_0 E^2}{2} dV = 4\pi \frac{\varepsilon_0}{2} \int_{r_e}^{\infty} \frac{e^2}{16\pi^2 \varepsilon_0^2 r^4} r^2 dr = \frac{e^2}{8\pi\varepsilon_0 r_e}\,. \tag{B.13}$$

Equating this energy to half of the electron rest energy, one defines the classical electron radius r_e

$$\frac{mc^2}{2} \equiv \frac{e^2}{8\pi\varepsilon_0 r_e} \qquad r_e \equiv \frac{e^2}{4\pi\varepsilon_0 mc^2}\,. \tag{B.14}$$

Why equating to only half of the rest energy? Because there is also a potential energy associated to having a diffused charge of integrated value e on the surface of a sphere of radius r_e. One can prove that for the classical electron radius this energy is equal to half of the electron rest energy:

$$U = \frac{1}{4\pi\varepsilon_0} \iint_{r=r_e} \frac{dq_1 dq_2}{r_1 - r_2} = \frac{mc^2}{2} \tag{B.15}$$

so the classical electron radius is the radius of a charged sphere whose electrostatic energy corresponds to the rest mass. Its value is six orders of magnitudes below the nanometer:

$$r_e \equiv \frac{e^2}{4\pi\varepsilon_0 mc^2} = \frac{\left(1.60\times10^{-19}\right)^2}{4\times3.14\times8.85\times10^{-12}\times9.11\times10^{-31}\times\left(3.00\times10^{8}\right)^2} = 2.8\times10^{-15}\ \text{m} \quad . \tag{B.16}$$

The classical electron radius appears in the cross-section for the Thomson scattering [3]

$$\sigma_T = \frac{8\pi}{3}\left(\frac{e^2}{4\pi\varepsilon_0 mc^2}\right)^2 , \tag{B.17}$$

which is the cross-section of the elastic scattering of electromagnetic radiation on a free charged particle, with the photon energy much smaller than the rest energy of the particle. This is a purely classical formula based on Maxwell equations. In this regime, the "size" of the electron is the classical electron radius.

B.3 Magnetic divergence of the inductance of a current loop

A second divergence of electromagnetism is related to the current density: Maxwell equations have a divergence in the magnetic field for infinitely large current densities. Let us compute the energy stored in the magnetic field created by a circular current loop of radius r_L. This circuit has a stored energy

$$U = \int \frac{B^2}{2\mu_0} dV \quad . \tag{B.18}$$

In the neighbourhood of the current line the field diverges, since it is proportional to the inverse of the distance to the current line (Biot-Savart law)

$$B = \frac{\mu_0 I}{2\pi r} . \tag{B.19}$$

Therefore if we consider a current loop made with a thin wire of radius r_w, the energy will be concentrated in the neighbourhood of the wire, and we can estimate the integral up to a distance R to the wire as

$$U \approx \frac{2\pi r_L}{2\mu_0} 2\pi \int_{r_w}^{R} \left(\frac{\mu_0 I}{2\pi r}\right)^2 r dr = \frac{\mu_0 r_L I^2}{2} \int_{r_w}^{R} \frac{dr}{r} = \frac{\mu_0 r_L I^2}{2} \ln\left(\frac{R}{r_w}\right) \tag{B.20}$$

and the energy diverges with the logarithm of the inverse of the wire radius r_w. Rewriting the equation in terms of the current density

$$j = \frac{I}{\pi r_w^2} \qquad r_w = \sqrt{\frac{I}{\pi j}} \tag{B.21}$$

one finds that the divergence in the stored energy is proportional to the logarithm of the current density:

$$U \approx \frac{\mu_0 r_L I^2}{2} \ln\left(\frac{R}{r_w}\right) = \frac{\mu_0 r_L I^2}{2} \ln\left(R\sqrt{\frac{\pi j}{I}}\right) \propto \ln j \,. \tag{B.22}$$

As a consequence, the inductance of a loop of current also has a logarithmic divergence: since

$$U = \frac{LI^2}{2} \tag{B.23}$$

one finds

$$L \approx \mu_0 r_L \ln\left(\frac{R}{r_w}\right). \tag{B.24}$$

In many numerical codes estimating the magnetic field generated by a coil, one considers the contributions coming from each wire of the coil as due to infinite current densities concentrated in the centre of the wire. This is very effective when the field is computed outside of the coil. In order to estimate the field inside the coil (we will see this is particularly important for superconducting magnets, since the current density is limited by the field in the coil) one can still use this approach, but to avoid the singularity one adds the contribution of the single wire (self field) given by the field on the edge of the wire (see Fig. B.3). Inside the wire, this contribution will decrease linearly up to reach zero in the centre of the wire.

$$B_{sf} = \frac{\mu_0 I_w}{2\pi r_w}. \tag{B.25}$$

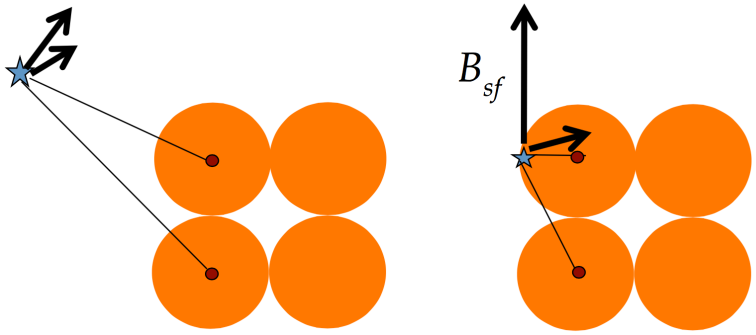


Fig. B.3: Estimating the field given by finite size current wires (orange circles) via infinite current densities: case outside the coil winding (left) and inside the winding, with estimate of the self field (right).

B.4 Losing the scale invariance: the revolution of 1865-1916, and the fine structure constant

In the second half of the nineteen century, a major conceptual revolution takes place in physics: the loss of the scale invariance. Since Galileo, physics has no ceiling neither floors, and all quantities can assume values arbitrarily small or large. It is true that the existence of a lower bound to temperatures was already since the eighteen century (Guillame Amontons was estimating it as -240° C in 1702): however, as the temperature is a measure of molecular motion, it is perfectly inuitive that there is a

minimal value for temperature, corresponding to the absence of motion. Here we discuss two main events breaking the scale invariance in a much less intuitive way, since they concern physical quantities that are a combination of space and time: an upper limit for the speed and a lower limit for the angular momentum.

Maxwell equations, written in 1865, contain the speed of light *c*, see Eq. (B.2) and therefore this quantity has to be invariant in all reference systems in uniform motion. This implies that Maxwell equations are not invariant for Galileo transformations; in 1892, Lorentz finds the transformations that make Maxwell equations invariants for reference systems in uniform translation [4]; note that this is not an isolated work but part of the general effort of the community, including prestigious names as Voigt, Heaviside, Thomson and Poincaré. Finally, Einstein gives a physical interpretation to Lorentz transformations building the special relativity in 1905 [5]. In Lorentz transformations, the speed of light cannot be overcome (it would give the square root of a negative number), and therefore an upper limit for velocity appears. The *c* in Maxwell equations is like an alien in the body of the 19th century physics announcing the new physics of the 20th century.

In the same years, a lower limit to angular momentum is discovered: in order to model the black body radiation, Max Planck has to make the hypothesis that the energy of photons is quantized according to

$$U = h\nu \tag{B.26}$$

where *h* is the Planck constant and ν the frequency of the photon [6]. This quantization turns out to be a quantization of angular momentum:

$$L = n\frac{h}{2\pi}; \tag{B.27}$$

any angular momentum can be an integer multipole of *h* (with the exception of the intrinsic angular momentum of certain particles, we will come back to this point in section B.8). From an intuitive point of view, a lower limit to angular momentum is as difficult to imagine as an upper limit to the speed.
The product of the speed of light and of the Planck constant has the physical dimension of a square of an electric charge: we prove it using the more confortable CGS system where the electrostatic energy is

$$U = \frac{e^2}{r} \qquad \left[e^2\right] = \left[Ur\right] = \text{g cm}^2\text{s}^{-2}\text{cm} = \text{g cm}^3\text{s}^{-2} \tag{B.28}$$

and the dimension of the Planck constant time the speed of light is

$$\left[hc\right] = \text{g cm}^2\text{s}^{-1}\text{cm s}^{-1} = \text{g cm}^3\text{s}^{-2}\ . \tag{B.29}$$

Therefore the quantity

$$\alpha \equiv 2\pi\frac{e^2}{hc} \tag{B.30}$$

is adimensional. The above expression holds in the CGS system: in the international system, that we use in these notes, unless explicitely stated, one has

$$\alpha \equiv \frac{e^2}{2\varepsilon_0 hc}. \tag{B.31}$$

This constant was introduced by A. Sommerfeld [7]; its numerical value is order of 0.01:

$$\alpha \equiv \frac{e^2}{2\varepsilon_0 hc} = \frac{\left(1.60\times10^{-19}\right)^2}{2\times8.85\times10^{-12}\times6.63\times10^{-34}\times3.00\times10^{8}} = \frac{1}{137.036...}. \tag{B.32}$$

R. Fenymann defined the fine structure constant as "one the biggest enigma of physics" [8]. We will see in the coming sections the origin of its name, its appearance in many occasions, and how its name is far from being adequate to describe its role in physics. Notwithstanding the relevance of his work, Sommerfeld was never awarded the Nobel prize.

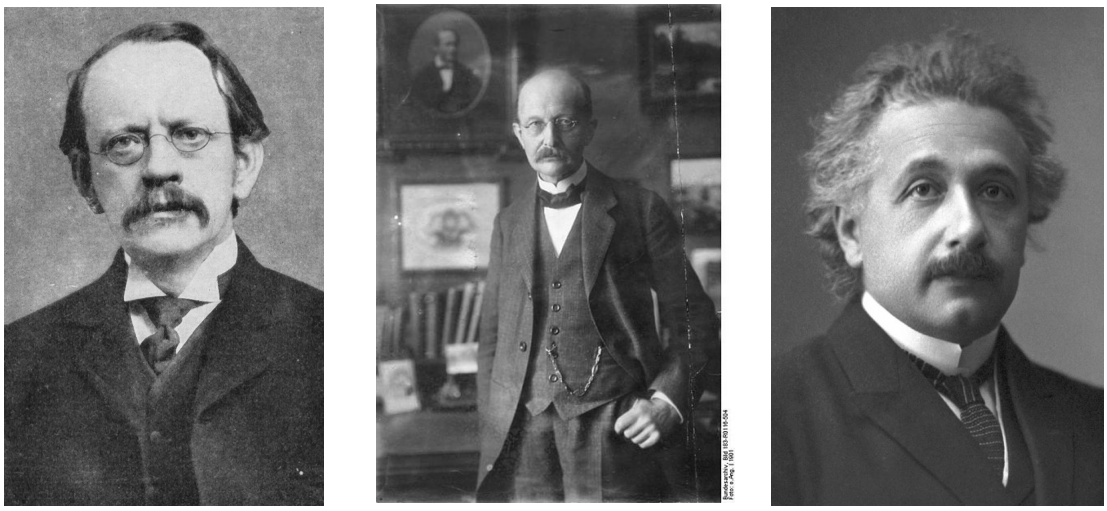


Fig. B.4: Joseph John Tompson, 1856-1940, Nobel prize in 1906, British (left), Max Planck, 1858-1947, Nobel prize in 1918, German (center), and Albert Einstein, 1879-1955, Nobel prize in 1921, German (left).

B.5 How far the classical electron radius is from atom size: a reminder on Bohr radius

The size of an atom is of the order of the Bohr radius, that can be derived by a semiclassical model (the Bohr atom) based on the electrostatic force between a charged particle *e* of mass *m* in the field of another charge *e*, of infinite mass

$$U = \frac{p^2}{2m} - \frac{e^2}{4\pi\varepsilon_0 r} \tag{B.33}$$

plus the quantization of the angular momentum (B.27), see for instance a classical textbook [3]. Considering a circular orbit of radius *R*, and the quantization of the angular momentum is

$$pR = n\frac{h}{2\pi}, \tag{B.34}$$

replacing the momentum in in the energy equation one has

$$U = n^2\frac{h^2}{8\pi^2 mR^2} - \frac{e^2}{4\pi\varepsilon_0 R} \tag{B.35}$$

and minimizing the energy and for *n*=1 one finds a minimum corresponding to the Bohr radius

$$\frac{dU}{dR} = -n^2 \frac{h^2}{16\pi^2 m R^3} + \frac{e^2}{4\pi\varepsilon_0 R^2} = -\frac{e^2}{4\pi\varepsilon_0 R^3}\left(n^2 \frac{\varepsilon_0 h^2}{\pi e^2 m} - R\right). \tag{B.36}$$

$$r_B \equiv \frac{\varepsilon_0 h^2}{\pi e^2 m} = \frac{8.85\times10^{-12}\times\left(6.63\times10^{-34}\right)^2}{3.14\times9.11\times10^{-31}\times\left(1.60\times10^{-19}\right)^2} = 5.3\times10^{-11}\ \text{m}. \tag{B.37}$$

The Bohr radius is about 1/20 of a nanometer, i.e. about four orders of magnitudes larger than the classical electron radius. It well agrees with the average amplitude of the lower energy orbital of the hydrogen atom that can be computed via the Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}\Psi = -\frac{\hbar^2}{2m}\nabla^2\Psi - \frac{e^2}{4\pi\varepsilon_0 R}\Psi. \tag{B.38}$$

The bound energy associated to the different orbital levels of Bohr model are

$$U\left(n^2 r_B\right) = \frac{1}{n^2}\frac{h^2}{8\pi^2 m r_B^2} - \frac{1}{n^2}\frac{e^2}{4\pi\varepsilon_0 r_B} = \frac{1}{n^2}\frac{e^2}{4\pi\varepsilon_0 r_B^2}\left(\frac{\varepsilon_0 h^2}{2\pi e^2 m} - r_B\right) = -\frac{1}{n^2}\frac{me^4}{8\varepsilon_0^2 h^2}. \tag{B.39}$$

The first evidence of the $1/n^2$ dependence came from spectroscopy; even though the energy can be written in terms of the constant α with the simple expression

$$U_n \equiv U\left(n^2 r_B\right) = -\frac{1}{n^2}\frac{me^4}{8\varepsilon_0^2 h^2} = -\frac{\alpha^2}{2n^2}mc^2, \tag{B.40}$$

as the observable is the frequency of the photons emitted/absorbed during a change of orbital, and not the energy, the measured quantity was not α but rather what was defined as the Rydberg constant R

$$\nu_n = \frac{U_n}{h} = -\frac{1}{n^2}\frac{me^4}{8\varepsilon_0^2 h^3} = -\frac{R}{n^2} \qquad R \equiv \frac{me^4}{8\varepsilon_0^2 h^3}. \tag{B.41}$$

Going to more and more precise measurements, it was found that there the spectra had a further splitting on finer levels, characterized by an additional number k, and given by

$$\nu_{n,k} = \frac{R}{n^2}\left(1 + \frac{\alpha^2}{n^2}\left(\frac{n}{k} - \frac{3}{4}\right) + O(\alpha^4)\right). \tag{B.42}$$

This equation was empirically found by Sommerfeld based on the measurements of spectral lines of hydrogen atom: in this case α becomes a direct observable from the splitting of the finer levels, and this is the origin of its name. As we will see in the next sections, this constant has a much deeper physical meaning than just a second order effect on energy levels of atoms. Note that the Eq. (B.42) can be derived as a solution of the Schrödinger equation.

Finally, we estimate the speed of the electron in the Bohr model: since

$$mr_B v_B = \frac{h}{2\pi} \tag{B.43}$$

one has

$$v_B = \frac{h}{2\pi m r_B} = \frac{e^2}{2\varepsilon_0 h} = \alpha c \tag{B.44}$$

i.e. the speed is about 137 times slower than the speed of light. This justifies the use of Schrödinger equation (based on non relativistic mechanics) for the estimate of the atomic levels. This is the third appearance of the fine structure constant in these notes, after (B.40) and (B.42).

B.6 Compton wavelength

We have shown in the previous section that the classical electron radius is about four order of magnitudes smaller than the Bohr radius. More precisely, the ratio between the two quantities is the square of α:

$$\frac{r_B}{r_e} = \frac{\varepsilon_0 h^2}{\pi m e^2} \frac{4\pi\varepsilon_0 m c^2}{e^2} = \frac{4\varepsilon_0^2 h^2 c^2}{e^4} = \frac{1}{\alpha^2} \tag{B.45}$$

and we find here the fourth appearance of the fine structure constant. In between (in a logarithmic sense) the Bohr radius and the classical electron radiu one finds a quantity having a pretty simple expression (see Fig. B.5)

$$\alpha r_B = \frac{e^2}{2\varepsilon_0 hc} \frac{\varepsilon_0 h^2}{\pi m e^2} = \frac{h}{2\pi mc} \tag{B.46}$$

usually defined as the Compton radius, or the reduced Compton wavelength

$$r_C \equiv \frac{h}{2\pi mc} = \frac{6.63\times10^{-34}}{2\times3.14\times9.11\times10^{-31}\times3.00\times10^{8}} = 3.9\times10^{-13}\ \text{m} \tag{B.47}$$

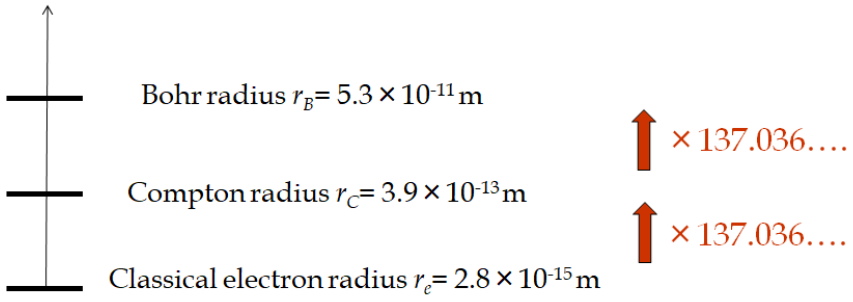


Fig. B.5: The scale of atomic lengths

The Compton wavelength is usually introduced in the Compton scattering [3,9], i.e. the same scattering between electromagnetic radiation (photon) and a charged electron as in Thomson scattering, but when there is a transfer of energy between the photon and the electron (inelastic scattering). Since the photons carry an energy proportional to their frequency ν, or inverse proportional to the wavelength λ

$$U = h\nu = \frac{hc}{\lambda} \tag{B.48}$$

the photon after the scattering has a smaller frequency (larger wavelength) according to

$$\Delta\lambda = \lambda' - \lambda = \frac{h}{mc}\left(1 - \cos\theta\right) \tag{B.49}$$

where θ is the angle of the scattered photon (see Fig. B.6). This equation can be deduced by conservation of momentum, assuming that the photon is a particle with momentum

$$p = \frac{h}{\lambda}. \tag{B.50}$$

Compton wavelength was defined as the observable of the experiment as

$$\lambda_C \equiv \frac{h}{mc} \tag{B.51}$$

allowing to rewrite (B.49) as

$$\Delta\lambda = \lambda' - \lambda = \lambda_C\left(1 - \cos\theta\right) \tag{B.52}$$

and is related to the Comption radius via

$$r_C = \frac{\lambda_C}{2\pi} = \frac{h}{2\pi mc} \tag{B.53}$$

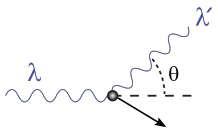


Fig. B.6: Compton scattering: change of photon wavelentgh during inelastic scattering with electron.

Fig. B.7: Arnold Sommerfeld, 1868-1951, German (left), Niels Bohr, 1885-1962, Nobel prize in 1922, Danish (centre), and Arthur Holly Compton, 1892-1962, Nobel prize in 1927, American (right).

B.7 Wrong naming

Both the Compton length and the fine structure constant suffer from the same disgrace: they keep their historical name that does not account for their real physical meaning. It would be much better to call the first one the scale length in electromagnetic interactions, and the second the electromagnetic coupling constant normalized through the two quantities breaking the scale invariance, i.e. the Planck constant and the speed of light.

Note that the equation of Schrödinger, usually written as

$$i\hbar\frac{\partial}{\partial t}\Psi = -\frac{\hbar^2}{2m}\nabla^2\Psi - \frac{e^2}{4\pi\varepsilon_0 r}\Psi \tag{B.54}$$

can be rewritten as

$$\frac{i}{c}\frac{\partial}{\partial t}\Psi = -\frac{\hbar}{2mc}\nabla^2\Psi - \frac{e^2}{4\pi\varepsilon_0\hbar c r}\Psi \qquad \frac{i}{c}\frac{\partial}{\partial t}\Psi = -\frac{r_C}{2}\nabla^2\Psi - \frac{\alpha}{r}\Psi \tag{B.55}$$

where the Compton radius is the fundamental length and the fine structure constant is the coupling constant of the electromagnetic field.

Fig. B.8: Erwin Schrödinger, 1887-1961, Nobel prize in 1933, Austrian (left), Paul Dirac, 1902-1984, Nobel prize in 1933, English (centre), and Richard Feynmann, 1918-1988, Nobel prize in 1965, American (right).

The Dirac equation, describing the intrinsic angular momentum of the electrons (spin, see next section) usually written as

$$i\gamma^\mu\partial_\mu\Psi + \frac{mc}{\hbar}\Psi = 0 \tag{B.56}$$

that can be cast in the more explicit form

$$i\gamma^\mu\partial_\mu\Psi + \frac{1}{r_C}\Psi = 0 \tag{B.57}$$

and finally, the equivalent of the Schödinger equation in the relativistic case, and in absence of any field is the Klein Gordon equation

$$\nabla^2\Psi - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\Psi = \left(\frac{mc}{\hbar}\right)^2 \Psi \tag{B.58}$$

which can also be rewritten as

$$\nabla^2\Psi - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\Psi = \frac{1}{r_C^2}\Psi\,. \tag{B.59}$$

Quantum electrodynamics is built as a perturbative expansion in powers of alpha (the Feynman diagrams), and the small value of the constant allows to effectively use a perturbative approach (i.e. the higher orders are smaller and smaller).

B.8 Intrinsic angular momentum and magnetic momentum of the electron

The electron, a structureless particle, has two intrinsic properties suggesting that something is spinning around. Firstly, the electron has intrinsic angular momentum equal to half of the quantum of angular momentum $h/2\pi$

$$L = \frac{1}{2}\frac{h}{2\pi}. \tag{B.60}$$

The spin of the electron, put in evidence once more by experimental measurements of atomic spectra, is such a puzzling feature that it goes untranslated in the other languages: to spin, that in English means to rotate around its axis, becomes as a proper noun in the other languages losing its original physical meaning for the non-English speaking community.

The electron also has an intrinsic magnetic momentum, that can be approximated by the Bohr magneton, defined as

$$m_e \approx m_B \equiv \frac{eh}{4\pi m}\,. \tag{B.61}$$

Note that the Bohr magneton corresponds to the magnetic momentum of a charge e spinning at the speed of light on a circle with Compton radius r_C:

$$m_B = SI = \pi r_C^2 \frac{c}{2\pi r_C} e = \frac{ecr_C}{2}\,. \tag{B.62}$$

Its numerical value is

$$m_B = \frac{ecr_C}{2} = \frac{1.6\times10^{-19}\times3.0\times10^{8}\times3.9\times10^{-13}}{2} = 9.4\times10^{-24}\ \text{A m}^2\,. \tag{B.63}$$

The intrinsic magnetic momentum of the electron is the main source of ferromagnetism, and therefore plays an important role in the design of superconducting magnets (see Chapter 9). Let us estimate the

field that can be obtained via an alignment of all these intrinsic magnet momentums; it will be given by

$$B = \mu_0 N_V m_B \ . \tag{B.64}$$

where N_V is the number of atoms per unity of volume. In the case of iron, atomic number is 56, density is 7800 kg/m^3, and therefore

$$N_V = \frac{7800}{56 \times 1.7 \times 10^{-27}} = 8.2 \times 10^{28} \ \text{m}^{-3} \ . \tag{B.65}$$

Finally, the maximum field achievable in the iron is order of

$$B = \mu_0 N_V m_B = 4\pi \times 10^{-7} \times 8.2 \times 10^{28} \times 9.3 \times 10^{-24} = 0.96 \ \text{T} \, . \tag{B.66}$$

Considering also the momentum given by the orbit, that in the simple case of the lowest level of the hydrogen atom is the same as the intrinsic magnetic momentum

$$m_o = SI = \pi r_B^2 \frac{c}{2\pi\alpha r_B} e = \frac{ecr_B}{2\alpha} = \frac{ecr_C}{2} = m_B \tag{B.67}$$

one doubles the value given in (B.66). Note that saturation in the iron is close to 2 T, but the contribution to this field is dominated given by the intrinsic momentum of the electron, and not equally shared between orbital momentum and intrinsic as in this semiclassical model. The accurate computation of the saturation fields in ferromagntic materials requires models based on quantum mechanics; however the semiclassical approach shown here gives the correct order of magnitude.

One striking feature of the electron is that a charged mass revolving on any orbit has a gyromagnetic ratio γ defined as

$$\gamma_e = \frac{m}{L} = \frac{SI}{L} = \frac{\pi r^2}{mrv} \frac{ev}{2\pi r} = \frac{e}{2m} \tag{B.68}$$

on the other hand, for the electron one has

$$\gamma_e = \frac{m_e}{L_e} \approx \frac{m_B}{L_e} = \frac{eh}{4\pi m} \frac{4\pi}{h} = \frac{e}{m} \tag{B.69}$$

and therefore the gyromagnetic ratio is about twice the value of a pointless particle of charge *e* and mass *m* on a orbit: this means that the electron cannot by modeled as a rotating charge. Among the numerous attempts to attribute a physical meaning to the intrinsic angular/magnetic momenta of the electron in the framework of semiclassical models, we wish to quote the work of McGregor [10], who assumes that the electron is a sphere with the dimension of the Compton radius, turning at the speed of light at its surface, and the electrical charge is localized in a spot on the surface; this awkward model gives the correct gyromagnetic ratio since charge and mass are not localized in the same locations. The same work puts in evidence the appearance of the fine structure constant in the relation between particle masses.

The g-factor ratio of the electron is defined as the ratio between the gyromagnetic ratio of the electron and its classical value $e/2m$; as we said at the beginning of this section, it is close to 2. The g-factor is one of the physical quantities known with higher precision: the measured value has twelve digits

$$g_e = \frac{\gamma_e}{e/2m} = 2.0023193043617\,. \tag{B.70}$$

The computation of the g-factor of the electron is one of the most spectacular success of the quantum electrodynamics, and it can be done in an approximate way via perturbative expansions (Feynmann diagrams) in a series expansion in powers of α. Since α is order of 1/100, the second order already provides three significant digits.

$$g_e \approx 2\left(1+\frac{\alpha}{2\pi}+\ldots\right) = 2\left(1+0.00116+O(\alpha^2)\right) = 2.00232+O(\alpha^2) \tag{B.71}$$

At the moment of writing these notes, experiments in Fermilab accelerator complex aiming at measuring the g-factor of the muon are pointing out a discrepancy with the values estimated through quantum electrodymanics [11]: this could reveal a new physics behind our present knowledge.

Acknowledgements

I wish to thank C. Benvenuti for numerous passionate discussions on the modeling in atomic physics.

References

1. B. Acharya, et al., "Search for magnetic monopoles produced via the Schwinger mechanism" *Nature* **602** (2022) 63-67
2. E. M. Purcell, et al., "Electricity and magnetism (Berkeley Physics Course, Vol. 2)" MacGraw-Hill (New York, 1973)
3. M. Born, "Atomic physics" Blackie & Son (London, 1937) 48
4. H. A. Lorentz, "La Théorie electromagnétique de Maxwell et son application aux corps mouvants"*Archives Néerlandaises des Sciences Exactes et Naturelles* **25** (1892) 363–552
5. A. Einstein, "On the electrodynamics of moving bodies" *Ann. Phys.* **17** (1905) 891-921
6. M. Planck, "On the law of distribution of energy in the normal spectrum" *Ann. Phys.* **4** (1901) 553
7. A. Sommerfeld, "Zur Quantentheorie der Spektrallinen" *Ann. Phys.* **4** (1916) 51.
8. R. Feynmann, "QED: the strange theory of light and matter" Princeton University Press (1985) 129
9. A. H. Compton "A quantum theory of the scattering of X-rays by light elements" *Phys. Rev.* **21** (1923) 483-502
10. M. MacGregor "The power of alpha" World Scientific (2007)
11. B. Abi, et al., "Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm" *Phys. Rev. Lett.* **126** (2021) 141801