

**ORGANISATION EUROPÉENNE POUR LA RECHERCHE NUCLÉAIRE
CERN EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH**



Magnets

Bruges, Belgium
16 – 25 June 2009

Proceedings

Editor: D. Brandt

GENEVA
2010

ISBN 978–92–9083–355–0

ISSN 0007–8328

Copyright © CERN, 2010

 Creative Commons Attribution 3.0

Knowledge transfer is an integral part of CERN’s mission. CERN publishes this report Open Access under the Creative Commons Attribution 3.0 license (<http://creativecommons.org/licenses/by/3.0/>) in order to permit its wide dissemination and use.

This report should be cited as:

Proceedings of the CAS–CERN Accelerator School: Magnets, Bruges, Belgium, 16–25 June 2009,
edited by D. Brandt, CERN-2010-004.

A contribution in this report should be cited as:

[Author name(s)], in Proceedings of the CAS–CERN Accelerator School: Magnets, Bruges, Belgium,
16–25 June 2009, edited by D. Brandt, CERN-2010-004, pp. [first page]–[lastpage]

Abstract

These proceedings present the lectures given at the twenty-third specialized course organized by the CERN Accelerator School (CAS), the topic being ‘Magnets’. The course was held in Bruges, Belgium, from 16 to 25 June 2009. This is the first time this topic has been selected for a specialized course. Taking into account the number of related applications currently in use in accelerators around the world, but, even more important, the worrying decrease in the corresponding expertise in the different laboratories, it was recognized that such a topic should definitively be incorporated into the CAS series of specialized courses. The specific aim of the course was to introduce the participants to the basics of resistive magnet design and its underlying theoretical concepts.

The first part of the school dealt with basic introductory courses such as Maxwell’s equations for magnets, beam optics, physics and measurement of magnetic materials, the different types of resistive magnets and their respective performance, an introduction to numerical field computation, and a core lecture on basic magnet design.

The second part of the course focused more on quality control, the different measurement systems with their electronics, calibration techniques and respective applications as well as the question of stability and reproducibility.

For the first time, in addition to the actual lectures, a Case Study was proposed to the participants. This consisted of eight hours of a guided practical exercise, where the participants had to propose their own design for a magnet fulfilling the boundary conditions corresponding to a combined-function magnet developed for the ALBA Synchrotron Light Source in Barcelona, Spain. This Case Study was enthusiastically received by the participants, who praised both the proposed approach and the amount of practical information acquired from this exercise.



Foreword

The aim of the CERN Accelerator School (CAS) to collect, preserve, and disseminate the knowledge accumulated in the world's accelerator laboratories applies not only to general accelerator physics, but also to related sub-systems, equipment, and technologies. This wider aim is achieved by means of specialized courses. The topic of the 2009 course was Resistive Magnets and was held at the Novotel Brugge Centrum, Bruges, Belgium, from 16 to 25 June 2009.

This was the first time that 'Resistive Magnets' was the topic of a CAS specialized course. The motivation for selecting this topic was twofold: first, resistive magnets represent an essential component of any accelerator and, secondly, a slow but constant decrease in the existing expertise in the different laboratories was a clear indication of the need for such training.

The present course was made possible by the active collaboration provided by the firm IBA (Louvain-la-Neuve, Belgium) and many individuals. In particular, the contributions of the Local Organizing Committee composed of Micheline Desmedt (IBA) and Dirk Vandeplassche (SCK-CEN) were most valuable. Similarly, the help of Louis Walckiers (CERN) in identifying contacts to propose possible venues was highly appreciated.

The generous financial support provided by Dr. Yves Jongen (IBA Chief Research Officer Managing Director) allowed CAS to offer scholarships to highly deserving young students, who would not have been able to attend the school otherwise.

As far as the course is concerned, it is worth mentioning that, for the first time at CAS, a 'Case Study' was proposed to the participants. It consisted of eight hours of a guided practical exercise, where the participants had to propose their own solution for the design of a magnet fulfilling the boundary conditions corresponding to a combined-function magnet developed for the ALBA Synchrotron Light Source in Barcelona, Spain. This Case Study proved to be extremely successful and similar practical approaches will certainly be considered for future specialized courses. The impressive success of the Case Study is most definitely related to the outstanding contributions of our colleagues Dieter Einfeld (ALBA- CELLS) and Davide Tommasini (CERN) who invested so much energy and enthusiasm in preparing, running, and evaluating the exercise.

As always, the backing of the CERN management, the guidance of the CAS Advisory and Programme Committees, the attention to detail of the Local Organizing Committee and the management and staff of the Novotel Brugge Centrum ensured that the School was held under optimum conditions. Special thanks must go to the lecturers for the preparation, presentation, and writing up of their topics.

Finally, the enthusiasm of the participants who came from more than 20 different countries was convincing proof of the usefulness and success of the course.

It is my pleasure and privilege to thank all those persons who helped in making the course a success, including Barbara Strasser and the CERN Desktop Publishing Service for their dedication and commitment to the production of this document.

Daniel Brandt
CERN Accelerator School

PROGRAMME Magnets
16-25 June 2009, Bruges, Belgium

Time	Wednesday 17 June	Thursday 18 June	Friday 19 June	Saturday 20 June	Sunday 21 June	Monday 22 June	Tuesday 23 June	Wednesday 24 June	Thursday 25 June
08:30	Maxwell Equations for Magnets I	Overview of Measurement Methods	Magnetic Field Computations using FEM	Injection & Extraction Magnets I	Permanent Magnets including Wigglers & Undulators III	Magnets including Wigglers & Undulators III	Metrology & Alignment I	Coils and Wires	Manufacturing & Calibration of Measurement Systems II
09:30	A. Wolski	L. Bottura	S. Russenschuck	M. Barnes	J. Bahrdt	A. Lestrange	L. Walckiers	M. Buzio	
10:30	Beam Optics I	Beam Optics II	Basic Magnet Design III	Magnets including Wigglers & Undulators I	E	Specification & Quality Control, Manufacturing & Acceptance Test I	Dielectric Insulation & High Voltage Issues I	Electronics for Measurement Systems I	Hall Probes: Physics, Measurements & Applications
	D. Brandt	D. Brandt	T. Zickler	J. Bahrdt	X	D. Einfeld	D. Tommasini	P. Arpaia	S. Sanfilippo
11:00	Maxwell Equations for Magnets II	Physics & Measurements of Magnetic Materials III	Basic Magnet Design IV	Injection & Extraction Magnets II	C	Specification & Quality Control, Manufacturing & Acceptance Test II	Metrology & Alignment II		Stability & Reproducibility
12:00	A. Wolski	S. Sgobba	T. Zickler	M. Barnes	U				
12:00	Physics & Measurements of Magnetic Materials I	Magnet Types & Performances II	SC Magnets I	Permanent Magnets including Wigglers & Undulators II	R	D. Einfeld	A. Lestrange	Case Study III	N Marks
13:00	S. Sgobba	A. Dael	L. Bottura	J. Bahrdt	S	Introduction	Dielectric Insulation & High Voltage Issues II		Case Study: Summing Up
				N	O	D. Einfeld/ D. Tommasini	D. Tommasini		
14:30	Magnet Types & Performances I	Foundation of Numerical Field Computation II	SC Magnets II	Solenoids				Harmonic Coils	
15:30	A. Dael	S. Russenschuck	L. Bottura	A. Dael		Case Study I	Case Study II	L. Walckiers	
15:30	Foundation of Numerical Field Computation I	Basic Magnet Design I	Eddy Currents I	Seminar An Accelerator Driven System: MYRRHA				Electronics for Measurement Systems II	
16:30	S. Russenschuck	T. Zickler	G. Moritz	D. Vandeplassche				P. Arpaia	
17:00	Physics & Measurements of Magnetic Materials II	Basic Magnet Design II	Discussion Introduction	Discussion Session					
18:00	S. Sgobba	T. Zickler	Dinner	Special Dinner	Dinner	Dinner	Dinner	Dinner	
20:00	Welcome Drink	Dinner							

Contents

Foreword	
<i>D. Brandt</i>	v
Maxwell's equations for magnets	
<i>A. Wolski</i>	1
Physics and measurements of magnetic materials	
<i>S. Sgobba</i>	39
Basic design and engineering of normal-conducting, iron-dominated electromagnets	
<i>Th. Zickler</i>	65
Eddy currents in accelerator magnets	
<i>G. Moritz</i>	103
Injection and extraction magnets: kicker magnets	
<i>M.J. Barnes, L. Ducimetière, T. Fowler, V. Senaj, L. Sermeus</i>	141
Injection and extraction magnets: septa	
<i>M.J. Barnes, J. Borburgh, B. Goddard, M. Hourican</i>	167
Permanent magnets including undulators and wigglers	
<i>J. Bahrdt</i>	185
Specifications, quality control, manufacturing, and testing of accelerator magnets	
<i>D. Einfeld</i>	231
Dimensional metrology and positioning operations: basics for a spatial layout analysis of measurement systems	
<i>A. Lestrade</i>	273
Dielectric insulation and high-voltage issues	
<i>D. Tommasini</i>	335
Magnetic measurement with coils and wires	
<i>L. Walckiers</i>	357
Fabrication and calibration of search coils	
<i>M. Buzio</i>	387
Hall probes: physics and application to magnetometry	
<i>S. Sanfilippo</i>	423
Magnet stability and reproducibility	
<i>N. Marks</i>	463
List of Participants	485

Maxwell's equations for magnets

A. Wolski

University of Liverpool, Liverpool, UK and the Cockcroft Institute, Daresbury, UK

Abstract

Magnetostatic fields in accelerators are conventionally described in terms of multipoles. We show that in two dimensions, multipole fields do provide solutions of Maxwell's equations, and we consider the distributions of electric currents and geometries of ferromagnetic materials required (in idealized situations) to generate specified multipole fields. Then, we consider how to determine the multipole components in a given field. Finally, we show how the two-dimensional multipole description may be extended to three dimensions; this allows fringe fields, or the main fields in such devices as undulators and wigglers, to be expressed in terms of a set of modes, where each mode provides a solution to Maxwell's equations.

1 Maxwell's equations

Maxwell's equations may be written in differential form as follows:

$$\operatorname{div} \vec{D} = \rho, \quad (1)$$

$$\operatorname{div} \vec{B} = 0, \quad (2)$$

$$\operatorname{curl} \vec{H} = \vec{J} + \frac{\partial \vec{D}}{\partial t}, \quad (3)$$

$$\operatorname{curl} \vec{E} = -\frac{\partial \vec{B}}{\partial t}. \quad (4)$$

The fields \vec{B} (magnetic flux density) and \vec{E} (electric field strength) determine the force on a particle of charge q travelling with velocity \vec{v} (the Lorentz force equation):

$$\vec{F} = q (\vec{E} + \vec{v} \times \vec{B}).$$

The electric displacement \vec{D} and magnetic intensity \vec{H} are related to the electric field and magnetic flux density by

$$\begin{aligned} \vec{D} &= \epsilon \vec{E}, \\ \vec{B} &= \mu \vec{H}. \end{aligned}$$

The electric permittivity ϵ and magnetic permeability μ depend on the medium within which the fields exist. The values of these quantities in vacuum are fundamental physical constants. In SI units:

$$\begin{aligned} \mu_0 &= 4\pi \times 10^{-7} \text{ H m}^{-1}, \\ \epsilon_0 &= \frac{1}{\mu_0 c^2}, \end{aligned}$$

where c is the speed of light in vacuum. The permittivity and permeability of a material characterize the response of that material to electric and magnetic fields. In simplified models, they are often regarded as constants for a given material; however, in reality the permittivity and permeability can have a complicated dependence on the fields that are present. Note that the *relative permittivity* ϵ_r and the *relative permeability* μ_r are frequently used. These are dimensionless quantities, defined by

$$\epsilon_r = \frac{\epsilon}{\epsilon_0}, \quad \mu_r = \frac{\mu}{\mu_0}. \quad (5)$$

That is, the relative permittivity is the permittivity of a material relative to the permittivity of free space, and similarly for the relative permeability.

The quantities ρ and \vec{J} are respectively the electric charge density (charge per unit volume) and electric current density ($\vec{J} \cdot \vec{n}$ is the charge crossing unit area perpendicular to unit vector \vec{n} per unit time). Equations (2) and (4) are independent of ρ and \vec{J} , and are generally referred to as the ‘homogeneous’ equations; the other two equations, (1) and (3) are dependent on ρ and \vec{J} , and are generally referred to as the ‘inhomogeneous’ equations. The charge density and current density may be regarded as *sources* of electromagnetic fields. When the charge density and current density are specified (as functions of space, and, generally, time), one can integrate Maxwell’s equations (1)–(3) to find possible electric and magnetic fields in the system. Usually, however, the solution one finds by integration is not unique: for example, the field within an accelerator dipole magnet may be modified by propagating an electromagnetic wave through the magnet. However, by imposing certain constraints (for example, that the fields within a magnet are independent of time) it is possible to obtain a unique solution for the fields in a given system of electric charges and currents.

Most realistic situations are sufficiently complicated that solutions to Maxwell’s equations cannot be obtained analytically. A variety of computer codes exist to provide solutions numerically, once the charges, currents, and properties of the materials present are all specified, see, for example, Refs. [1–3]. Solving for the fields in realistic (three-dimensional) systems often requires a reasonable amount of computing power; some sophisticated techniques have been developed for solving Maxwell’s equations numerically with good efficiency [4]. We do not consider such techniques here, but focus instead on the analytical solutions that may be obtained in idealized situations. Although the solutions in such cases may not be sufficiently accurate to complete the design of a real accelerator magnet, the analytical solutions do provide a useful basis for describing the fields in real magnets, and provide also some important connections with the beam dynamics in an accelerator.

An important feature of Maxwell’s equations is that, for systems containing materials with constant permittivity and permeability (i.e., permittivity and permeability that are independent of the fields present), the equations are *linear* in the fields and sources. That is, each term in the equations involves a field or a source to (at most) the first power, and products of fields or sources do not appear. As a consequence, the *principle of superposition* applies: if \vec{B}_1 and \vec{B}_2 are solutions of Maxwell’s equations with the current densities \vec{J}_1 and \vec{J}_2 , then the field $\vec{B}_T = \vec{B}_1 + \vec{B}_2$ will be a solution of Maxwell’s equations, with the source given by the total current density $\vec{J}_T = \vec{J}_1 + \vec{J}_2$. This means that it is possible to represent complicated fields as superpositions of simpler fields. An important and widely used analysis technique for accelerator magnets is to decompose the field (determined from either a magnetic model, or from measurements of the field in an actual magnet) into a set of multipoles. While it is often the ideal to produce a field consisting of a single multipole component, this is never perfectly achieved in practice: the multipole decomposition indicates the extent to which components other than the ‘desired’ multipole are present. Multipole decompositions also produce useful information for modelling the beam dynamics. Although the principle of superposition strictly only applies in systems where the permittivity and permeability are independent of the fields, it is always possible to perform a multipole decomposition of the fields in free space (e.g., in the interior of a vacuum chamber), since in that region the permittivity and permeability are constants. However, it should be remembered that for nonlinear materials (where the permeability, for example, depends on the magnetic field strength), the field inside the material comprising the magnet will not necessarily be that expected if one were simply to add together the fields corresponding to the multipole components.

Solutions to Maxwell’s equations lead to a rich diversity of phenomena, including the fields around charges and currents in certain simple configurations, and the generation, transmission and absorption of electromagnetic radiation. Many existing texts cover these phenomena in detail; see, for example, the authoritative text by Jackson [5]. Therefore, we consider only briefly the electric field around a point charge and the magnetic field around a long straight wire carrying a uniform current: our main purpose

here is to remind the reader of two important integral theorems (Gauss's theorem, and Stokes's theorem), of which we shall make use later. In the following sections, we discuss analytical solutions to Maxwell's equations for situations relevant to some of the types of magnets commonly used in accelerators. These include multipoles (dipoles, quadrupoles, sextupoles, and so on), solenoids, and insertion devices (undulators and wigglers). We consider only static fields. We begin with two-dimensional fields, that is fields that are independent of one coordinate (generally, the coordinate representing the direction of motion of the beam). We will show that multipole fields are indeed solutions of Maxwell's equations, and we will derive the current distributions needed to generate 'pure' multipole fields. We then discuss multipole decompositions, and compare techniques for determining the multipole components present in a given field from numerical field data (from a model, or from measurements). Finally, we consider how the two-dimensional multipole decomposition may be extended to three-dimensional fields, to include (for example) insertion devices, and fringe fields in multipole magnets.

2 Integral theorems and the physical interpretation of Maxwell's equations

2.1 Gauss's theorem and Coulomb's law

Gauss's theorem states that for any smooth vector field \vec{a} :

$$\int_V \operatorname{div} \vec{a} dV = \int_{\partial V} \vec{a} \cdot d\vec{S},$$

where V is a volume bounded by the closed surface ∂V . Note that the area element $d\vec{S}$ is oriented to point *out* of V .

Gauss's theorem is helpful for obtaining physical interpretations of two of Maxwell's equations, (1) and (2). First, applying Gauss's theorem to (1) gives:

$$\int_V \operatorname{div} \vec{D} dV = \int_{\partial V} \vec{D} \cdot d\vec{S} = q, \quad (6)$$

where $q = \int_V \rho dV$ is the total charge enclosed by ∂V .

Suppose that we have a single isolated point charge in an homogeneous, isotropic medium with constant permittivity ϵ . In this case, it is interesting to take ∂V to be a sphere of radius r . By symmetry, the magnitude of the electric field must be the same at all points on ∂V , and must be normal to the surface at each point. Then, we can perform the surface integral in (6):

$$\int_{\partial V} \vec{D} \cdot d\vec{S} = 4\pi r^2 D.$$

This is illustrated in Fig. 1: the outer circle represents a cross-section of a sphere (∂V) enclosing volume V , with the charge q at its centre. The black arrows in Fig. 1 represent the electric field lines, which are everywhere perpendicular to the surface ∂V . Since $\vec{D} = \epsilon \vec{E}$, we find Coulomb's law for the magnitude of the electric field around a point charge:

$$E = \frac{q}{4\pi\epsilon r^2}.$$

Applied to Maxwell's equation (2), Gauss' theorem leads to:

$$\int_V \operatorname{div} \vec{B} dV = \int_{\partial V} \vec{B} \cdot d\vec{S} = 0.$$

In other words, the magnetic flux integrated over any closed surface must equal zero—at least, until we discover magnetic monopoles. Lines of magnetic flux occur in closed loops; whereas lines of electric field can start (and end) on electric charges.

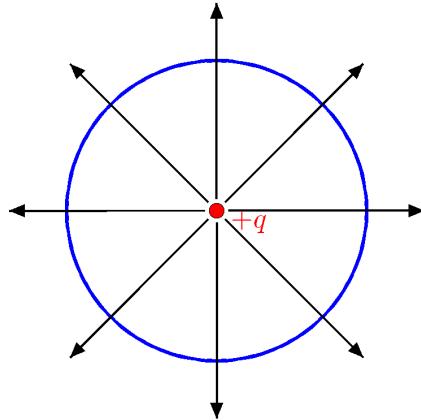


Fig. 1: Electric field lines from a point charge q . The field lines are everywhere perpendicular to a spherical surface centred on the charge.

2.2 Stokes's theorem and Ampère's law

Stokes's theorem states that for any smooth vector field \vec{a} :

$$\int_S \text{curl } \vec{a} \cdot d\vec{S} = \int_{\partial S} \vec{a} \cdot d\vec{l}, \quad (7)$$

where the loop ∂S bounds the surface S . Applied to Maxwell's equation (3), Stokes's theorem leads to

$$\int_{\partial S} \vec{H} \cdot d\vec{l} = \int_S \vec{J} \cdot d\vec{S}, \quad (8)$$

which is Ampère's law. From Ampère's law, we can derive an expression for the strength of the magnetic field around a long, straight wire carrying current I . The magnetic field must have rotational symmetry around the wire. There are two possibilities: a radial field, or a field consisting of closed concentric loops centred on the wire (or some superposition of these fields). A radial field would violate Maxwell's equation (2). Therefore, the field must consist of closed concentric loops; and by considering a circular loop of radius r , we can perform the integral in Eq. (8):

$$2\pi r H = I,$$

where I is the total current carried by the wire. In this case, the line integral is performed around a loop ∂S centred on the wire, and in a plane perpendicular to the wire: essentially, this corresponds to one of the magnetic field lines, see Fig. 2. The total current passing through the surface S bounded by the loop ∂S is simply the total current I .

In an homogeneous, isotropic medium with constant permeability μ , $\vec{B} = \mu_0 \vec{H}$, we obtain the expression for the magnetic flux density at distance r from the wire:

$$B = \frac{I}{2\pi\mu r}. \quad \boxed{\text{---}}$$
 (9)

This result will be useful when we come to consider how to generate specified multipole fields from current distributions.

Finally, applying Stokes's theorem to the homogeneous Maxwell's equation (4), we find

$$\int_{\partial S} \vec{E} \cdot d\vec{l} = -\frac{\partial}{\partial t} \int_S \vec{B} \cdot d\vec{S}. \quad (10)$$

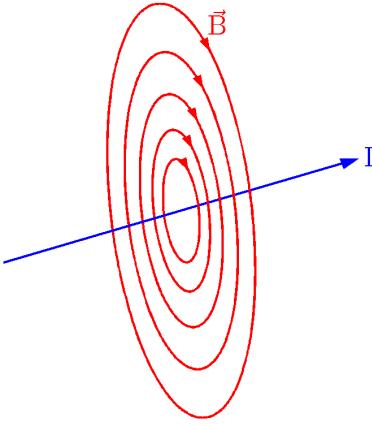


Fig. 2: Magnetic field lines around a long straight wire carrying a current I

Defining the electromotive force \mathcal{E} as the integral of the electric field around a closed loop, and the magnetic flux Φ as the integral of the magnetic flux density over the surface bounded by the loop, Eq. (10) gives

$$\mathcal{E} = -\frac{\partial \Phi}{\partial t}, \quad (11)$$

which is Faraday's law of electromagnetic induction. Faraday's law is significant for magnets with time-dependent fields, such as pulsed magnets (used for injection and extraction), and magnets that are 'ramped' (for example, when changing the beam energy in a storage ring). The change in magnetic field will induce a voltage across the coil of the magnet that must be taken into account when designing the power supply. Also, the induced voltages can induce eddy currents in the core of the magnet, or in the coils themselves, leading to heating. This is an issue for superconducting magnets, which must be ramped slowly to avoid quenching [6].

2.3 Boundary conditions

Gauss's theorem and Stokes's theorem can be applied to Maxwell's equations to derive constraints on the behaviour of electromagnetic fields at boundaries between different materials. Here, we shall focus on the boundary conditions on the magnetic field: these conditions will be useful when we consider multipole fields in iron-dominated magnets.

Consider first a short cylinder or 'pill box' that crosses the boundary between two media, with the flat ends of the cylinder parallel to the boundary, see Fig. 3 (a). Applying Gauss's theorem to Maxwell's equation (2) gives

$$\int_V \operatorname{div} \vec{B} dV = \int_{\partial V} \vec{B} \cdot d\vec{S} = 0,$$

where the boundary ∂V encloses the volume V within the cylinder. If we take the limit where the length of the cylinder ($2h$) approaches zero, then the only contributions to the surface integral come from the flat ends; if these have infinitesimal area dS , then since the orientations of these surfaces are in opposite directions on opposite sides of the boundary, and parallel to the normal component of the magnetic field, we find

$$-B_{1\perp} dS + B_{2\perp} dS = 0,$$

where $B_{1\perp}$ and $B_{2\perp}$ are the normal components of the magnetic flux density on either side of the boundary. Hence

$$B_{1\perp} = B_{2\perp}. \quad (12)$$

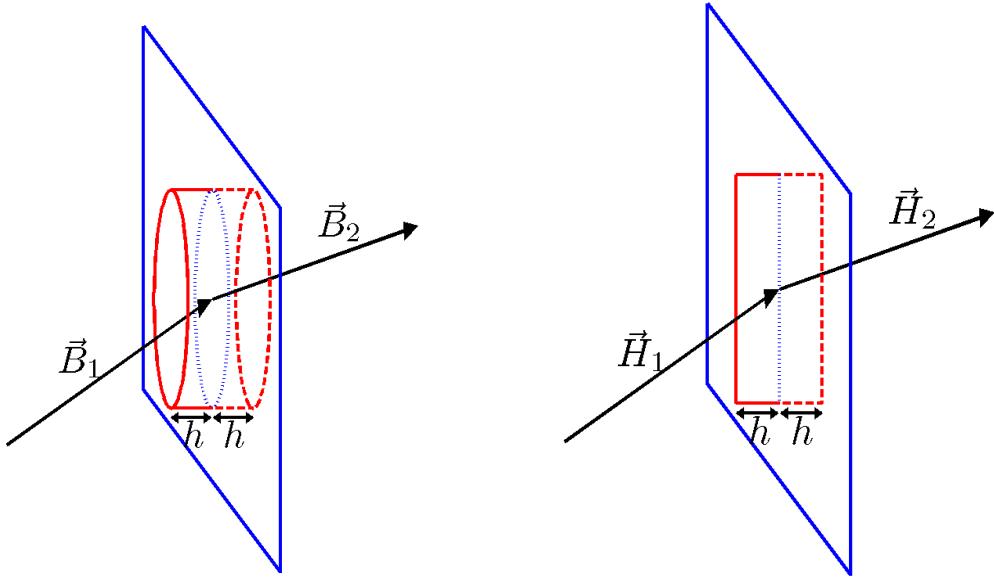


Fig. 3: (a) Left: ‘pill box’ surface for derivation of the boundary conditions on the normal component of the magnetic flux density at the interface between two media. (b) Right: geometry for derivation of the boundary conditions on the tangential component of the magnetic intensity at the interface between two media.

In other words, the normal component of the magnetic flux density is continuous across a boundary.

A second boundary condition, this time on the component of the magnetic field parallel to a boundary, can be obtained by applying Stokes’s theorem to Maxwell’s equation (3). In particular, we consider a surface S bounded by a loop ∂S that crosses the boundary of the material, see Fig. 3 (b). If we integrate both sides of Eq. (3) over that surface, and apply Stokes’s theorem (7), we find

$$\int_S \operatorname{curl} \vec{H} \cdot d\vec{S} = \int_{\partial S} \vec{H} \cdot d\vec{l} = I + \frac{\partial}{\partial t} \int_S \vec{D} \cdot d\vec{S},$$

where I is the total current flowing through the surface S . Now, let the surface S take the form of a thin strip, with the short ends perpendicular to the boundary, and the long ends parallel to the boundary. In the limit that the length of the short ends goes to zero, the area of S goes to zero: both the current flowing through the surface S , and the electric displacement integrated over S become zero. However, there are still contributions to the integral of \vec{H} around ∂S from the long sides of the strip. Thus we find that

$$H_{1\parallel} = H_{2\parallel}, \quad (13)$$

where $H_{1\parallel}$ is the component of the magnetic intensity parallel to the boundary at a point on one side of the boundary, and $H_{2\parallel}$ is the component of the magnetic intensity parallel to the boundary at a nearby point on the other side of the boundary. In other words, the *tangential* component of the magnetic intensity \vec{H} is continuous across a boundary.

We can derive a stronger constraint on the magnetic field at a boundary in the case where the material on one side of the boundary has infinite permeability (which can provide a reasonable model for some ferromagnetic materials). Since $\vec{B} = \mu \vec{H}$, it follows from (13) that

$$\frac{B_{1\parallel}}{\mu_1} = \frac{B_{2\parallel}}{\mu_2},$$

and in the limit $\mu_2 \rightarrow \infty$, while μ_1 remains finite, we must have

$$B_{1\parallel} = 0. \quad (14)$$

In other words, the magnetic flux density at the surface of a material of infinite permeability must be perpendicular to that surface. Of course, the permeability of a material characterizes its response to an applied external magnetic field: in the case where the permeability is infinite, a material placed in an external magnetic field acquires a magnetization that exactly cancels any component of the external field at the surface of the material.

3 Two-dimensional multipole fields

Consider a region of space free of charges and currents; for example, the interior of an accelerator vacuum chamber (at least, in an ideal case, and when the beam is not present). If we further exclude propagating electromagnetic waves, then any magnetic field generated by steady currents outside the vacuum chamber must satisfy

$$\operatorname{div} \vec{B} = 0, \quad (15)$$

$$\operatorname{curl} \vec{B} = 0. \quad (16)$$

Equation (15) is just Maxwell's equation (2), and Eq. (16) follows from Maxwell's equation (3) given that $\vec{J} = 0$, $\vec{B} = \mu_0 \vec{H}$, and derivatives with respect to time vanish.

We shall show that a magnetic field $\vec{B} = (B_x, B_y, B_z)$ with B_z constant, and B_x, B_y given by

$$B_y + iB_x = C_n(x + iy)^{n-1} \quad (17)$$

where $i = \sqrt{-1}$ and C_n is a (complex) constant, satisfies Eqs. (15) and (16). Note that the field components B_x and B_y are real, and are obtained from the imaginary and real parts of the right-hand side of Eq. (17). To show that the above field satisfies Eqs. (15) and (16), we apply the differential operator

$$\frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \quad (18)$$

to each side of Eq. (17). Applied to the left-hand side, we find

$$\left(\frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right) (B_y + iB_x) = \left(\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y} \right) + i \left(\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} \right). \quad (19)$$

Applied to the right-hand side of Eq. (17), the differential operator (18) gives

$$\left(\frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \right) C_n(x + iy)^{n-1} = C_n(n-1)(x + iy)^{n-2} + i^2 C_n(n-1)(x + iy)^{n-2} = 0. \quad (20)$$

Combining Eqs. (17), (19) and (20), we find

$$\begin{aligned} \frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} &= 0, \\ \frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y} &= 0. \end{aligned}$$

Finally, we note that B_z is constant, so any derivatives of B_z vanish; furthermore, B_x and B_y are independent of z , so any derivatives of these coordinates with respect to z vanish. Thus we conclude that for the field (17)

$$\operatorname{div} \vec{B} = 0, \quad (21)$$

$$\operatorname{curl} \vec{B} = 0, \quad (22)$$

and that this field is therefore a solution to Maxwell's equations within the vacuum chamber. Of course, this analysis tells us only that the field is a *possible* physical field: it does not tell us how to generate such a field. The problem of generating a field of the form Eq. (17) we shall consider in Section 4.

Fields of the form (17) are known as *multipole fields*. The index n (an integer) indicates the *order* of the multipole: $n = 1$ is a dipole field, $n = 2$ is a quadrupole field, $n = 3$ is a sextupole field, and so on. A solenoid field has $C_n = 0$ for all n , and B_z non-zero; usually, a solenoid field is not considered a multipole field, and we assume (unless stated otherwise) that $B_z = 0$ in a multipole magnet. Note that we can apply the principle of superposition to deduce that a more general magnetic field can be constructed by adding together a set of multipole fields:

$$B_y + iB_x = \sum_{n=1}^{\infty} C_n (x + iy)^{n-1}. \quad (23)$$

A ‘pure’ multipole field of order n has $C_n \neq 0$ for only that one value of n .

The coefficients C_n in Eq. (23) characterize the strength and orientation of each multipole component in a two-dimensional magnetic field. It is sometimes more convenient to express the field using polar coordinates, rather than Cartesian coordinates. Writing $x = r \cos \theta$ and $y = r \sin \theta$, we see that Eq. (23) becomes

$$B_y + iB_x = \sum_{n=1}^{\infty} C_n r^{n-1} e^{i(n-1)\theta}.$$

By writing the multipole expansion in this form, we see immediately that the strength of the field in a pure multipole of order n varies as r^{n-1} with distance from the magnetic axis. We can go a stage further, and express the field in terms of polar components:

$$B_y + iB_x = B_r \sin \theta + B_\theta \cos \theta + iB_r \cos \theta - iB_\theta \sin \theta = (B_\theta + iB_r) e^{-i\theta},$$

thus:

$$B_\theta + iB_r = \sum_{n=1}^{\infty} C_n r^{n-1} e^{in\theta}. \quad (24)$$

By writing the field in this form, we see that for a pure multipole of order n , rotation of the magnet through π/n around the z axis simply changes the sign of the field. We also see that if we write

$$C_n = |C_n| e^{in\phi_n} \quad \text{[Note]}$$

then the value of ϕ_n (the phase of C_n) determines the orientation of the field. Conventionally, a pure multipole with $\phi_n = 0$ is known as a ‘normal’ multipole, while a pure multipole with $\phi_n = \pi/2$ is known as a ‘skew’ multipole (Fig. 4).

The units of C_n depend on the order of the multipole. In SI units, for a dipole, the units of C_1 are tesla (T); for a quadrupole, the units of C_2 are Tm⁻¹; for a sextupole, the units of C_3 are Tm⁻², and so on. It is sometimes preferred to specify multipole components in dimensionless units. In that case, we introduce a reference field, B_{ref} , and a reference radius, R_{ref} . The multipole expansion is then written

$$B_y + iB_x = B_{\text{ref}} \sum_{n=1}^{\infty} (a_n + ib_n) \left(\frac{x + iy}{R_{\text{ref}}} \right)^{n-1}. \quad (25)$$

This is a standard notation for multipole fields, see, for example, Ref. [7]. In polar coordinates

$$B_y + iB_x = B_{\text{ref}} \sum_{n=1}^{\infty} (a_n + ib_n) \left(\frac{r}{R_{\text{ref}}} \right)^{n-1} e^{i(n-1)\theta}. \quad (26)$$

The reference field and reference radius can be chosen arbitrarily, but must be specified if the coefficients a_n and b_n are to be interpreted fully.

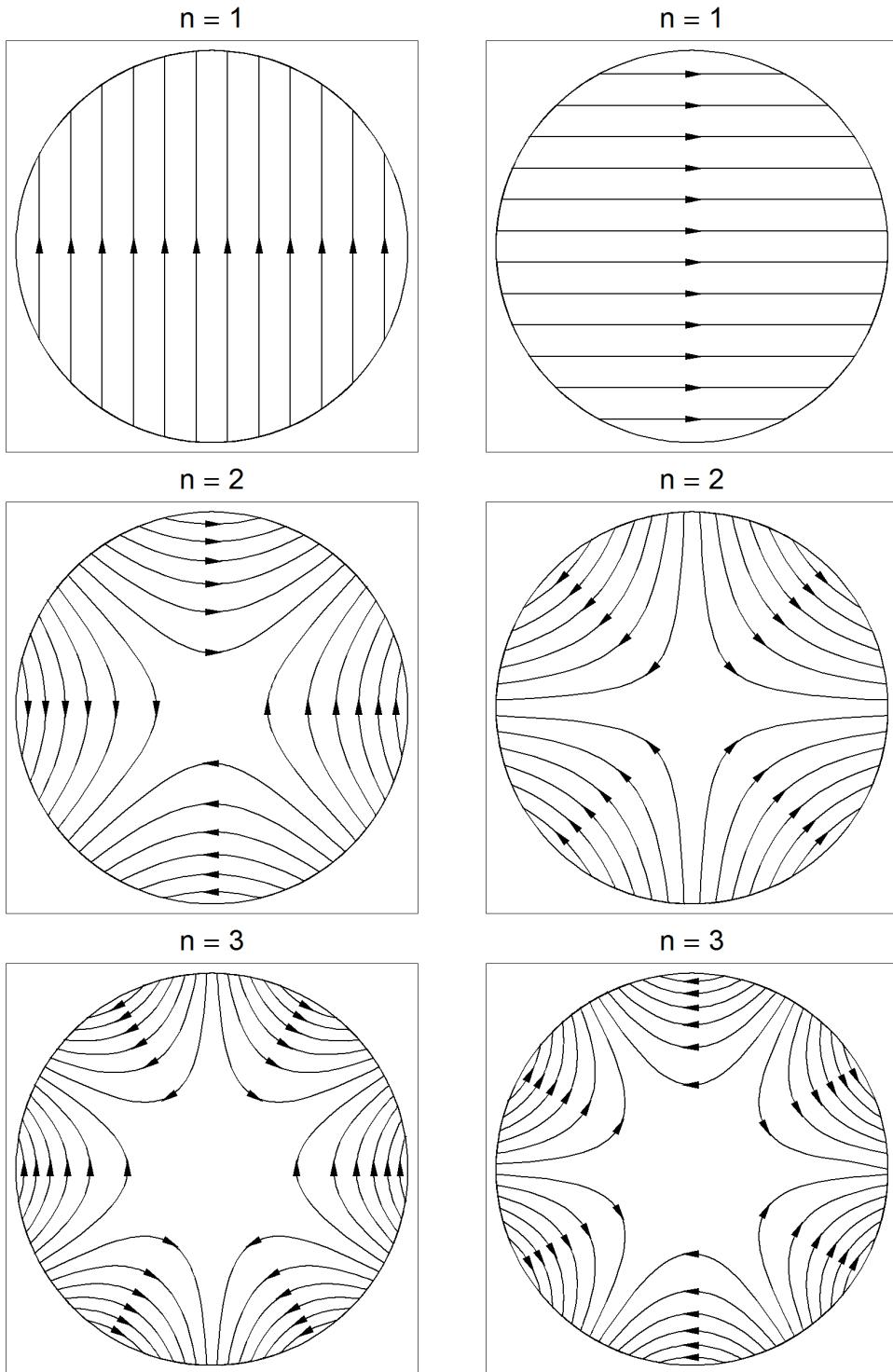


Fig. 4: ‘Pure’ multipole fields. Top: dipole. Middle: quadrupole. Bottom: sextupole. Fields on the left are normal (a_n positive); those on the right are skew (b_n positive). The positive y axis is vertically up; the positive x axis is horizontal and to the right.

Note that for a pure multipole field of order n , the coefficients a_n and b_n are related to the derivates of the field components with respect to the x and y coordinates. Thus, for a normal multipole:

$$\frac{\partial^{n-1} B_y}{\partial x^{n-1}} = (n-1)! \frac{B_{\text{ref}}}{R_{\text{ref}}^{n-1}} a_n,$$

and for a skew multipole:

$$\frac{\partial^{n-1} B_x}{\partial x^{n-1}} = (n-1)! \frac{B_{\text{ref}}}{R_{\text{ref}}^{n-1}} b_n.$$

A normal dipole has a uniform vertical field; a normal quadrupole has a vertical field for $y = 0$, that increases linearly with x ; a normal sextupole has a vertical field for $y = 0$ that increases as the square of x ; and so on.

4 Generating multipole fields

Given a system of electric charges and currents, we can integrate Maxwell's equations to find the electric and magnetic fields generated by those charges and currents. In general, the integration must be done numerically; but for simple systems it is possible to find analytical solutions. We considered two such cases in Section 2: the electric field around an isolated point charge, and the magnetic field around a long straight wire carrying a constant current.

It turns out that we can combine the magnetic fields from long, straight, parallel wires to generate pure multipole fields. It is also possible to generate pure multipole fields using high-permeability materials with the appropriate geometry. We consider both methods in this section. For the moment, we deal with ‘idealized’ geometries without practical constraints. We discuss the impact of some of the practical limitations in later sections.

4.1 Current distribution for a multipole field

Our goal is to determine a current distribution that will generate a pure multipole field of specified order. As a first step, we derive the multipole components in the field around a long straight wire carrying a uniform current. We already know from Ampère's law (9) that the field at distance r from a long straight wire carrying current I in free space has magnitude given by

$$B = \frac{I}{2\pi\mu_0 r},$$

and that the direction of the field describes a circle centred on the wire. To derive the multipole components in the field, we first derive an expression for the field components at an arbitrary point (x, y) from a wire carrying current I , passing through a point (x_0, y_0) and parallel to the z axis.

Since we are working in two dimensions, we can represent the components of a vector by the real and imaginary parts of a complex number. Thus, the vector from (x_0, y_0) to a point (x, y) is given by $re^{i\theta} - r_0 e^{i\theta_0}$, and the magnitude of the field at (x, y) is

$$B = \frac{I}{2\pi\mu_0} \frac{1}{|re^{i\theta} - r_0 e^{i\theta_0}|}.$$

The geometry is shown in Fig. 5. The direction of the field is perpendicular to the line from (x_0, y_0) to (x, y) . Since a rotation through 90° can be represented by a multiplication by i , we can write

$$B_x + iB_y = \frac{I}{2\pi\mu_0} \frac{i(re^{i\theta} - r_0 e^{i\theta_0})}{|re^{i\theta} - r_0 e^{i\theta_0}|^2},$$

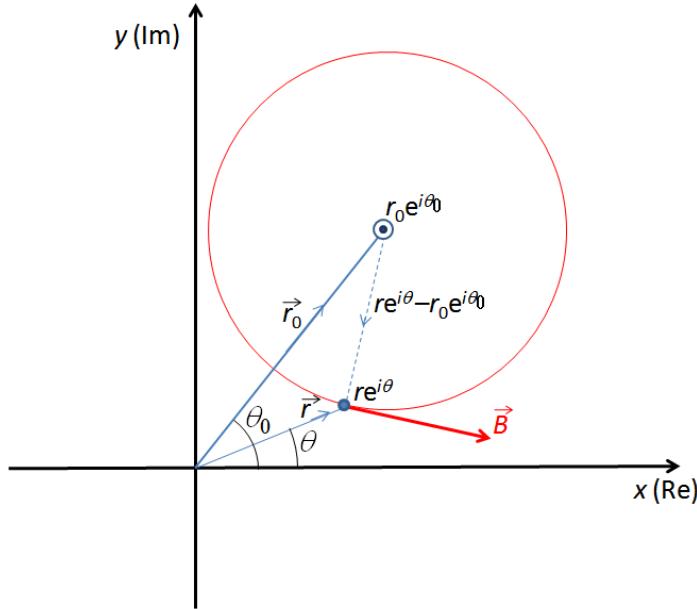


Fig. 5: Geometry for calculation of multipole components in the field around a long, straight wire carrying a uniform current. The wire passes through $r_0 e^{i\theta_0}$, and is parallel to the z axis (the direction of the current is pointing out of the page).

and hence

$$B_y + iB_x = \frac{I}{2\pi\mu_0} \frac{(re^{-i\theta} - r_0 e^{-i\theta_0})}{|re^{i\theta} - r_0 e^{i\theta_0}|^2} = \frac{I}{2\pi\mu_0} \frac{1}{re^{i\theta} - r_0 e^{i\theta_0}}.$$

Now, we write the magnetic field as

$$B_y + iB_x = -\frac{I}{2\pi\mu_0 r_0} \frac{e^{-i\theta_0}}{1 - \frac{r}{r_0} e^{i(\theta-\theta_0)}},$$

and use the Taylor series expansion for $(1 - \zeta)^{-1}$, where ζ is a complex number with $|\zeta| < 1$:

$$\frac{1}{1 - \zeta} = \sum_{n=0}^{\infty} \zeta^n,$$

to write

$$B_y + iB_x = -\frac{I}{2\pi\mu_0 r_0} e^{-i\theta_0} \sum_{n=1}^{\infty} \left(\frac{r}{r_0} \right)^{n-1} e^{i(n-1)(\theta-\theta_0)}. \quad (27)$$

Equation (27) is valid for $r < r_0$. Comparing with the standard multipole expansion, Eq. (26), we see that if we choose for the reference field B_{ref} and the reference radius R_{ref}

$$\begin{aligned} B_{\text{ref}} &= \frac{I}{2\pi\mu_0 r_0}, \\ R_{\text{ref}} &= r_0, \end{aligned}$$

then the coefficients for the multipole components in the field are given by

$$b_n + ia_n = -e^{-in\theta_0}.$$

The field around a long straight wire can be represented as an infinite sum over all multipoles.

Now we consider a current flowing on the surface of a cylinder of radius r_0 . Suppose that the current flowing in a section of the cylinder at angle θ_0 and subtending angle $d\theta_0$ at the origin is $I(\theta_0) d\theta_0$. By the principle of superposition, we can obtain the total field by summing the contributions from the currents at all values of θ_0 :

$$B_y + iB_x = -\frac{1}{2\pi\mu_0 r_0} \sum_{n=1}^{\infty} \left(\frac{r}{r_0}\right)^{n-1} e^{i(n-1)\theta} \int_0^{2\pi} e^{-in\theta_0} I(\theta_0) d\theta_0. \quad (28)$$

We see that the multipole components are related to the Fourier components in the current distribution over the cylinder of radius r_0 . In particular, if we consider a current distribution with just a single Fourier component

$$I(\theta_0) = I_0 \cos(n_0\theta_0 - \phi), \quad (29)$$

the integral in the right-hand side of Eq. (28) vanishes except for $n = n_0$, and we find

$$B_y + iB_x = -\frac{I_0}{2\pi\mu_0 r_0} \left(\frac{r}{r_0}\right)^{n_0-1} e^{i(n_0-1)\theta} \pi e^{-i\phi}.$$

The current distribution (29) generates a pure multipole field of order n_0 . If we choose, as before

$$\begin{aligned} B_{\text{ref}} &= \frac{I_0}{2\pi\mu_0 r_0}, \\ R_{\text{ref}} &= r_0, \end{aligned}$$

then the multipole coefficients are

$$b_n + ia_n = -\pi e^{-i\phi}.$$

The parameter ϕ gives the ‘angle’ of the current distribution. For $\phi = 0$ or $\phi = \pi$, the current generates a normal multipole; for $\phi = \pm\pi/2$, the current generates a skew multipole (Fig 6).

The fact that a sinusoidal current distribution on a cylinder can generate a pure multipole field is not simply of academic interest. By winding wires in an appropriate pattern on a cylinder, it is possible to approximate a sinusoidal current distribution closely enough to produce a multipole field of acceptable quality for many applications. Usually, several layers of windings are used with a different pattern of wires in each layer, to improve the approximation to a sinusoidal current distribution. Superconducting wires can be used to achieve strong fields: an example of superconducting quadrupoles in the LHC is shown in Fig. 7.

4.2 Geometry of an iron-dominated multipole magnet

Normal-conducting magnets usually use iron cores to increase the flux density achieved by a given current. In such a magnet, the shape of the magnetic field depends mainly on the geometry of the iron. In this section, we shall derive the geometry required to generate a pure multipole of given order. To simplify the problem, we will make some approximations: in particular, we shall assume that the iron core has uniform cross-section and infinite extent along z ; that there are no limits to the iron in x or y ; and that the iron has infinite permeability. The field in a more realistic magnet will generally need to be calculated numerically; however, the characteristics derived from our idealized model are often a good starting point for the design of an iron-dominated multipole magnet.

We base our analysis on the magnetic scalar potential, φ , which is related to the magnetic field \vec{B} by

$$\vec{B} = -\text{grad } \varphi. \quad (30)$$

Note that the curl of the field in this case is zero, for any function φ : this is a consequence of the mathematical properties of the grad and curl operators. Therefore, it follows from Maxwell’s equation (3)

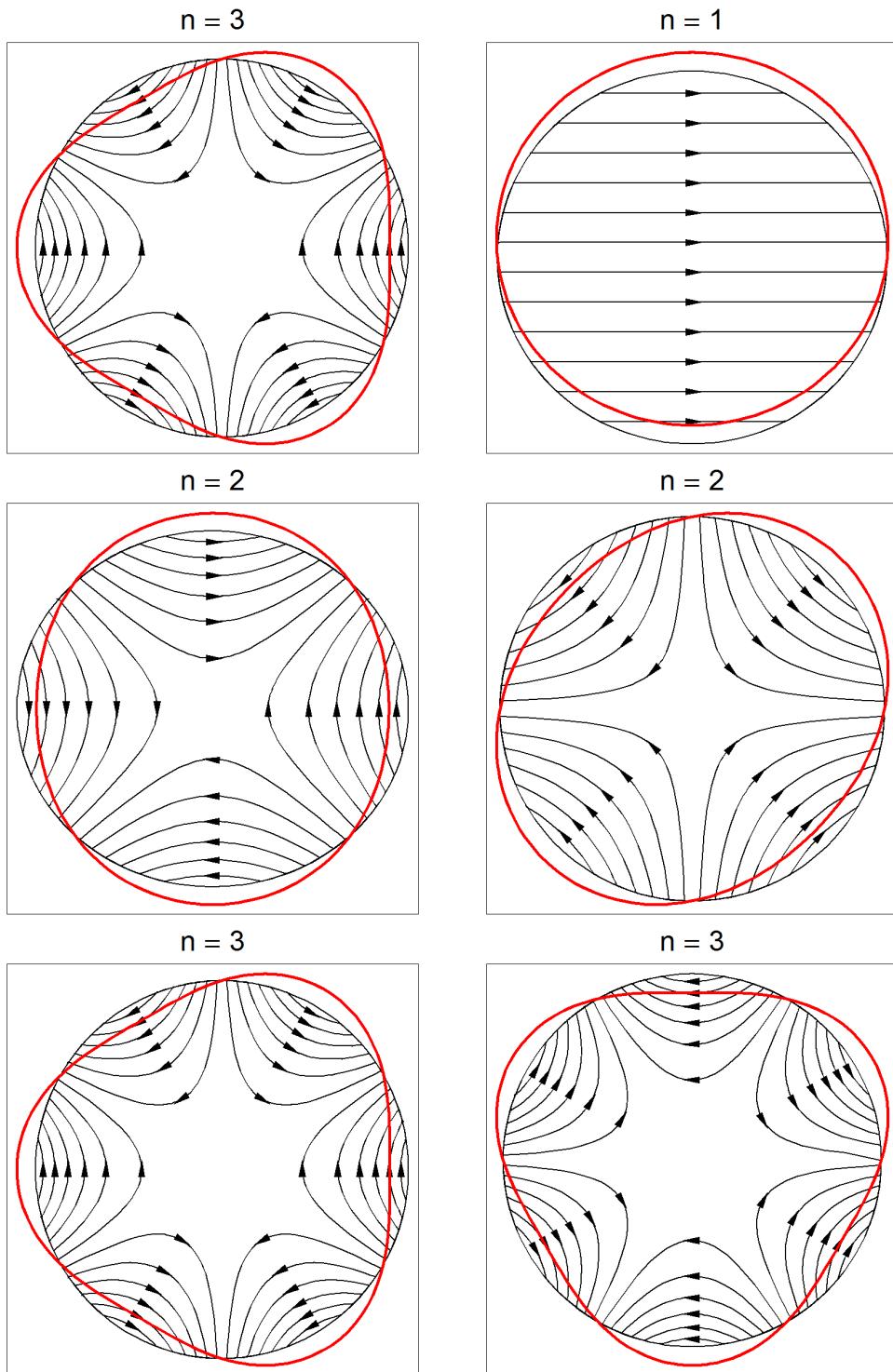


Fig. 6: Current distributions for generating pure multipole fields. Top: dipole. Middle: quadrupole. Bottom: sextupole. Fields on the left are normal (a_n positive); those on the right are skew (b_n positive). The positive y axis is vertically up; the positive x axis is horizontal and to the right. The deviation of the red line from the circular boundary shows the local current density. Current is flowing in the positive z direction (out of the page) for increased radius, and in the negative z direction for reduced radius.

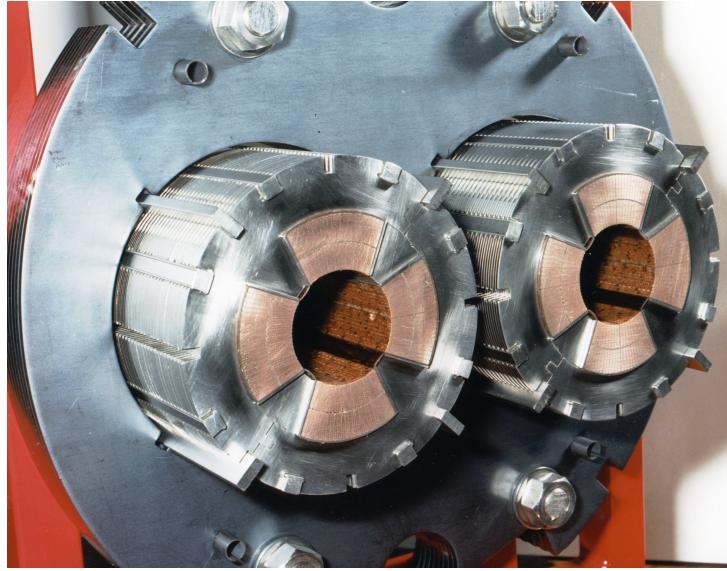


Fig. 7: Superconducting quadrupoles in the LHC

that a magnetic field can only be derived from a scalar potential if: (i) there is no current density at the location where the field is to be calculated; (ii) there is no time-dependent electric displacement at the location where the field is to be calculated. Where there exists an electric current or a time-dependent electric field, it is more appropriate to use a vector potential (in which case, the magnetic flux density is found from the curl of the vector potential). However, for multipole fields, we have already shown that both the divergence and the curl of the field vanish, Eqs. (21) and (22). Since the curl of the grad of any function is identically zero, Eq. (22) is automatically satisfied for any field \vec{B} derived using (30). From Eq. (21), we find

$$\nabla^2 \varphi = 0, \quad (31)$$

where ∇^2 is the Laplacian operator. Equation (31) is Poisson's equation: the scalar potential in a particular case is found by solving this equation with given boundary conditions.

To determine the geometry of iron required to generate a pure multipole field, we shall start by writing down the scalar potential for a pure multipole field. Since the magnetic flux density \vec{B} is obtained from the gradient of the scalar potential, the flux density at any point must be perpendicular to a surface of constant scalar potential. However, we already know from Eq. (14) that the magnetic flux density at the surface of a material with infinite permeability must be perpendicular to that surface. Hence to generate a pure multipole field in a magnet containing material of infinite permeability, we just need to shape the material so that its surface follows a surface of constant magnetic scalar potential for the required field.

We therefore look for a potential φ that satisfies

$$-\left(\frac{\partial}{\partial y} + i\frac{\partial}{\partial x}\right)\varphi = B_y + iB_x = C_n(x + iy)^{n-1}.$$

As we shall now show, an appropriate solution is

$$\varphi = -|C_n| \frac{r^n}{n} \sin(n\theta - \phi_n) \quad (32)$$

where

$$x + iy = re^{i\theta},$$

and so

$$\begin{aligned} x &= r \cos \theta, \\ y &= r \sin \theta. \end{aligned}$$

That Eq. (32) is indeed the potential for a pure multipole of order n can be shown as follows. In polar coordinates, the gradient can be written

$$\text{grad } \varphi = \hat{r} \frac{\partial \varphi}{\partial r} + \frac{\hat{\theta}}{r} \frac{\partial \varphi}{\partial \theta}, \quad (33)$$

where \hat{r} and $\hat{\theta}$ are unit vectors in the directions of increasing r and θ , respectively. Using

$$\begin{aligned} \hat{r} &= \hat{x} \cos \theta + \hat{y} \sin \theta, \\ \hat{\theta} &= -\hat{x} \sin \theta + \hat{y} \cos \theta, \end{aligned}$$

it follows from Eq. (33) that

$$\begin{aligned} -\text{grad } \varphi &= (\hat{x} \cos \theta + \hat{y} \sin \theta) |C_n| r^{n-1} \sin(n\theta - \phi_n) - (\hat{x} \sin \theta - \hat{y} \cos \theta) |C_n| r^{n-1} \cos(n\theta - \phi_n), \\ &= \hat{x} \sin((n-1)\theta - \phi_n) |C_n| r^{n-1} + \hat{y} \cos((n-1)\theta - \phi_n) |C_n| r^{n-1}. \end{aligned}$$

Thus the field derived from the potential (32) can be written

$$B_y + iB_x = |C_n| e^{-i\phi_n} r^{n-1} e^{i(n-1)\theta}.$$

Therefore, if

$$C_n = |C_n| e^{-i\phi_n},$$

then

$$B_y + iB_x = C_n r^{n-1} e^{i(n-1)\theta} = C_n (x + iy)^{n-1},$$

and we see that the potential (32) does indeed generate a pure multipole field of order n .

From the above argument, we can immediately conclude that to generate a pure multipole field, we can shape a high permeability material such that the surface of the material follows the curve given (in parametric form, with parameter θ) by

$$r^n \sin(n\theta - \phi_n) = r_0^n, \quad (34)$$

where r_0 is a constant giving the minimum distance between the surface of the material and the origin. The cross-sections of iron-dominated multipole magnets of orders 1, 2 and 3 are shown in Fig. 8. Note that $r \rightarrow \infty$ for $n\theta - \phi_n \rightarrow \text{integer} \times \pi$. Treating each region between infinite values of r as a separate pole, we see that a pure multipole of order n has $2n$ poles. We also see that the potential changes sign when moving from one pole to either adjacent pole: that is, poles alternate between ‘north’ and ‘south’. The field must be generated by currents flowing along wires between the poles, parallel to the z axis: to avoid direct contribution from the field around the wires, these wires should be located a large (in fact, infinite) distance from the origin.

Note that it is possible to determine the shape of the pole face for a magnet containing any specified set of multipoles by summing the potentials for the different multipole components, and then solving for r as a function of θ , for a fixed value of the scalar potential. Magnets designed to have more than one multipole component are often known as ‘combined function’ magnets. Perhaps the most common type of combined function magnet is a dipole with a quadrupole component: such magnets can be used to steer and focus a beam simultaneously. The shape of the pole faces and the field lines in a dipole with (strong) quadrupole component is shown in Fig. 9.

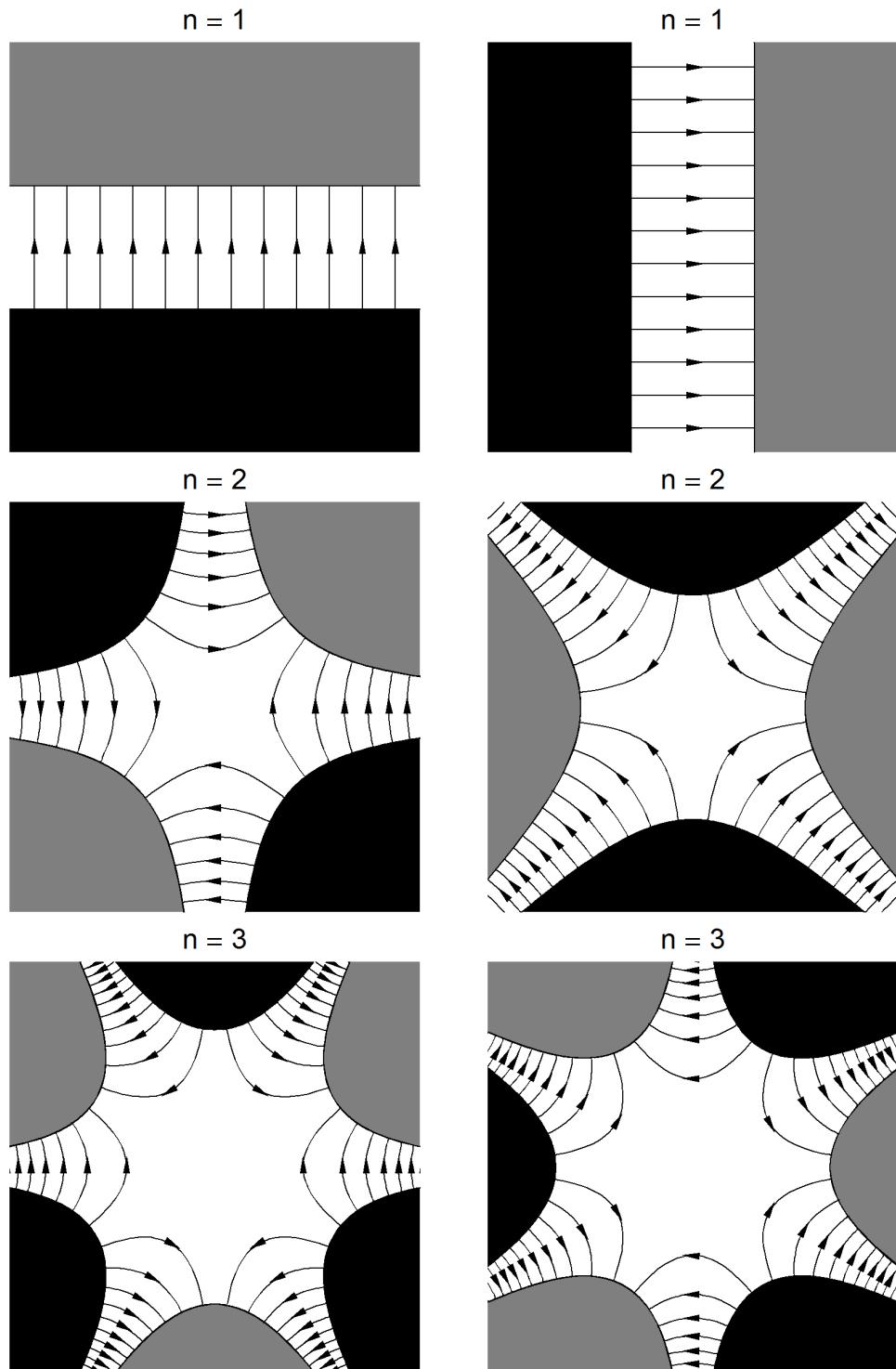


Fig. 8: Pole shapes for generating pure multipole fields. Top: dipole. Middle: quadrupole. Bottom: sextupole. Fields on the left are normal (a_n positive); those on the right are skew (b_n positive). The positive y axis is vertically up; the positive x axis is horizontal and to the right. The poles, shown as black (north) or grey (south), are constructed from material with infinite permeability.

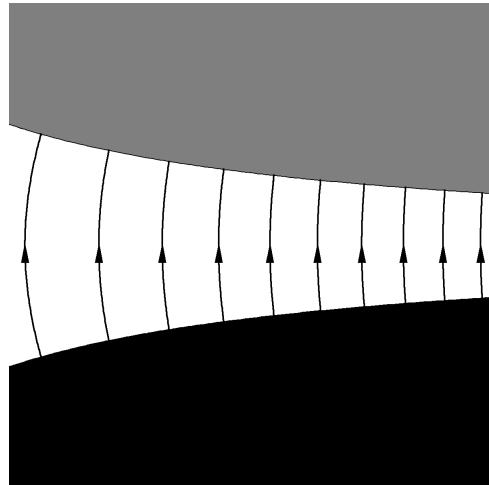


Fig. 9: Pole shapes for dipole magnet with additional quadrupole component

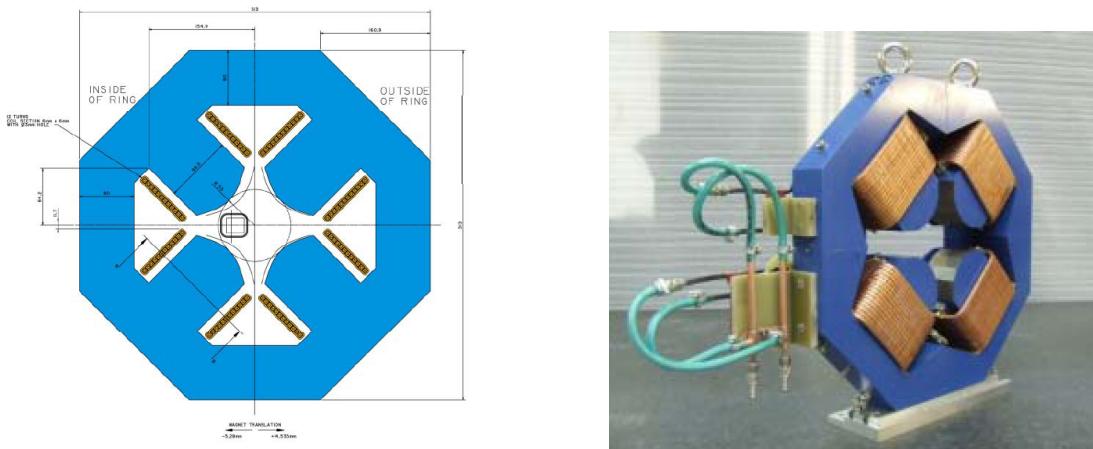


Fig. 10: Iron-dominated quadrupole magnet for the EMMA Fixed-Field Alternating Gradient accelerator at Daresbury Laboratory. Left: magnet cross-section [8]. Right: magnet prototype [9].

In practice, some variation from the ‘ideal’ geometry is needed to account for the fact that the material used in the magnet has finite permeability, and finite extent transversely and longitudinally. The wires carrying the current that generates the magnetic flux are arranged in coils around each pole; as we shall see, the strength of the field is determined by the number of ampere-turns in each coil. An iron-dominated electromagnetic quadrupole is shown in Fig. 10.

To complete our discussion of methods to generate multipole fields, we derive an expression for the field strength in an iron-dominated magnet with a given number of ampere-turns in the coil around each pole. To do this, we consider a line integral as shown in Fig. 11. In the figure, we show a quadrupole; however the generalization to other orders of multipole is straightforward. Note that, in principle, the coils carrying the electric current, and the line segment C_3 , are an infinite distance from the origin (the centre of the magnet).

Using Maxwell’s equation (3), with constant (zero) electric displacement, and integrating over the surface S bounded by the curve $C_1 + C_2 + C_3$ gives

$$\int_S \operatorname{curl} \vec{H} \cdot d\vec{S} = \int_S \vec{J} \cdot d\vec{S} = -NI.$$

Note that the surface is oriented so that the normal is parallel to the positive z axis; and the coil around

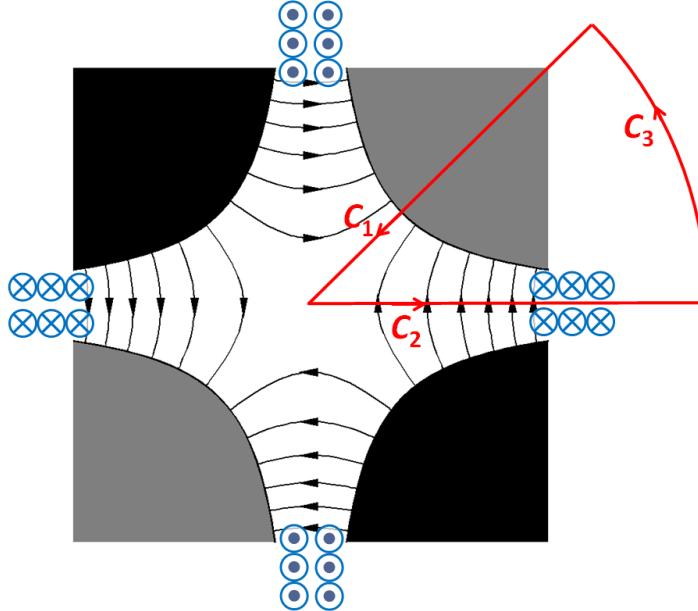


Fig. 11: Contour for line integral used to calculate the field strength in an iron-dominated quadrupole

each pole consists of N turns of wire carrying current I . Applying Stokes's theorem (7) gives

$$\int_{C_1} \vec{H} \cdot d\vec{l} + \int_{C_2} \vec{H} \cdot d\vec{l} + \int_{C_3} \vec{H} \cdot d\vec{l} = -NI.$$

We know, from Eq. (12), that the normal component of the magnetic flux density \vec{B} is continuous across a boundary. Then, since $\vec{B} = \mu \vec{H}$, it follows that for a finite field between the poles, and for $\mu \rightarrow \infty$, the magnetic intensity \vec{H} vanishes within the poles. Also, the field is perpendicular to the line segment C_2 . Thus, the only part of the integral that makes a non-zero contribution, is the integral along C_1 from the face of the pole to the origin. Hence

$$\int_0^{r_0} \frac{B_r(r)}{\mu_0} dr = NI. \quad (35)$$

The contour C_1 is chosen so that along this contour the field has only a radial component, parallel to the contour. From Eq. (24), we see that for a multipole of order n , along this contour we have

$$B_r = |C_n| r^{n-1}.$$

Thus we find by performing the integral in Eq. (35)

$$|C_n| = \mu_0 N I \frac{n}{r_0^n}.$$

For a normal multipole, the field is given by

$$B_y + iB_x = \frac{\mu_0 n N I}{r_0} \left(\frac{x + iy}{r_0} \right)^{n-1}.$$

For example, in a normal quadrupole ($n = 2$), the field gradient is given by

$$\frac{\partial B_y}{\partial x} = \frac{2\mu_0 N I}{r_0^2}. \quad (36)$$

5 Multipole decomposition

In the previous section, we derived the current density distributions and material geometries needed to generate a pure multipole field of a given order. However, the distributions and geometries required are not perfectly achievable in practice: the currents and materials have infinite longitudinal extent; and we require either a current that exists purely on the surface of a cylinder, or infinite permeability materials with infinite transverse extent.

Real multipole magnets, therefore, will not consist of a single multipole component, but a superposition of (in general) an infinite number of multipole fields. The exact shape of the field can have a significant impact on the beam dynamics in an accelerator. In many simulation codes for accelerator beam dynamics, the magnets are specified by the multipole coefficients: this is because simple techniques exist for approximating the effect, for example, of sextupole, octopole, and other higher-order components in the field of a quadrupole magnet. The question then arises how to determine the multipole components in a given magnetic field.

At this point, we can make a distinction between the *design* field of a magnet, and the field that exists within a fabricated magnet. The design field is one that is still in some sense ‘ideal’; though the design field for a quadrupole magnet (for example) will contain other multipole components, because the design has to respect practical constraints, i.e., the magnet will have finite longitudinal and transverse extent, any currents will flow in wires of non-zero dimension, and any materials present will have finite (and often non-linear) permeability. Usually, one attempts to optimize the design to minimize the strengths of the multipole components apart from the one required: the residual strengths are generally known as *systematic* multipole errors. These errors will be present in any fabricated magnet, although, because of construction tolerances, the errors will vary between any two magnets of the same nominal design. The differences between the multipole components in the design field and the components in a particular magnet are known as *random* multipole errors.

The effects of systematic and random multipole errors on an accelerator, and hence the specification of upper limits on these quantities, can usually only be properly understood by running beam dynamics simulations. Therefore, accelerator magnet (and lattice) design often proceeds iteratively. Some initial estimate of the limits on the errors is often needed to guide the magnet design; but then any design that is developed must be studied by further beam dynamics simulations to determine whether improvements are needed.

It is therefore important to be able to determine the multipole components in a magnetic field from numerical field data: these data may come from either a magnetic model (i.e., from the design of a magnet), or from measurements on a real device. There are different procedures that can be used to achieve the ‘decomposition’ of a field into its multipole components. In this section, we shall consider methods based on Cartesian and polar representations of two-dimensional fields (i.e., fields that are independent of the longitudinal coordinate). In Section 6 we shall consider decompositions of three-dimensional fields (i.e., fields that have explicit dependence on longitudinal as well as transverse coordinates). However, we first consider an important concept in the discussion of multipole field errors, namely how the symmetry of a multipole magnet leads to ‘allowed’ and ‘forbidden’ higher-order multipoles.

5.1 Multipole symmetry, ‘allowed’ and ‘forbidden’ higher-order multipoles

A pure multipole field of order n can be written



$$B_y + iB_x = |C_n| e^{-i\phi_n} r^{n-1} e^{i(n-1)\theta}. \quad (37)$$

The parameter ϕ_n characterizes the angular orientation of the magnet around the z axis. In particular, from Eq. (34), we see that a change in ϕ_n by $n\alpha$ is equivalent to a rotation of the coordinates (a change in θ) by $-\alpha$. Thus, a rotation of a magnet around the z axis by angle α may be represented by a change in ϕ_n by $n\alpha$. In particular, if the magnet is rotated by π/n , then from Eq. (37), we see that the field at

any point simply changes sign:

$$\text{if } \phi_n \mapsto \phi_n + \pi, \text{ then } \vec{B} \mapsto -\vec{B}. \quad (38)$$

This property of the magnetic field is imposed by the symmetry of the magnet. In a real magnet, it will not be satisfied exactly, because random variations in the geometry will break the symmetry. However, it is possible to maintain the symmetry exactly in the design of the magnet; this means that although higher order multipoles will in general be present, only those multipoles satisfying the symmetry constraint (38) can be present. These are the ‘allowed’ multipoles. Other multipoles, which must be completely absent, are the ‘forbidden’ multipoles.

We can derive a simple expression for the allowed multipoles in a magnet designed with symmetry for a multipole of order n . Consider an additional multipole (a ‘systematic error’) in this field, of order m . By the principle of superposition, the total field can be written as

$$B_y + iB_x = |C_n| e^{-i\phi_n} r^{n-1} e^{i(n-1)\theta} + |C_m| e^{-i\phi_m} r^{m-1} e^{i(m-1)\theta}.$$

The geometry is such that under a rotation about the z axis through π/n , the magnet looks the same, except that all currents have reversed direction: therefore the field simply changes sign. Under this rotation $\phi_n \mapsto \phi_n + \pi$; however, $\phi_m \mapsto \phi_m + m\pi/n$. This means that we must have:

$$e^{-i\frac{m}{n}\pi} = -1.$$

Therefore m/n must be an odd integer. Assuming that $m \neq n$ (i.e., the multipole error is of a different order than the ‘main’ multipole field), then

$$\frac{m}{n} = 3, 5, 7, \dots \quad (39)$$

Thus, for a dipole, the allowed higher order multipoles are sextupole, decapole, etc.; for a quadrupole, the allowed higher order multipoles are dodecapole, 20-pole, etc. The fact that the allowed higher order multipoles have an order given by an odd integer multiplied by the order of the main multipole is a consequence of the fact that magnetic poles always occur in north–south pairs. This is illustrated for a quadrupole in Fig. 12; here we see that to maintain the correct rotational symmetry (with the field changing sign under a rotation through $\pi/2$) the first higher-order multipole must be constructed by ‘splitting’ each main pole into three, then into five, and so on.

The field in a real magnet will contain all higher order multipoles, not just the ones allowed by symmetry. However, it is often the case that the allowed multipoles dominate over the forbidden multipoles.

5.2 Fitting multipoles: Cartesian basis

Suppose we have obtained a set of numerical field data, either from a magnetic model, or from measurements on a real magnet. To determine the effect of the field on the beam dynamics in an accelerator, it is helpful to know the multipole components in the field. One way to compute the multipole components is to fit a polynomial to the field data. For example, if we consider a normal multipole (coefficients C_n are all real), the vertical field along the x axis (i.e., for $y = 0$) is given by

$$B_y = \sum_{n=1}^{\infty} C_n x^{n-1}. \quad (40)$$

The number of data points determines the highest order multipole that can be fitted. Fitting may be achieved using, for example, a routine that minimizes the squares of the residuals between the data and the fitted function. However, although this procedure can, in principle, produce good results, it is not

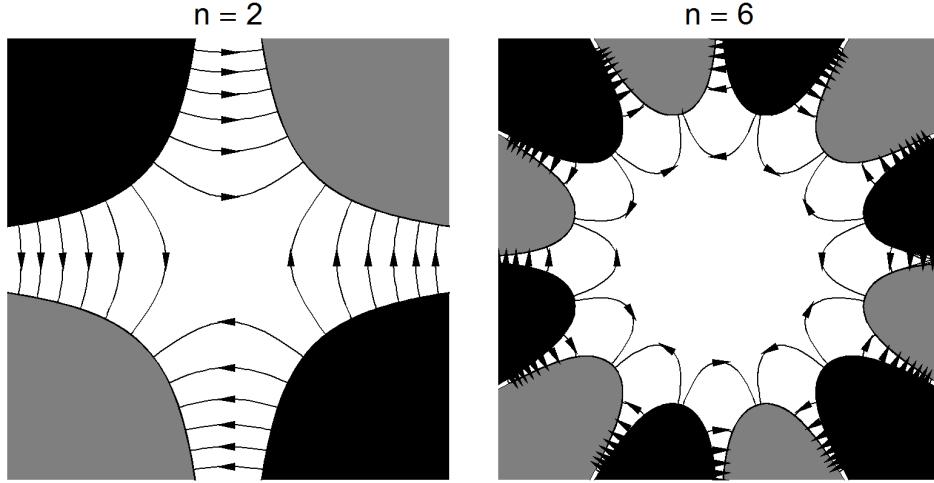


Fig. 12: Normal quadrupole field (left) and dodecapole field (right). The dodecapole is the first higher-order multipole with the same rotational symmetry as the quadrupole (under a rotation by $\pi/2$, north and south poles interchange).

very robust. In particular, the presence of multipoles of higher order than those included in the fit can affect the values determined for those multipoles that are included in the fit. We can illustrate this as follows.

Let us construct a quadrupole field ($n = 2$), and add to it higher order multipoles of order 3, 4, 5 and 6. The values of the coefficients a_n (actual values, and fitted values in two different cases) are given in Table 1. The field B_y/B_{ref} is plotted as a function of x/R_{ref} in Fig. 13: the field data are shown as points, while the fit, including multipoles up to order 6, is shown as a line. Also shown is the deviation $\Delta B_y/B_{\text{ref}}$ from an ideal quadrupole field, i.e. ΔB_y is the contribution of the higher order multipoles. We see that if we base the fit on all the multipoles that are present (i.e. up to order 6), then we obtain accurate values for all multipole coefficients.

However, in general, multipoles of all orders are present, while our fit is based on a finite number of multipoles. If we try to fit the data in our illustrative case using multipoles up to order 5 only (i.e. omitting the order 6 multipole that is present), then we see that there is an impact on the accuracy with which we determine the lower-order multipoles. This can be seen in the final column of Table 1: there is even an error in the value that we determine for the quadrupole strength. When we plot the fit against the field data, we see that there is some small residual deviation between the data and the fit: this is to be expected, since the function we are using to obtain the fit does not match exactly the function used to generate the data. Although not visible in the total field, plotted in Fig. 14, the difference between the fit and the data is apparent in the plot of the deviation from the quadrupole field.

Our concern is that the presence of higher-order multipoles has affected the accuracy with which

Table 1: Actual and fitted multipole values for a quadrupole field with artificially constructed multipole errors

n	actual coefficient a_n	fitted coefficient a_n	
		$(n \leq 6)$	$(n \leq 5)$
2	1.000	1.000	0.9972
3	0.010	0.010	0.0100
4	0.001	0.001	0.0131
5	0.010	0.010	0.0100
6	0.010	0.010	—

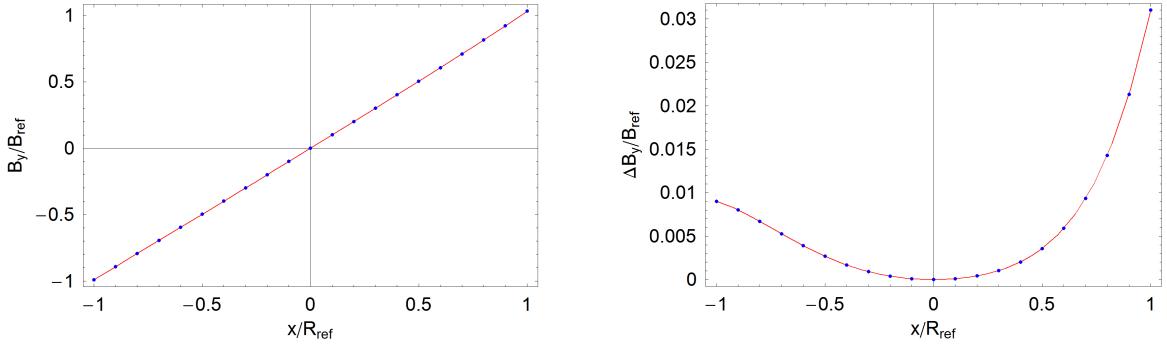


Fig. 13: Measured (points) and fitted (line) field in a quadrupole with higher-order multipole errors of order 3, 4, 5 and 6. Multipoles up to order 6 are fitted. Left: total field. Right: deviation from quadrupole field.

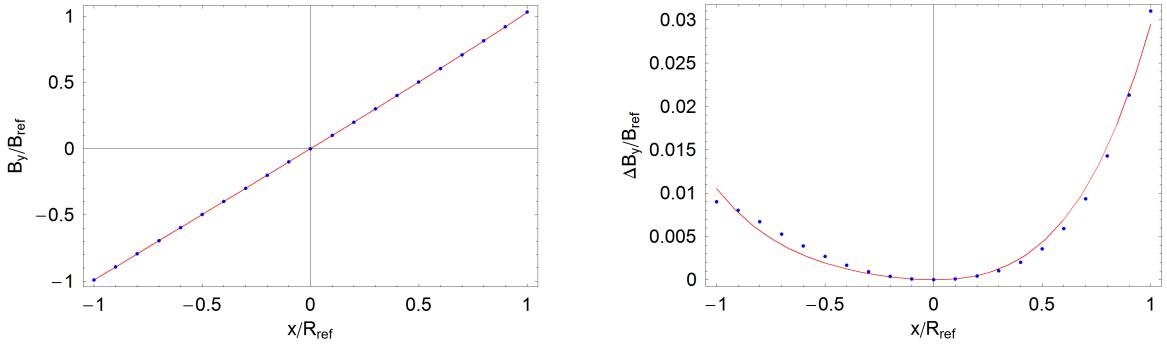


Fig. 14: Measured (points) and fitted (line) field in a quadrupole with higher-order multipole errors of order 3, 4, 5 and 6. Multipoles up to order 5 are fitted. Left: total field. Right: deviation from quadrupole field.

we determine the lower-order multipoles, even down to the quadrupole field strength. This can have significant implications for beam dynamics: the effect of a linear focusing error in a beam line (from some variation in the quadrupole strength) can be very different from the effects of higher-order multipole errors. For example, if one is measuring the betatron tunes or the beta functions in a storage ring, these can be very sensitive to linear focusing errors, and relatively insensitive to higher order multipoles. Determining the multipole coefficients using a polynomial fit can lead to inaccurate predictions of the linear behaviour of the beam line, depending on the higher-order multipoles present in the magnets.

The problem is that we have based our fit on monomials, i.e., powers of x . Our fit is a sum of these monomials, with coefficients determined from the data. However, it is possible to obtain a fit to data generated using one monomial, with a different monomial. For example, if one constructs data which is purely linear in x , then one can obtain a fit using a monomial x^3 (even though the fit will not be as good as one obtained using a monomial x). Mathematically, the basis functions we are using (monomials in this case) are not orthogonal: the coefficients we determine depend on which set of basis functions we choose to use. A more robust technique would use basis functions that are orthogonal, i.e., the coefficients we determine will be the same, no matter which set of functions we choose. Fortunately, there exists an appropriate set of functions that provides an orthogonal basis for multipole fields. We discuss this basis in the following section, 5.3. The advantage of orthogonal basis functions is that the coefficients we determine for different terms in the fit are *independent* of which terms we include in the fit; for example, the quadrupole strength that we find in a particular magnet will be the same, irrespective of which higher-order terms we include in the fit, and which higher-order terms are actually present.

5.3 Fitting multipoles: polar basis

From Eq. (24) we know that the field in a multipole magnet can be written in polar coordinates as

$$B_\theta + iB_r = \sum_{n=1}^{\infty} C_n r^{n-1} e^{in\theta}.$$

We see that if we make a set of measurements of B_r and B_θ at different values of θ and fixed radial distance r , then we can obtain the coefficients C_n by a discrete Fourier transform.

Suppose we make M measurements of the field, at $\theta = \theta_m$, where

$$\theta_m = 2\pi \frac{m}{M}, \quad m = 0, 1, 2, \dots, M-1. \quad (41)$$

We write the measurement at $\theta = \theta_m$ as B_m ; note that B_m is a complex number, whose real and imaginary parts are given by the azimuthal and radial components of the field at $\theta = \theta_m$.

Now we construct, for a chosen integer n'

$$\sum_{m=0}^{M-1} B_m e^{-2\pi i n' \frac{m}{M}} = \sum_{m=0}^{M-1} \sum_{n=1}^{\infty} C_n r_0^{n-1} e^{2\pi i (n-n') \frac{m}{M}},$$

where r_0 is the radial distance at which the field measurements are made. The summation over m on the right-hand side vanishes, unless $n = n'$. Thus we can write

$$\sum_{m=0}^{M-1} B_m e^{-2\pi i n' \frac{m}{M}} = MC_{n'} r_0^{n'-1}.$$

If we relabel n' as n , then we see that the multipole coefficients C_n are given by

$$C_n = \frac{1}{Mr_0^{n-1}} \sum_{m=0}^{M-1} B_m e^{-2\pi i n \frac{m}{M}}. \quad (42)$$

The advantage of this technique over that in Section 5.2 is that the basis functions used to construct the fit are of the form $e^{in\theta}$, for integer n . These functions are orthogonal: mathematically, this means that

$$\int_0^{2\pi} e^{in\theta} e^{-in'\theta} d\theta = 2\pi \delta_{nn'},$$

where the Kronecker delta function $\delta_{nn'} = 1$ if $n = n'$, and $\delta_{nn'} = 0$ if $n \neq n'$. The important consequence for us is that the value we determine for any given multipole using Eq. (42) is independent of the presence of any other multipoles, of higher or lower order.

A further advantage of using the polar basis instead of the Cartesian basis comes from the dependence of the field on the radial distance. Suppose that the field data are measured (or obtained from a model) with accuracy ΔB_m . Then the accuracy in the multipole coefficients will be

$$\Delta C_n \approx \frac{\Delta B_m}{r_0^{n-1}}.$$

We obtain better accuracy in the multipole coefficients if we choose the radius r_0 , on which the measurements are made, to be as large as possible. Furthermore, the accuracy in the fitted field will be

$$\Delta B \approx \Delta C_n \left(\frac{r}{r_0} \right)^{n-1}.$$

We obtain *improved* accuracy in the field for $r < r_0$; but the accuracy reduces quickly (particularly for higher-order multipoles) for $r > r_0$. It is important to choose the radial distance r_0 large enough to enclose all particles likely to pass through the magnet, otherwise results from tracking may not be accurate.

5.4 Multipole decomposition: some comments

In this section, we have considered two techniques for deriving the multipole components of two-dimensional magnetostatic fields. We have seen that while the multipole components can be obtained, in principle, from a simple least-squares fit of a polynomial to the field components along one or other of the coordinate axes, there are advantages to basing the fit on field data obtained on a circle enclosing the origin, with as large a radius as possible. In the next section, we shall see how the idea of a multipole expansion can be generalized to three dimensions, and how a multipole decomposition can be performed in that case. However, it is worth pausing to consider in a little more detail some of the reasons for wishing to represent a field as a set of (multipole) modes.

It is of course possible to represent a magnetic field using a set of numerical field data, giving the three field components on points forming a ‘mesh’ covering the region of interest. In some ways, this is a very convenient representation, since it is the one usually provided directly by a magnetic modelling code: further processing is usually required to arrive at other representations. However, while a numerical field map in two dimensions is often a practical representation, in three dimensions the amount of data in even a relatively simple magnet can become extremely large, especially if a high resolution is required for the mesh. A multipole representation, on the other hand, provides the description of a magnetic field as a relatively small set of coefficients, from which the field components at any point can be reconstructed, using the basis functions. In other words, a multipole representation is more ‘portable’ than a numerical field map.

Secondly, a representation based on a multipole expansion lends itself to further manipulation in ways that a numerical field map does not. For example, any noise in the data (from measurement or computational errors) can be ‘smoothed’ by suppression of higher-order modes. Conversely, random errors can be introduced into data based on a model with perfect symmetry by introducing multipole coefficients corresponding to ‘forbidden’ harmonics. There will of course be issues surrounding the suppression or enhancement of errors by adjusting the multipole coefficients; however, one benefit of this approach is that for *any* set of multipole coefficients, the field is at least a physical field, in the sense of satisfying Maxwell’s equations. The same will not usually be true if, for example, a general smoothing algorithm is applied to a numerical field map.

Finally, one of the main motivations for performing a multipole decomposition of a field is to provide data in a format appropriate for many beam dynamics codes. Accuracy is one criterion often important for beam dynamics codes: efficiency is another. Characterization of a storage ring frequently requires tracking of thousands of particles over hundreds or thousands of turns, through a beam line that can easily consist of hundreds of magnetic elements. Numerical integration of the equations of motion for a particle in a numerical field map is generally too slow to be a practical method. There are many techniques that can be used to improve the efficiency of particle tracking in accelerators: one of the most common is the ‘thin lens’ method. The dynamical effects of dipoles and quadrupoles usually need to be represented with high accuracy. Fortunately, for these magnets, it is possible to write down accurate solutions to the equations of motion in closed form, allowing tracking through a magnet of given length to be performed in a single step. The same is not true for sextupoles, or higher-order multipoles; however, it is usually sufficiently accurate to represent such magnets by a model in which the length of the magnet approaches zero, but where the integrated strength (the multipole coefficient multiplied by the length) remains constant. For such a ‘thin lens’ it is possible to write down exact solutions to the equations of motion, allowing tracking again to be performed in a single step. A quadrupole with higher-order multipole errors can be represented as a ‘long’ perfect quadrupole field, with a set of ‘thin’ multipoles at one end, or at the centre. However, construction of such a model for a tracking code requires a multipole decomposition of the field obtained from a magnet modelling code.

We should emphasize that in our discussion of multipole decomposition, here and in Sections 5.2 and 5.3, we have made no clear distinction between field data obtained from a computational model, or from measurement of a real magnet. Of course, it is much easier to obtain the data required from a

computational model: it is then quite straightforward to perform the required decomposition to determine the values of the various multipole coefficients. Unfortunately, the data do not include manufacturing errors, which can be very important. Measurements provide more realistic data: however, many other issues need to be addressed, including accuracy of field measurements, alignment of the measurement instruments with respect to the magnet, etc. Such issues are beyond the scope of our discussion.

6 Three-dimensional fields

In the previous sections, we have restricted ourselves to the case where the magnetic field is independent of the longitudinal coordinate. The multipole modes that we can use for such fields actually provide a good description for many accelerator multipole magnets, even though such magnets of course have finite length. The ends or ‘fringe fields’ of dipoles, quadrupoles and so on, where the field strengths often vary rapidly with longitudinal position, cannot be accurately represented by two-dimensional fields; however, in many accelerators, only the fringe fields of dipoles have a significant impact on the dynamics.

But there are cases where a full three-dimensional description of a magnetic field is desirable, or even necessary. For example, the fields of insertion devices (wiggler and undulators) are often represented as a sequence of short dipoles of alternating polarity; however, where the period becomes small compared with the aperture, the three-dimensional nature of the field can start to have effects that cannot be ignored. There can even be cases where ‘conventional’ multipoles designed for special situations (for example, where very wide aperture is required, and where the length of the magnet needs to be short, because of space constraints) can have fringe fields that affect the dynamics to a significant extent.

It is therefore of somewhat more than purely academic interest to consider how two-dimensional multipole representations may be generalized to three dimensions. As usual, there are many different ways to approach the problem: the method that is used will often depend on the problem to be solved. In the following sections, we describe two rather general methods that may be of use in many situations arising in accelerators. First, we consider a field expansion based on Cartesian modes. While this provides some nice illustrations, the Cartesian expansion does have some disadvantages. To address these disadvantages, we describe how a field expansion based on polar coordinates can be performed.

6.1 Cartesian modes

Consider the field given by

$$B_x = -B_0 \frac{k_x}{k_y} \sin k_x x \sinh k_y y \sin k_z z, \quad (43)$$

$$B_y = B_0 \cos k_x x \cosh k_y y \sin k_z z, \quad (44)$$

$$B_z = B_0 \frac{k_z}{k_y} \cos k_x x \sinh k_y y \cos k_z z. \quad (45)$$

As may easily be verified, this field satisfies

$$\text{curl } \vec{B} = 0.$$

Furthermore, the equation

$$\text{div } \vec{B} = 0$$

is satisfied if

$$k_y^2 = k_x^2 + k_z^2. \quad (46)$$

We conclude that, as long as the constraint (46) is satisfied, the fields (43)–(45) provide solutions to Maxwell’s equations in regions with constant permeability, and static (or zero) electric fields. Of course, it is possible to find similar sets of equations but with different ‘phase’ along each of the coordinate axes, and with the hyperbolic trigonometric function appearing for the dependence on x or z , rather than y .

By superposing fields, with appropriate variations on the form given by Eqs. (43)–(45), it is possible to construct quite general three-dimensional magnetic fields. For example, a slightly more general field than that given by Eqs. (43)–(45) can be obtained simply by superposing fields with different mode numbers and amplitudes:

$$B_x = - \iint \tilde{B}(k_x, k_z) \frac{k_x}{k_y} \sin k_x x \sinh k_y y \sin k_z z dk_x dk_z, \quad (47)$$

$$B_y = \iint \tilde{B}(k_x, k_z) \cos k_x x \cosh k_y y \sin k_z z dk_x dk_z, \quad (48)$$

$$B_z = \iint \tilde{B}(k_x, k_z) \frac{k_z}{k_y} \cos k_x x \sinh k_y y \cos k_z z dk_x dk_z. \quad (49)$$

In this form, we see already how to perform a mode decomposition, i.e., how we can determine the coefficients $\tilde{B}(k_x, k_z)$ as functions of the ‘mode numbers’ k_x and k_z . If we consider in particular the vertical field component on the plane $y = y_0$, then we have from (48)

$$\frac{B_y}{\cosh k_y y_0} = \iint \tilde{B}(k_x, k_z) \cos k_x x \sin k_z z dk_x dk_z.$$

Hence $\tilde{B}(k_x, k_z)$ may be obtained from an inverse Fourier transform of $B_y(x, z)/\cosh k_y y_0$. Given field data on a grid over x and z , we can then perform numerically an inverse discrete Fourier transform, to obtain a set of coefficients $\tilde{B}(k_x, k_z)$. Note that once we have obtained these coefficients, we can then reconstruct all field components at all points in space. This is an important consequence of the strong constraints on the fields provided by Maxwell’s equations: in general, for a static field, if we know how one field component varies over a two-dimensional plane, then we can deduce how all the field components vary over all space (on and off the plane).

Let us consider an example. To keep things simple, we shall again work with the case where the field is independent of one coordinate: now, however, we shall assume that the fields are independent of the horizontal transverse, rather than the longitudinal coordinate. This may be a suitable model for a planar wiggler or undulator with very wide poles. The model may of course be extended to include dependence of the fields on the horizontal transverse coordinate: although our immediate example strictly deals with a two-dimensional field, the extension to three dimensions is quite straightforward.

Suppose that the mode amplitude function $\tilde{B}(k_x, k_z)$ has the form

$$\tilde{B}(k_x, k_z) = \delta(k_x) \tilde{B}(k_z), \quad (50)$$

where $\delta(k_x)$ is the Dirac delta function. The delta function has the property that, for any function $f(k_x)$,

$$\int_{-\infty}^{\infty} \delta(k_x) f(k_x) dk_x = f(0).$$

Using (50) in Eqs. (47)–(49) gives

$$\begin{aligned} B_x &= 0, \\ B_y &= \int \tilde{B}(k_z) \cosh k_z y \sin k_z z dk_z, \\ B_z &= \int \tilde{B}(k_z) \sinh k_z y \cos k_z z dk_z. \end{aligned}$$

There is no horizontal transverse field component, and the vertical and longitudinal field components have no dependence on x : we have a two-dimensional field. In a plane defined by a particular value for the vertical coordinate, $y = y_0$, the vertical field component is given by

$$B_y(z) = \cosh k_z y_0 \int \tilde{B}(k_z) \sin k_z z dk_z.$$

The mode amplitude function can be obtained from a Fourier transform of the vertical component of the magnetic field on the plane $y = y_0$. Usually, we will have a finite set of field data, obtained from a magnet modelling code, or from measurements on a real device.

Suppose that we have a data set of $2M + 1$ vertical field measurements, taken at locations

$$y = y_0, \quad z = \frac{m}{M} \hat{z}, \quad (51)$$

where m is an integer in the range $-M \leq m \leq M$. The field at any point is given by

$$B_y(y, z) = \sum_{m=-M}^M \tilde{B}_m \cosh m k_z y \sin m k_z z,$$

where

$$k_z = \frac{2\pi}{2\hat{z}}.$$

Note that in this case, the field is antisymmetric about $z = 0$, i.e.,

$$B_y(y, -z) = -B_y(y, z).$$

The mode amplitudes \tilde{B}_m are obtained by

$$\tilde{B}_m = \frac{1}{\cosh m k_z y_0} \frac{1}{2M} \sum_{m'=-M}^M B_y(y_0, z) \sin m' k_z z.$$

Note that, because of the antisymmetry of the field

$$\tilde{B}_{-m} = -\tilde{B}_m.$$

As a specific numerical example, let us construct an ‘artificial’ data set along a line $y = y_0 = 0.25$, and with $\hat{z} = 3$. The data are constructed using a function that gives a sinusoidal variation in the field along z up to $|z| < 1.25$; then a continuous and smooth (continuous first derivative) fall-off to zero field for $|z| > 1.5$. For this numerical example, we do not worry unduly about units: the reader may assume lengths in cm, fields in kG, or any other preferred units. Initially, we take $M = 40$, i.e., we assume we have 81 measurements of the field (or, we have computed the field from a model at 81 equally-spaced points along z ; strictly speaking, because we are dealing with the case where the field is antisymmetric in z , we need only half this number of field measurements or computations). The field ‘data’, the fitted field (reconstructed using the mode amplitudes) in $y = 0.25$, and the mode amplitudes, are shown in Fig. 15.

It is interesting to compare with the situation where we have only 31 field measurements or computations, i.e., $M = 15$. Using the same function that we used to construct the data set with 81 data points, we produce the fit and the mode amplitudes shown in Fig. 16. Comparing Figs. 15 and 16, we see that in both cases the fitted field does pass exactly through all the data points. This is a necessary consequence of the fit, which is based on a discrete Fourier transform of the data points. However, using only 31 data points, there is a significant oscillation of the fitted field between the data points in the region $1.5 < |z| < 3$, where the field is actually zero (by construction). This is a consequence of the fact that we have ‘truncated’ some modes with non-negligible amplitude. The mode amplitudes in both cases are the same for mode numbers $-15 \leq m \leq m$; but with 81 data points we can determine amplitudes for a larger number of modes, which gives us a more accurate interpolation between the data points.

Having obtained fits to the field in the plane $y = 0.25$, we can reconstruct the field at any point, on or off the plane. It is often of interest to look at the mid-plane; usually, this is defined by $y = 0$. In this plane, we do not have any field data. However, we can compare the field produced by the fits with

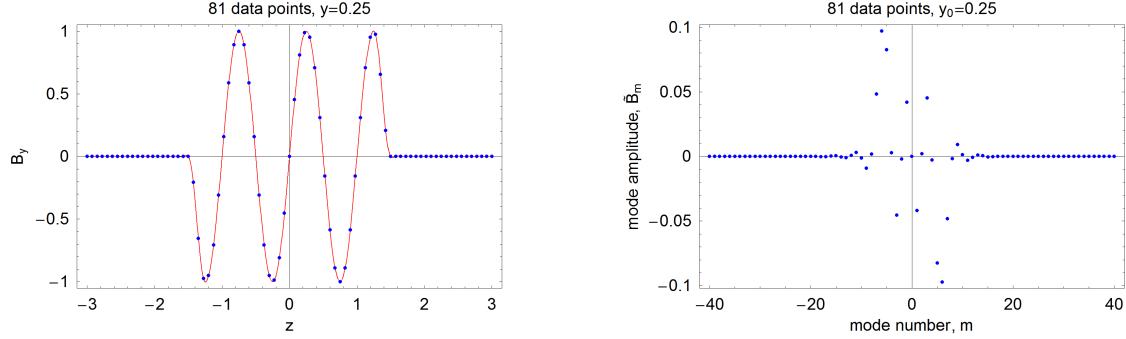


Fig. 15: Left: Field data (points) and fit (line) in a magnet with dependence of the field on longitudinal coordinate z . Right: Mode amplitudes. The field data consist of 81 measurements (or computations) at equally-spaced points from $z = -3$ to $z = +3$, and $y = 0.25$.

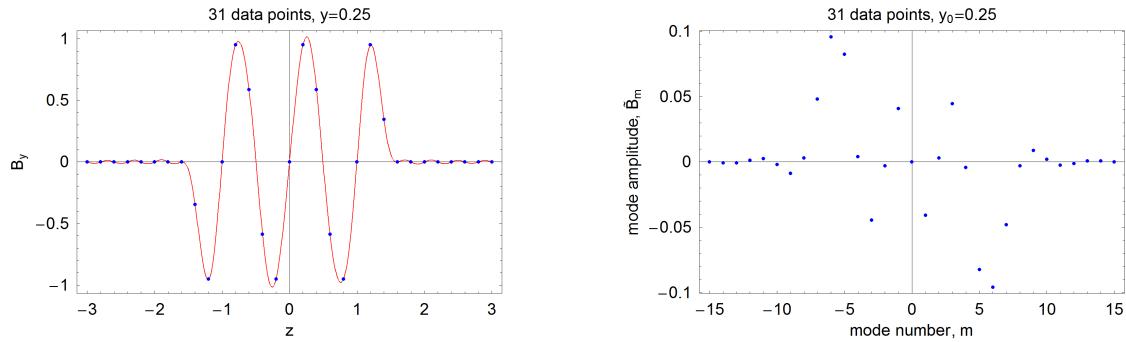


Fig. 16: Left: Field data (points) and fit (line) in a magnet with dependence of the field on longitudinal coordinate z . Right: Mode amplitudes. The field data consist of 31 measurements (or computations) at equally-spaced points from $z = -3$ to $z = +3$, and $y = 0.25$.

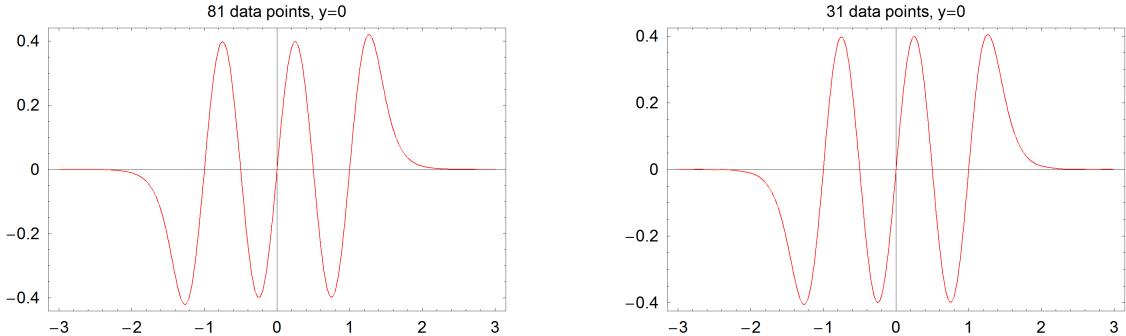


Fig. 17: Field on the plane $y = 0$ determined from fits to the data shown in Figs. 15 and 16. Left: fit determined from data set with 81 data points. Right: fit determined from data set with 31 data points.

81 data points and with 31 data points: these fields are shown in Fig. 17. We see that the fit based on 31 data points produces an essentially identical field on $y = 0$ as the fit based on 81 data points. (The data points in each case are taken on the plane $y = 0.25$.) This is a consequence of the ‘suppression’ of higher-order modes, that arises from the hyperbolic dependence of the field on the y coordinate.

To emphasize the significance of the hyperbolic dependence of the field on the vertical coordinate, we can look at the variation of B_y with y , for a given value of z . We choose $z = 0.25$, which corresponds to a peak in the vertical field component as a function of z . The variation of B_y with y for the two cases (fit based on 81 data points, and fit based on 31 data points) is shown in Fig. 18.

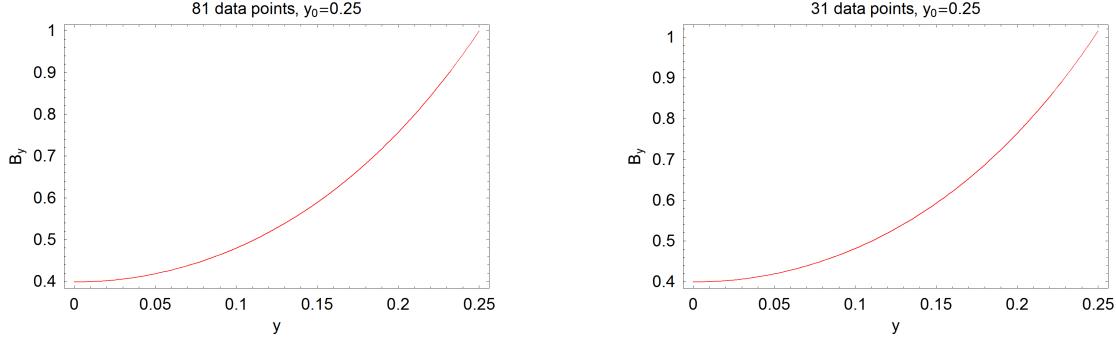


Fig. 18: Vertical field component as a function of y , for $z = 0.25$. The field is determined from fits to the data shown in Figs. 15 and 16. Left: fit determined from data set with 81 data points. Right: fit determined from data set with 31 data points.

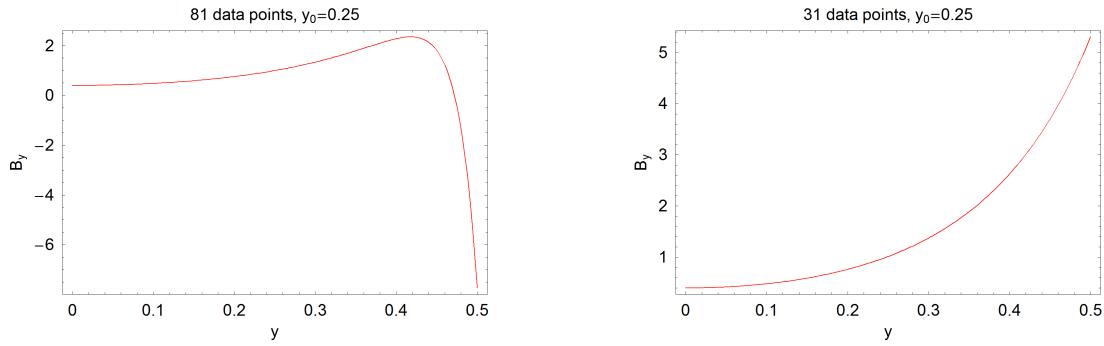


Fig. 19: Vertical field component as a function of y , for $z = 0.25$. The field is determined from fits to the data shown in Figs. 15 and 16. Left: fit determined from data set with 81 data points. Right: fit determined from data set with 31 data points.

Up to $y = 0.25$, the two fits give essentially the same field. However, if we try to extrapolate beyond this plane (the plane on which the fit was performed), we see dramatically different behaviour. Figure 19 compares the vertical field component obtained from the two fits (81 data points and 31 data points), again at $z = 0.25$, but now with a range of y from 0 to 0.5. In one case (81 data points), the field increases to a maximum before dropping rapidly. In the other case (31 data points), the field increases monotonically over the range. The reason for the different behaviour is the additional modes in the fit to the set of 81 data points. These higher order modes make only a small additional contribution to the field for $|y| < 0.25$; but for values of the vertical coordinate beyond this value, because of the hyperbolic dependence of y , the contribution of these modes becomes increasingly significant, and eventually, dominant.

The behaviour of the field fits for $|y| > 0.25$ is a clear illustration of why it is dangerous to extrapolate the fit beyond the region enclosed by the plane of the fit. In this case, because of the symmetry in the vertical direction, the region enclosed is between the planes $y = -0.25$ and $y = +0.25$. The ‘safe’ region is also bounded in z , by $z = -0.3$ and $z = +0.3$; because we use discrete mode numbers in z , the fitted fields will in fact be periodic in z , and will repeat with period $z = 0.6$. In general, there will be similar periodicity in x ; however, in this particular example, we analysed a field that was independent of x , so the ‘safe’ region of the fit is unbounded in x .

6.2 Cylindrical modes

The Cartesian modes discussed in Section 6.1 are often useful for describing fields in insertion devices, particularly those that have weak variation of the field with x , and periodic behaviour in z (over some range): because the modes ‘reflect’ the geometry, it is often possible to achieve good fits to a given field using a small number of modes. To maximize the region over which the fit is reliable, one needs to choose a plane with a value of y as large as possible, with x and z extending out as far as possible on this plane. For a planar undulator or wiggler, it is often possible to choose a plane close to the pole tips in which x in particular extends over the entire vacuum chamber.

However, for other geometries, the Cartesian modes may not provide a convenient basis. For example, if the magnet has a circular aperture, then the plane that provides the largest range in x is the mid-plane, $y = 0$, and as y increases, the available range in x decreases. To base the fit on the Cartesian basis requires some compromise between the range of reliability in the horizontal transverse and vertical directions.

Fortunately, it is possible to choose an alternative basis for magnets with circular aperture, in which the field fit can be based on the surface of a cylinder inscribed through the magnet. In that case, the radius of the cylinder can be close to the aperture limit, maximizing the range of reliability of the fit. The appropriate modes in this case are most easily expressed in cylindrical polar coordinates.

A field with zero divergence and curl (and hence satisfying Maxwell’s equations for static fields in regions with uniform permeability) is given by

$$B_r = \int dk_z \sum_n \tilde{B}_n(k_z) I'_n(k_z r) \sin n\theta \cos k_z z, \quad (52)$$

$$B_\theta = \int dk_z \sum_n \tilde{B}_n(k_z) \frac{n}{k_z r} I_n(k_z r) \cos n\theta \cos k_z z, \quad (53)$$

$$B_z = - \int dk_z \sum_n \tilde{B}_n(k_z) I_n(k_z r) \sin n\theta \sin k_z z. \quad (54)$$

Here, $I_n(k_z r)$ is the modified Bessel function of the first kind, of order n . Modified Bessel functions of the first kind for order $n = 0$ to $n = 3$ are plotted in Fig. 20. For small values of the argument ξ , the modified Bessel function of order n has the series expansion

$$I_n(\xi) = \frac{\xi^n}{2^n \Gamma(1+n)} + O(n+1). \quad (55)$$

For larger values of the argument, the modified Bessel functions $I_n(\xi)$ increase exponentially. This is significant: it means that if we fit a field to data on the surface of a cylinder of given radius, then residuals of the fit will decrease exponentially within the cylinder towards $r = 0$, and increase exponentially outside the cylinder with increasing r . The ‘safe’ region of the fit will be within the cylinder.

Note that Eqs. (52)–(54) may be generalized to include different ‘phases’ in the azimuthal angle θ and the longitudinal coordinate z .

An attractive feature of the polar basis is that it is possible to draw a direct connection between the three-dimensional modes in this basis and the multipole components in a two-dimensional field. Consider a mode amplitude $\tilde{B}_n(k_z)$ given (for some particular value of n) by

$$\tilde{B}_n(k_z) = 2^n \Gamma(1+n) C_n \frac{\delta(k_z)}{nk_z^{n-1}}, \quad (56)$$

where $\delta(k)$ is the Dirac delta function, and C_n is a constant. Substituting these mode amplitudes into Eqs. (52)–(54), using the expansion (55), and performing the integral over k_z gives

$$B_r = \sum_n C_n r^{n-1} \sin n\theta,$$

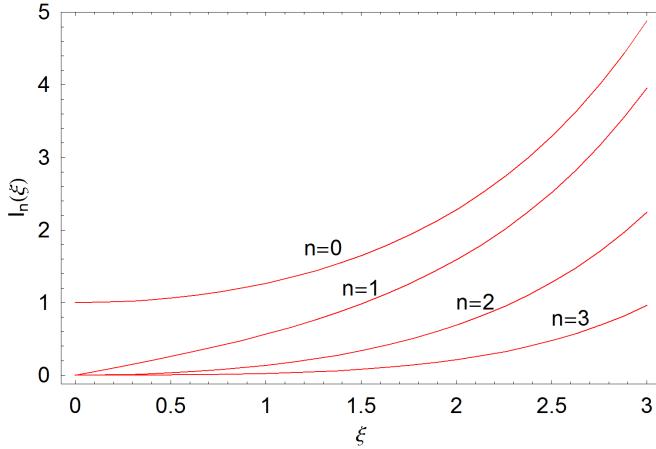


Fig. 20: Modified Bessel functions of the first kind, of order $n = 0$ to $n = 3$

$$\begin{aligned} B_\theta &= \sum_n C_n r^{n-1} \cos n\theta, \\ B_z &= 0. \end{aligned}$$

Comparing with Eq. (24), we see that this is a multipole field of order n . Thus a two-dimensional multipole field is a special case of a three-dimensional field (52)–(54), with mode coefficient given by Eq. (56).

In general, the mode coefficients $\tilde{B}_n(k_z)$ may be obtained by a Fourier transform of the field on the surface of a cylinder of given radius. For example, it follows from Eq. (52) that

$$\frac{B_r}{I'_n(k_z r)} = \int dk_z \sum_n \tilde{B}_n(k_z) \sin n\theta \cos k_z z.$$

An elegant feature of the polar basis, as compared to the Cartesian basis discussed in Section 6.1, is that the modes reflect the real periodicity of the field in the angle coordinate θ . In the Cartesian basis, the modes were periodic in x , although the field, in general, would not have any periodicity in x .

Since the mode coefficients $\tilde{B}_n(k_z)$ are related to the multipole coefficients in a two-dimensional field, we can use these coefficients to extend the idea of a multipole to a three-dimensional field. Strictly speaking, the mode coefficients $\tilde{B}_n(k_z)$ are related to the field by a two-dimensional Fourier transform; however, we can perform a one-dimensional inverse Fourier transform (in the z variable) to obtain a set of functions which represent, in some sense, the ‘multipole components’ of a three-dimensional field as a function of z . Here, we use the term ‘multipole components’ rather loosely, since a multipole field is strictly defined only in the two-dimensional case (i.e., for a field that is independent of the longitudinal coordinate). A quantity that is perhaps easier to interpret is the contribution to the field at any point made by the mode coefficients $\tilde{B}_n(k_z)$ with a given n . For $n = 1$, the field components B_r and B_θ at any point in z will behave as for a dipole field; for $n = 2$, B_r and B_θ will behave as for a quadrupole field, and so on.

As an illustrative example, we consider the field in a specific device: the wiggler in a damping ring for TESLA (a proposed linear collider) [10]. This wiggler has a peak field of 1.6 T and period 400 mm; the total length of wiggler in each of the TESLA damping rings would be over 400 m. The field in the wiggler has been studied extensively, because of concerns that dynamical effects associated with the nonlinear components in the field would limit the acceptance of the damping ring [11]. A model was constructed for one quarter period of the magnet, which allowed the field at any point within the body of the magnet to be computed. Effects associated with the ends of the wiggler were neglected, but could in principle be included in the study. By performing a mode decomposition using the techniques described

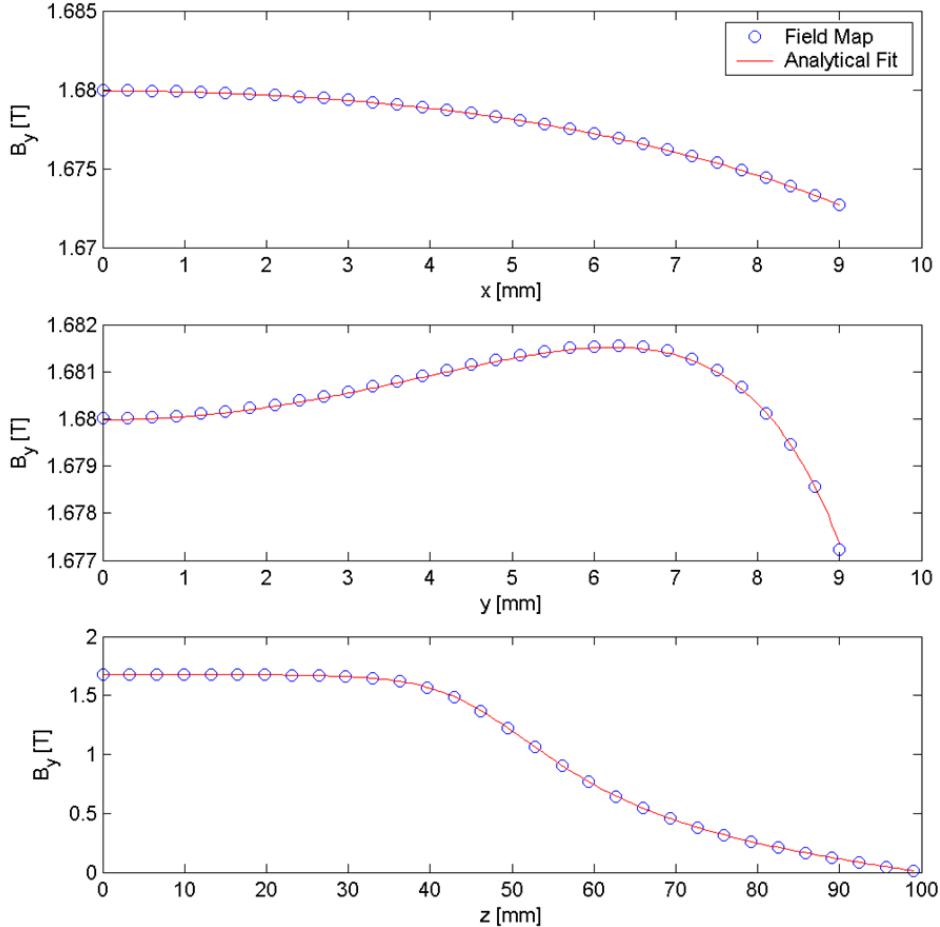


Fig. 21: Fit to the field of one quarter of one period of the TESLA damping ring wiggler

above, it was possible to construct an accurate dynamical model allowing fast tracking to characterize the acceptance of the damping ring. The methods used for the dynamical analysis are beyond the scope of the present discussion; however, we present the results of the analysis relating directly to the field, to illustrate the methods described in this section.

A model of the wiggler was used to compute the magnetic field on a mesh of points bounded by a cylinder of radius 9 mm, within one quarter period of the wiggler. Although all field components were computed on the mesh, which covered the interior of the cylinder as well as the surface, only the radial field component on the surface of the cylinder was used to calculate the mode amplitudes. The fit can be validated by comparing the field ‘predicted’ by the fit with the field data (from the computational model) not used directly in the fitting procedure. A fit achieved using 7 azimuthal and 100 longitudinal modes is shown in Fig. 21. Each plot shows the variation of the vertical field as a function of one Cartesian coordinate, with the other two coordinates fixed at zero. In the vertical direction, the range shown is from the mid-plane of the wiggler to close to the pole tip. Note that the variation in the field in the transverse (x and y) directions is very small, less than 0.1% of the maximum field. It appears from Fig. 21, that there is very good agreement between the fit (line) and the field data (circles) within the cylinder on the surface of which the fit was performed.

The quality of the fit can be further illustrated by plotting the residuals, i.e., the difference between the fitted field and the field data. The residuals for the vertical field component on two horizontal planes, $y = 0$ mm and $y = 6$ mm are shown in Fig. 22. Note that to produce ‘smooth’ surface plots, we interpolate between the mesh points used in the computational model. On the mid-plane of the wiggler, the

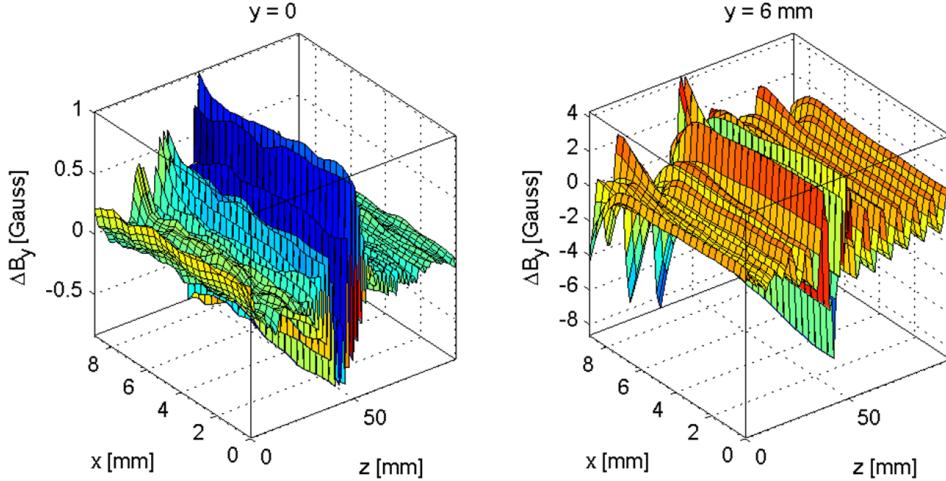


Fig. 22: Residuals of the fit to the field of one quarter of one period of the TESLA damping ring wiggler

residuals are less than 1 gauss (recall that the peak field is 1.6 T); the region shown in the left-hand plot in Fig. 22 lies entirely within the surface of the cylinder used in the fit. On the plane $y = 6$ mm, the residuals are somewhat larger, and show an exponential increase for large values of the horizontal transverse coordinate: but note that for values of x larger than about 6.7 mm, the points in the plot are *outside* the surface of the cylinder used for the fit. In the longitudinal direction, the residuals appear to be dominated by very high frequency modes: this suggests that it may be possible to reduce the residuals still further by increasing the number of longitudinal modes used in the fit. However, this fit was considered to be of sufficient quality to allow an accurate determination of the effect of the wiggler on the beam dynamics to be made.

The tools used for study of the beam dynamics were based on the mode coefficients determined by the fitting procedure. Once a fit has been obtained and shown to be of good quality, then, strictly speaking, further analysis of the field is not required. However, it is interesting to compute, from the mode amplitudes, the contribution to the field in the wiggler from different ‘multipole’ components, as a function of longitudinal position. As described above, the contribution of a multipole of order n is obtained by a one-dimensional (in the longitudinal dimension) inverse Fourier transform of the mode amplitudes $\tilde{B}_n(k_z)$. To obtain non-zero values for the contributions from multipoles higher than order $n = 1$ (dipole), we need to choose non-zero values for either the x or y coordinates at which we compute the field. We choose (arbitrarily) $x = 8$ mm, and $y = 0$ mm. The contributions to the vertical field component from multipoles of order 1 through 7 are shown in Fig. 23. Note that multipoles of even order are forbidden by the symmetry of the wiggler (see Section 5.1). We see from Fig. 23 that the dominant contribution by far is, as expected, the dipole component. The sextupole component is not insignificant; the contributions of higher order multipoles are extremely small, and the high-frequency ‘oscillation’ as a function of longitudinal position is probably unphysical, and the result of noise in the fitting.

It is worth making a few final remarks about mode decompositions for three-dimensional fields. First, as already mentioned, in many cases a full three-dimensional mode decomposition will not be necessary. While this does provide a detailed description of the field in a form suitable for beam dynamics studies, three-dimensional decompositions do rely on a large number of accurate and detailed field measurements. While such ‘measurements’ may be conveniently obtained from a model, it may be difficult or impractical to make such measurements on a real magnet. Fortunately, in many cases, a two-dimensional field description in terms of multipoles is sufficient. Generally, a three-dimensional analysis only need be undertaken where there are grounds to believe that the three-dimensional nature of

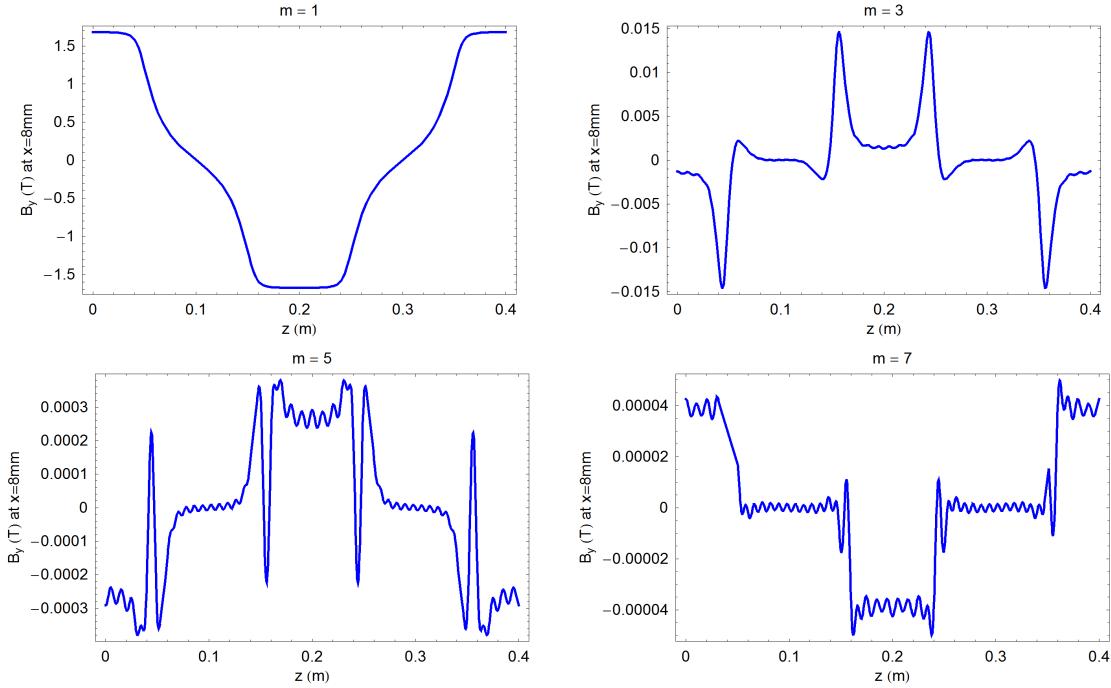


Fig. 23: Multipole contributions to the field in the TESLA damping wiggler as a function of longitudinal position, at $x = 8$ mm and $y = 0$ mm, for orders 1 (dipole), 3 (sextupole), 5 (decapole) and 7

the field is likely to have a significant impact on the beam dynamics.

Second, we have already emphasized that to obtain an accurate description of the field within some region in terms of a mode decomposition, the mode amplitudes should be determined by a fit on a surface enclosing the region of interest. Outside the region bounded by the surface of the fit, the fitted field can be expected to diverge exponentially from the real field. However, in choosing the surface for the fit, the geometry of the magnet will impose some constraints. A magnet with a wide rectangular aperture may lend itself to a description using a Cartesian basis (fitting on the surface of a rectangular box); a circular aperture, however, is more likely to require use of a polar basis (fitting on the surface of a cylinder with circular cross-section). Both cases have been described above. It may be appropriate in other cases to perform a fit on the surface of a cylinder with elliptical cross-section. The basis functions in this case involve Mathieu functions. For further details, the reader is referred to work by Dragt [12] and by Dragt and Mitchell [13].

Appendices

A The vector potential

Our analysis of iron-dominated multipole magnets in Section 4.2 was based on the magnetic scalar potential φ . The magnetic flux density can be derived from a scalar potential

$$\vec{B} = -\text{grad } \varphi$$

in the case where the flux density has vanishing divergence and curl:

$$\text{div } \vec{B} = \text{curl } \vec{B} = 0.$$

More generally (in particular, where the flux density has non-vanishing curl) one derives the magnetic flux density from a vector potential \vec{A} , using

$$\vec{B} = \text{curl } \vec{A}. \quad (\text{A.1})$$

Although we have not required the vector potential in our discussion of Maxwell's equations for accelerator magnets, it is sometimes used in analysis of beam dynamics. In particular, descriptions of the dynamics based on Hamiltonian mechanics generally use the vector potential rather than the magnetic flux density or the magnetic scalar potential. We therefore include here a brief discussion of the vector potential, paying attention to aspects relevant to the descriptions we have developed for two-dimensional and three-dimensional magnet fields.

First, we note that the divergence of any curl is identically zero:

$$\operatorname{div} \operatorname{curl} \vec{V} \equiv 0,$$

for any differentiable vector field \vec{V} . Thus, if we write $\vec{B} = \operatorname{curl} \vec{A}$, then Maxwell's equation (2):

$$\operatorname{div} \vec{B} = 0,$$

is automatically satisfied. Maxwell's equation (3) in uniform media (constant permeability), with zero current and static electric fields gives

$$\operatorname{curl} \vec{B} = \mu \vec{J}, \quad (\text{A.2})$$

where \vec{J} is the current density. This leads to the equation for the vector potential:

$$\operatorname{curl} \operatorname{curl} \vec{A} \equiv \operatorname{grad}(\operatorname{div} \vec{A}) - \nabla^2 \vec{A} = \mu \vec{J}. \quad (\text{A.3})$$

Equation (A.3) is a second-order differential equation for the vector potential in a medium with permeability μ , where the current density is \vec{J} . This appears harder to solve than the first-order differential equation for the magnetic flux density, Eq. (A.2). However, Eq. (A.3) may be simplified significantly, if we apply an appropriate *gauge condition*. To understand what this means, recall that the magnetic flux density is given by the curl of the vector potential, and that the curl of the gradient of any scalar field is identically zero. Thus, we can add the gradient of a scalar field to a vector potential, and obtain a new vector potential that gives the same flux density as the old one. That is, if

$$\vec{B} = \operatorname{curl} \vec{A},$$

and

$$\vec{A}' = \vec{A} + \operatorname{grad} \psi, \quad (\text{A.4})$$

for an arbitrary differentiable scalar field ψ , then

$$\operatorname{curl} \vec{A}' = \operatorname{curl} \vec{A} + \operatorname{curl} \operatorname{grad} \psi = \operatorname{curl} \vec{A} = \vec{B}.$$

In other words, the vector potential \vec{A}' leads to exactly the same flux density as the vector potential \vec{A} . Since the dynamics of a given system are determined by the fields rather than the potentials, either \vec{A}' or \vec{A} is a valid choice for the description of the system. Equation (A.4) is known as a *gauge transformation*. The consequence of having the freedom to make a gauge transformation means that the vector potential for any given system is not uniquely defined: given some particular vector potential, it is always possible to make a gauge transformation without any change in the physical observables of a system. The analogue in the case of electric fields, of course, is that the ‘zero’ of the electric scalar potential can be chosen arbitrarily: only *changes* in potential (i.e., energy) are observable, so given some particular scalar potential field, it is possible to add a constant (that is, a quantity independent of position) and obtain a new scalar potential that gives the same physical observables as the original scalar potential.

For magnetostatic fields, we can use a gauge transformation to simplify Eq. (A.3). Suppose we have obtained a vector potential \vec{A} for some particular physical system. Define a scalar field ψ , which satisfies

$$\nabla^2 \psi = -\operatorname{div} \vec{A}. \quad (\text{A.5})$$

Then define

$$\vec{A}' = \vec{A} + \text{grad } \psi.$$

Since \vec{A}' and \vec{A} are related by a gauge transformation, they lead to the same magnetic flux density, and the same physical observables for the system. However, the divergence of \vec{A}' vanishes:

$$\text{div } \vec{A}' = \text{div } \vec{A} + \text{div grad } \psi = -\nabla^2 \psi + \nabla^2 \psi = 0,$$

where we have used Eq. (A.5). Thus, given any vector potential, we can make a gauge transformation to find a new vector potential that gives the same magnetic flux density, but has vanishing divergence. The *gauge condition*

$$\text{div } \vec{A} = 0 \quad (\text{A.6})$$

is known as the *Coulomb gauge*. It is possible to work with other gauge conditions (for example, for time-dependent electromagnetic fields the Lorenz gauge condition is often more appropriate); however, for our present purposes, the Coulomb gauge leads to a simplification of Eq. (A.3), which now becomes

$$\nabla^2 \vec{A} = -\mu \vec{J}. \quad (\text{A.7})$$

Equation (A.7) is Poisson's equation for a vector field. Note that despite being a second-order differential equation, it is in a sense simpler than Maxwell's equation (3), since we have 'decoupled' the components of the vectors; that is, we have a set of three uncoupled second-order differential equations, where each equation relates a component of the vector potential to the corresponding component of the current density. Equation (A.7) has the solution

$$\vec{A}(\vec{r}) = -\frac{\mu}{4\pi} \int \frac{\vec{J}(\vec{r}')}{|\vec{r} - \vec{r}'|} d^3 r'.$$

In this form, we see that the potential at a point in space is inversely proportional to the distance from the source.

Now, consider the potential given by

$$A_x = 0, \quad A_y = 0, \quad A_z = -\text{Re} \frac{C_n(x+iy)^n}{n}. \quad (\text{A.8})$$

Taking derivatives, we find that

$$\begin{aligned} \frac{\partial A_z}{\partial x} &= -\text{Re} C_n(x+iy)^{n-1}, \\ \frac{\partial A_z}{\partial y} &= \text{Im} C_n(x+iy)^{n-1}. \end{aligned}$$

Then, since A_x and A_y are zero, we have

$$\vec{B} = \text{curl } \vec{A} = \left(\frac{\partial A_z}{\partial y}, -\frac{\partial A_z}{\partial x}, 0 \right).$$

Hence

$$B_y + iB_x = C_n(x+iy)^{n-1}, \quad (\text{A.9})$$

which is just the multipole field. Thus Eq. (A.8) is a potential that gives a multipole field. Note also that, since A_z is independent of z , this potential satisfies the Coulomb gauge condition (A.6):

$$\text{div } \vec{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} = 0.$$

An advantage of working with the vector potential in the Coulomb gauge is that, for multipole fields, the transverse components of the vector potential are both zero. This simplifies, to some extent, the Hamiltonian equations of motion for a particle moving through a multipole field. However, note that the longitudinal component B_z of the magnetic flux density is zero in this case. To generate a solenoidal field, with B_z equal to a non-zero constant, we need to introduce non-zero components for A_x , or A_y , or both. For example, a solenoid field with flux density B_{sol} may be derived from the vector potential:

$$A_x = -\frac{1}{2}B_{\text{sol}}y, \quad A_y = \frac{1}{2}B_{\text{sol}}x.$$

Let us return for a moment to the case of multipole fields. If we work in a gauge in which the transverse components of the vector potential are both zero, then the field components are given by

$$B_y = -\frac{\partial A_z}{\partial x}, \quad B_x = \frac{\partial A_z}{\partial y}.$$

From these expressions, we see that if we take any two points with the same y coordinate, then the difference in the vector potential between these two points is given by the ‘flux’ passing through a line between these points:

$$\Delta A_z = - \int B_y dx.$$

Similarly for any two points with the same x coordinate

$$\Delta A_z = \int B_x dy.$$

In general, for a field that is independent of z , and working in a gauge where $A_x = A_y = 0$, we can write

$$\Delta A_z = \frac{\Delta \Phi}{\Delta z}, \tag{A.10}$$

where ΔA_z is the change in the vector potential between two points P_1 and P_2 in a given plane $z = z_0$; and $\Delta \Phi$ is the magnetic flux through a rectangular ‘loop’ with vertices P_1 , P_2 , P_3 and P_4 : see Fig. A.1. P_3 and P_4 are points obtained by transporting P_1 and P_2 a distance Δz parallel to the z axis. Equation (A.10) can also be obtained by applying Stokes’s theorem to the loop $P_1P_2P_3P_4$, with the relationship (A.1) between \vec{B} and \vec{A} :

$$\int \vec{A} \cdot d\vec{l} = \int \text{curl } \vec{A} \cdot d\vec{S} = \int \vec{B} \cdot d\vec{S},$$

hence

$$A_z(P_2)\Delta z - A_z(P_1)\Delta z = \Delta \Phi.$$

Finally, we give the vector potentials corresponding to three-dimensional fields. In the Cartesian basis, with the field given by Eqs. (43)–(45), a possible vector potential (in the Coulomb gauge) is

$$\begin{aligned} A_x &= 0, \\ A_y &= B_0 \frac{k_z}{k_x k_y} \sin k_x x \sinh k_y y \cos k_z z, \\ A_z &= -B_0 \frac{1}{k_x} \sin k_x x \cosh k_y y \sin k_z z. \end{aligned}$$

In the polar basis, with the field given by Eqs. (52)–(54), a possible vector potential is

$$\begin{aligned} A_r &= -\frac{r}{m} I_m(k_z r) \cos n\theta \sin k_z z, \\ A_\theta &= 0, \\ A_z &= -\frac{r}{2m} I'_m(k_z r) \cos n\theta \sin k_z z. \end{aligned}$$

However, note that this potential does not satisfy the Coulomb gauge condition.

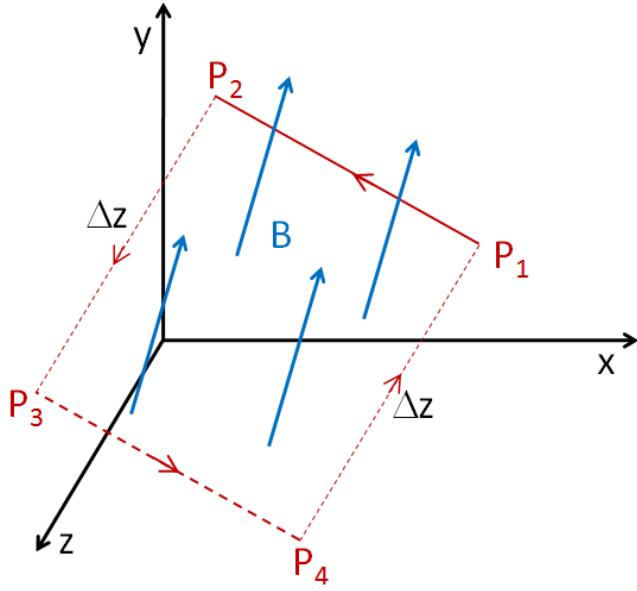


Fig. A.1: Interpretation of the vector potential in a two-dimensional magnetic field (i.e., a field that is independent of z). The change in the vector potential between P_1 and P_2 is equal to the flux of the magnetic field through the loop $P_1P_2P_3P_4$, divided by Δz .

References

- [1] Vector Fields Software, <http://www.cobham.com>.
- [2] Computer Simulation Technology, <http://www.cst.com>.
- [3] ESRF Insertion Devices Group, <http://www.esrf.eu/Accelerators/Groups/InsertionDevices/Software/Radia>.
- [4] S. Russenschuck, Foundation of numerical field computation, Proceedings of the CERN Accelerator School Course on Magnets, Bruges, Belgium (2009).
- [5] J. D. Jackson, Classical Electrodynamics, 3rd edition (Wiley, New York, 1999).
- [6] L. Bottura, Superconducting magnets, Proceedings of the CERN Accelerator School Course on Magnets, Bruges, Belgium (2009).
- [7] A. Chao and M. Tigner (editors), Handbook of Accelerator Physics and Engineering (World Scientific, Singapore, 1999).
- [8] B. J. A. Shepherd, N. Marks, Quadrupole magnets for the 20 MeV FFAG, EMMA, Proceedings of PAC07, Albuquerque, New Mexico, USA (2007).
- [9] N. Marks et al., Development and adjustment of the EMMA quadrupoles, Proceedings of EPAC08, Genoa, Italy (2008).
- [10] TESLA Technical Design Report (2001). http://tesla.desy.de/new_pages/TDR_CD/start.html.
- [11] A. Wolski, J. Gao and S. Guiducci, (editors) Configuration Studies and Recommendations for the ILC Damping Rings, LBNL-59449 (2006).
- [12] A. J. Dragt, ‘Lie methods for nonlinear dynamics with applications to accelerator physics,’ available at the URL <http://www.physics.umd.edu/dsat/dsatliemethods.html>.
- [13] C. E. Mitchell and A. J. Dragt, Computation of transfer maps from magnetic field data in wigglers and undulators, ICFA Beam Dynamics Newsletter, 42, (April 2007) p. 65.

Physics and measurements of magnetic materials

S. Sgobba

CERN, Geneva, Switzerland

Abstract

Magnetic materials, both hard and soft, are used extensively in several components of particle accelerators. Magnetically soft iron–nickel alloys are used as shields for the vacuum chambers of accelerator injection and extraction septa; Fe-based material is widely employed for cores of accelerator and experiment magnets; soft spinel ferrites are used in collimators to damp trapped modes; innovative materials such as amorphous or nanocrystalline core materials are envisaged in transformers for high-frequency polyphase resonant convertors for application to the International Linear Collider (ILC). In the field of fusion, for induction cores of the linac of heavy-ion inertial fusion energy accelerators, based on induction accelerators requiring some 10^7 kg of magnetic materials, nanocrystalline materials would show the best performance in terms of core losses for magnetization rates as high as 10^5 T/s to 10^7 T/s. After a review of the magnetic properties of materials and the different types of magnetic behaviour, this paper deals with metallurgical aspects of magnetism. The influence of the metallurgy and metalworking processes of materials on their microstructure and magnetic properties is studied for different categories of soft magnetic materials relevant for accelerator technology. Their metallurgy is extensively treated. Innovative materials such as iron powder core materials, amorphous and nanocrystalline materials are also studied. A section considers the measurement, both destructive and non-destructive, of magnetic properties. Finally, a section discusses magnetic lag effects.

1 Magnetic properties of materials: types of magnetic behaviour

The sense of the word ‘lodestone’ (waystone) as magnetic oxide of iron (magnetite, Fe_3O_4) is from 1515, while the old name ‘lodestar’ for the pole star, as the star leading the way in navigation, is from 1374. Both words are based on the original ‘lode’ spelling of ‘load’, issued from the old (1225) English ‘lad’, guide, way, course [1]. According to tradition, the mariner Flavio Gioia of Amalfi, born 1302, first discovered the ‘power of the lodestone’ enabling the manufacture of the first compass and replacing the lodestar in navigation. Nevertheless magnetite, known according to tradition to the Chinese since 2600 B.C., is cited first in Europe by Homer, relating that lodestone was already used by the Greeks to direct navigation at the time of the siege of Troy [2].

Magnetic properties of several materials are discussed in the text by Bozorth [3]. Conventional soft and hard magnetic materials are treated in Ref. [4]. The volume of O’Handley [5] covers a number of advanced materials, including amorphous and nanocrystalline materials. A general introduction to magnetic properties of materials can be found in the recent textbook by Cullity and Graham [6]. The comprehensive *Handbook of Magnetism and Advanced Magnetic Materials* [7] systematically covers very novel materials of technological and scientific interest in volume 4, including advanced soft magnetic materials for power applications.

Diamagnetism is due to induced currents opposing an applied field resulting in a small negative magnetic susceptibility κ . Diamagnetic contributions are present in all atoms, but are generally negligible in technical materials, except superconducting materials under some conditions. Monoatomic rare gases such He are diamagnetic, as well as most polyatomic gases such as N₂ (that might show nevertheless a net paramagnetic behaviour because of O₂ contamination). Since He is repelled by magnetic fields, operation of superconducting magnets in a weightless environment during orbital flights imposes a significant difficulty not present in laboratory experiments, already discussed and quantified in 1977 [8]. This concern is still present today: the effect of a magnetic field on diamagnetic liquid helium will be studied in the very near future in the cryogenic system of the cryomagnet of the Alpha Magnetic Spectrometer (AMS) experiment, foreseen on the International Space Station (ISS) [9].

Paramagnetism, corresponding to a positive susceptibility, is observed in many metals and substances including ferromagnetic and antiferromagnetic materials above their Curie (T_c) and Néel (T_N) temperature, respectively [10]. Particular care should be taken for some Ni-basis superalloys for non-magnetic application at very low T . Incoloy 800 (32.5Ni-21Cr-46Fe) features a magnetic permeability as low as 1.0092 at room temperature (annealed state, under a field of 15.9 kA/m). Nevertheless, due to a $T_c = -115$ °C, the alloy is ferromagnetic at cryogenic temperatures.

Ferromagnetism is due to the ordered array of magnetic moments, caused by the interaction of atomic spin moments occurring in certain conditions. Field-dependent permeability and persistent magnetization after the removal of magnetic field are observed for hysteretic ferromagnetic materials. Ordered ferromagnetic phase occurs for ferromagnets at $T < T_c$. Here T_c is the temperature above which spontaneous magnetization ‘vanishes’ [6]. The T_c of Fe, Ni and Co are 1043 K, 631 K and 1394 K, respectively. In general, ferrous alloys with body centred cubic (bcc) crystalline structure are ferromagnetic, while face centred cubic (fcc) are not. Nevertheless, rapidly solidified metastable alloys such as Fe-Cu alloys can show ferromagnetism in a wider composition range than expected, even in the fcc phase formed below 70% Fe content [11].

Antiferromagnetism corresponds to an antiparallel arrangement with zero net magnetic moment at $T < T_N$. ‘Non-magnetic’ austenitic stainless steels such as AISI 304L, 316L, 316LN, high Mn – high N stainless steels are antiferromagnetic under T_N and paramagnetic above T_N , where they obey a Curie–Weiss law ($\kappa = C/(T-\Theta)$, where Θ is a negative critical temperature and C is a constant (Fig. 1a).

Magnetic susceptibility of high Mn – high N grades such as P506 (approx. 0.012% C, 19% Cr, 11% Ni, 12% Mn, 0.9% Mo, 0.33% N) and UNS 21904 (approx. 0.028% C, 20% Cr, 7% Ni, 9% Mn, 0.35% N), particularly at 4.2 K, is lower than any traditional steel of the 300-series. As known, this is essentially due to the higher Mn content of the alloys (P506, Mn = 12%; UNS21904, Mn = 9%), stabilizing austenite (fcc ‘non-magnetic’ phase), and increasing T_N . Higher T_N allows for lower values of $\kappa (< 3 \cdot 10^{-3})$ at 4.2 K. Measured values of T_N are in agreement with the Warnes [12] law:

$$T_N/K = 90 - 1.25\text{Cr} - 2.75\text{Ni} - 5.5\text{Mo} - 14\text{Si} + 7.75\text{Mn} . \quad (1)$$

As an example, for steel P506, predicted $T_N = 121.5$ K, measured T_N is 125.7 K. Owing to the absence of precipitated δ-ferrite (bcc magnetic phase) in the weld, the presence of a laser weld has no measurable influence on the magnetic susceptibility of P506 (Fig. 1b). On the other hand, in welds of UNS21904, δ-ferrite contributes a significant increase of susceptibility in the whole T range [13].

Diamagnetism and paramagnetism can be considered as mainly due to the magnetic contribution of isolated atoms or molecules (in reality the existence of a Curie temperature T_c is

explained by interaction of elementary moments in the paramagnetic range). Ferromagnetism and antiferromagnetism are due to a larger order arrangement of electron spins and/or magnetic moments.

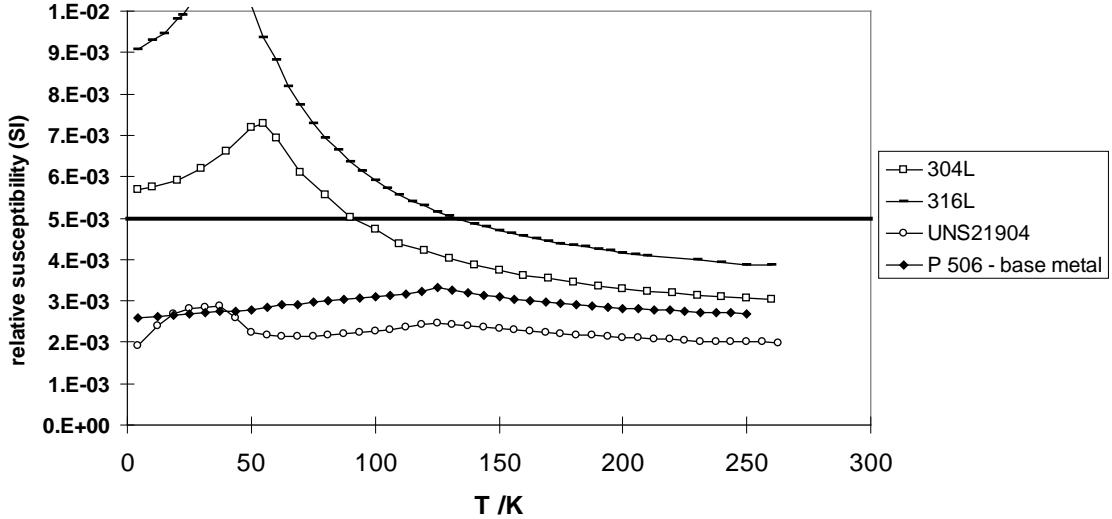


Fig. 1a: Magnetic susceptibility of different steels of the AISI 300 series, compared to high Mn – high N steels P506 and UNS 21904. Maximum allowed limit at CERN for non-magnetic applications is $5 \cdot 10^{-3}$. Peaks are at the respective Néel temperatures T_N . Above T_N , susceptibility obeys a Curie–Weiss law $\kappa = C/(T-\Theta)$.

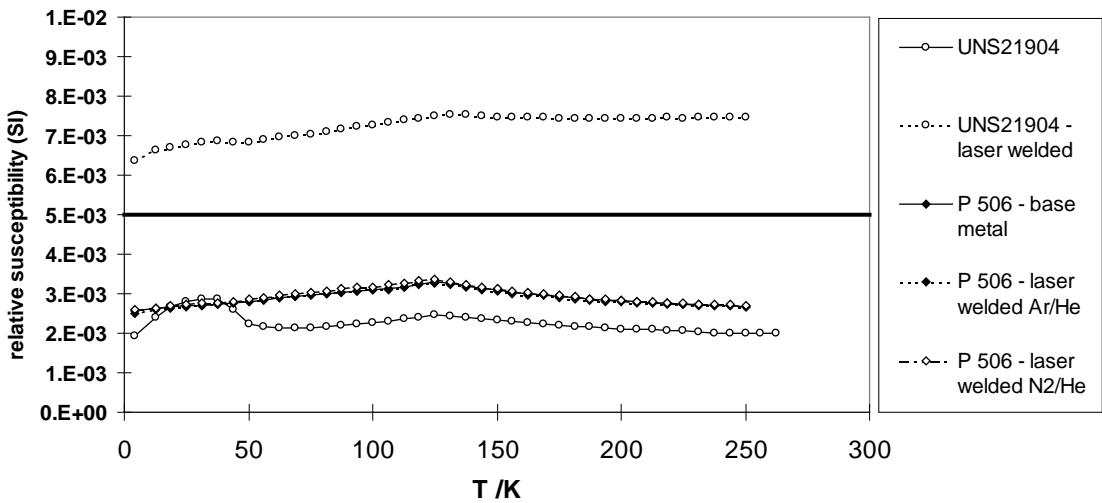


Fig. 1b: Compared magnetic susceptibility of steels P506 (base metal and weld) and UNS 21904 (idem). Maximum allowed limit at CERN in the welds for non-magnetic applications is $5 \cdot 10^{-3}$.

2 Soft ferromagnetic materials of interest for accelerator technology

2.1 Some definitions and units

Working in SI, we define the flux density or magnetic induction B (measured in T) and the magnetic field strength H ($\text{A} \cdot \text{m}^{-1}$). The permeability μ ($\text{H} \cdot \text{m}^{-1}$) is defined by

$$B = \mu \cdot H \quad (2)$$

The magnetization M ($\text{A} \cdot \text{m}^{-1}$) is defined as

$$B = \mu_0 \cdot (H + M) \quad (3)$$

where μ_0 ($\text{H} \cdot \text{m}^{-1}$) is the permeability of free space. The susceptibility κ (dimensionless) is the ratio M/H . From the above

$$\mu = \mu_0 \cdot (1 + \kappa) \quad (4)$$

The relative permeability is defined as $\mu_r = \mu / \mu_0$. From the above relationships, $\mu_r = 1 + \kappa$. A relative permeability of 1.005 corresponds to a susceptibility of $5 \cdot 10^{-3}$. The relative permeability μ_r and susceptibility κ are material properties, frequently reported for both magnetic and ‘non-magnetic’ materials.

2.2 Magnetization curves of soft ferromagnetic materials

A magnetization curve is the plot of the intensity of magnetization M or the magnetic induction B against the field strength H .

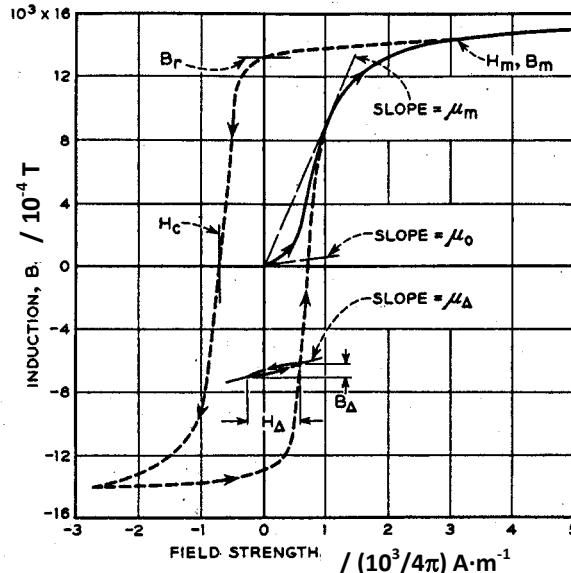


Fig. 2: Magnetization curve and hysteresis loop of iron (from Bozorth [3])

In the example of Fig. 2, the values of the field strength H_m and the magnetic induction B_m at the tip of the loop are defined. In the hysteresis loop, are also defined the residual induction B_r for which $H = 0$ (called retentivity if the tip corresponds to saturation) and the coercive force H_c for which $B = 0$ (called coercivity if the tip corresponds to saturation). For ferromagnetic materials, permeability is strongly dependent on field and tends to 1 for saturation. The initial and maximum permeability are easily identified in the curve. The magnetic properties of ferromagnetic materials are significantly

affected by their purity, the metalworking processes applied to the material (hot and cold working, subsequent annealing), and the resulting microstructure. Anisotropy effects due to texture can occur (effect of rolling, extrusion). By definition, soft ferromagnetic materials are easily magnetized and demagnetized (materials for transformer cores, for shielding of magnetic fields, magnetically soft ferrites for ac shielding applications, etc.). They have narrow hysteresis loops (low values of H_c), high permeability, low eddy-current losses, high magnetic saturation inductions.

2.3 High-purity iron

Iron is referred to as ‘high purity’ when the total concentration of impurities (mainly C, N, O, P, S, Si and Al) does not exceed a few hundred ppm. Otherwise it is rather referred to as low carbon steel or non-alloyed steel [14]. Very pure Fe features a high electrical conductivity and is unsuitable for ac applications. Typical impurity contents of different grades of iron are shown in Table 1.

Table 1: Impurity content of different iron grades. So-called ‘Armco irons’ can correspond to very different purity grades (from Ref. [15]).

Element	Concentration (at.%) $\times 10^3$		
	High-purity iron	ORNL Armco iron	BMI Armco iron
Al	0.21–2.1	<105	<0.4
Ca	0.14–1.4	<54	<0.9
Cr		90	0.09–0.9
Cu	0.09–0.9	51	1–10
Mn		<29	<6
Mo		95	1–10
Ni	0.95–9.5	<40	0.2–2
Si	2.0 –20	<12	<1.1
Ti		<22	<1.1
V		61	75
C	14	11	16
P	2.0	40	51
S	5.2	<5.6	95
H ₂	8.8	304	210
O ₂	2.0	20	328
N ₂			
Totals:			
Minimum including oxygen	40.99	672	777
Minimum without oxygen	32.19	368	567
Maximum including oxygen	71.50	940	807
Maximum without oxygen	62.7	636	597

* Al, Ca, Cr, Cu, Mn, Mo, Ni, Si, Ti, V analyzed by emission spectrometry (semiquantitative); C, P, S, H₂, O₂, N₂ analyzed by quantitative analysis.

Table 2 summarizes magnetic properties of various grades of iron. Saturation magnetization (≈ 2.15 T) is not strongly influenced by purity, while coercivity H_c and achievable magnetic permeability do strongly depend on purity and crystallographic features. Values of initial and maximum permeability drop for cold worked material. In order to restore magnetic properties, annealing cycles are required, allowing internal strains to be reduced, grain size to be increased, as well as the annealing of dislocations. Iron has various phases with different stability domains: α - and δ -iron, corresponding to the ferromagnetic ferritic phase of bcc structure, which are present up to

912°C ($T_{\alpha,\gamma}$) and in the ranges between 1394°C and 1538°C , respectively, and γ -iron, corresponding to the austenitic phase of fcc structure, in the range between 912°C and 1394°C . This phase is non-ferromagnetic. For this reason, there are two classes of anneals used commercially [16]:

- 1) Anneals below 900°C
- 2) Anneals at or about 925°C or higher to promote grain growth and to further improve magnetic properties.

These anneals, particularly the high T ones, should be followed by slow cool. Higher maximum permeability is obtained by exceeding $T_{\alpha,\gamma}$, allowing the material to enter the γ stability domain and subsequently revert γ by slow cooling. For high maximum permeability, annealing should be performed at between 925°C and 1000°C (above $T_{\alpha,\gamma}$) followed by cooling at a rate $< 5^{\circ}\text{C}/\text{min}$. For high permeability at $B \geq 1.2 \text{ T}$, it is advisable to anneal at a maximum of 800°C and to cool slowly [17].

Table 2: Magnetic properties of various grades of iron (from Ref. [14]).

Material	$H_c (\text{A m}^{-1})$	$\mu_i (\mu_0)$	$\mu_{\max} (\mu_0)$
Ingot (99.8% Fe)	112	10	1 000
Armco	80	200	7 000
Commercially pure	20–100	200–500	3 500–20 000
Carbonyl iron powder	6	3 000	20 000
Vacuum-melted	25	—	21 000
Electrolytic	7	1 000	26 000
Electrolytic annealed	18	—	41 500
Vacuum-smelted and hydrogen-annealed	3	—	88 400
Purified Armco (99.95% Fe)	4	10 000	227 000
Vacuum-annealed	—	14 000	280 000
Single-crystal	—	—	680 000
Single-crystal, magnetically annealed	12	—	1 430 000

Saturation magnetic polarization $J_s = 2.15 \text{ T}$ at 20°C , except for ingot with $J_s = 2.05 \text{ T}$.

2.4 Low-carbon steels

For applications that require ‘less than superior magnetic properties’ [4], low-C steels are frequently used, including in magnet construction where in several cases they are purchased to magnetic specification. One of the most common grades is the structural-constructional steel 1010¹ used for the VINCY cyclotron, in the OPERA and ATLAS experiments, and recently proposed for the CLIC main beam quadrupole prototypes. Owing to large ranges and allowed impurity contents, the composition of a specific grade of low-C steel is not sufficiently reproducible between different producers and heats to closely guarantee magnetic properties. For the 1010 steel, Si varies in different possible content ranges depending even on the form of the product (bars, rods, etc.). This explains the large spread in magnetization curves for different heats of the same grade of steel (Fig. 3). The recommended magnetic annealing cycle for this steel is at $815^{\circ}\text{C} \leq T \leq 980^{\circ}\text{C}$, for a duration of between 1 h and 6 h followed by furnace cooling [4]. As for pure Fe, lower values of T are intended for stress relief, and higher range for full annealing. During cooling, particularly in the critical temperature range (between 849°C and 682°C), slower cooling rates should be applied (for 1010 steel at a rate of $28^{\circ}\text{C}/\text{h}$).

Contrary to high-purity irons (for C concentrations less than 20 ppm), low-C steels are subject to magnetic ageing. An increase of coercivity occurs with time, due to formation of cementite precipitates giving rise to domain wall pinning. For magnetic cores that may operate at between 50°C

¹ Roughly equivalent to European grades 1.1121, 1.0301, 1.0308, 1.0032.

and 100°C, ageing can be an issue. Ageing is mainly due to C, but also to N (due to formation of AlN precipitates), and S. In order to avoid ageing, C, N, S should be reduced below the range 20 ppm to 30 ppm. This reduction is possible by degassing of the melt followed by a final purification of the steel under pure hydrogen, at a T as high as 1475°C. Indeed, at high T , H₂ reacts with the C present in the steel through the reaction Fe + C + 2H₂ = Fe + CH₄ and with other impurities such as oxygen, sulfur, nitrogen by forming H₂O, H₂S, NH₃, respectively, thus reducing their content under critical concentrations.

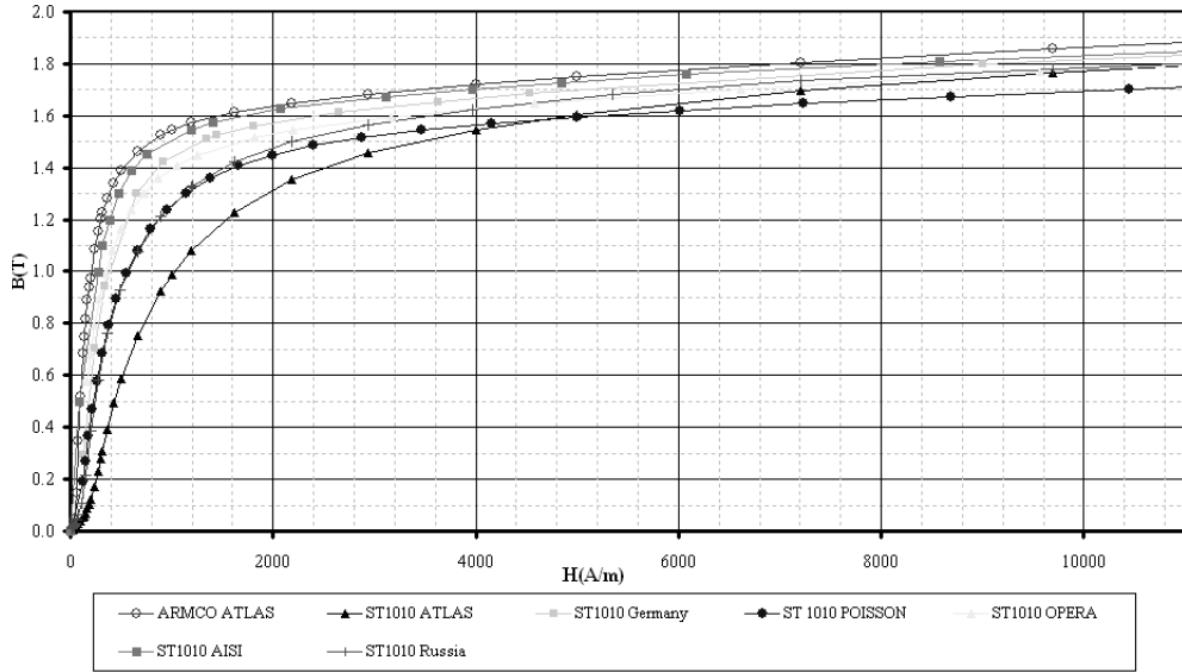


Fig. 3: Magnetization curves of different heats of 1010 constructional steels used as magnetic steel in different experiments (courtesy A. Vorozhtsov)

2.5 Non-oriented silicon steels

The Fe-Si alloys (phase diagram in Fig. 4) were accidentally discovered by Hadfield in 1882 (1.5% Si content). Magnetic properties were reported in 1900. Properties of earlier alloys (Hopkinson, 1885) were hindered by an excessive C content. Industrial production started in Germany in 1903 (2–2.5% Si). Their commercial use started in the US in 1905 and in England in 1906 [3].

Alloying with Si allows for an increase in permeability and decrease in hysteresis loss. Also thanks to additions of Al and Mn, eddy current losses decrease due to higher resistivity. With Al addition that reacts with N to form AlN, no ageing is experienced by silicon steels. On the other hand, with respect to pure Fe, saturation magnetization decreases with increasing Si content (2 T for 3.5% Si).

Silicon steels, also called electrical steels, are industrially produced in casting-hot rolling lines. Hot-rolled strips are subsequently pickled, cold rolled, continuous annealed (annealed products are called ‘fully processed’), coated and slotted on line to the required width [19]. Coatings play an important role for the adhesive bonding of magnet laminations. They provide electrical insulation in addition to mechanical bonding. Coatings are designated according to the IEC 60404-1-1 standard [20], can be organic, inorganic with organic components, can be applied on one or both sides. Advanced coatings based on active organic bonder lacquers such as STABOLIT 70® require a delicate curing operation at a later stage, after stacking of the laminations. This coating can reach a mean shear strength above 22 MPa when cured at 190°C for 15 min [21], which stays above 20 MPa

for doses up to 10^8 rad. Fully processed non-oriented silicon steels can be easily specified according to EN 10106 [22].

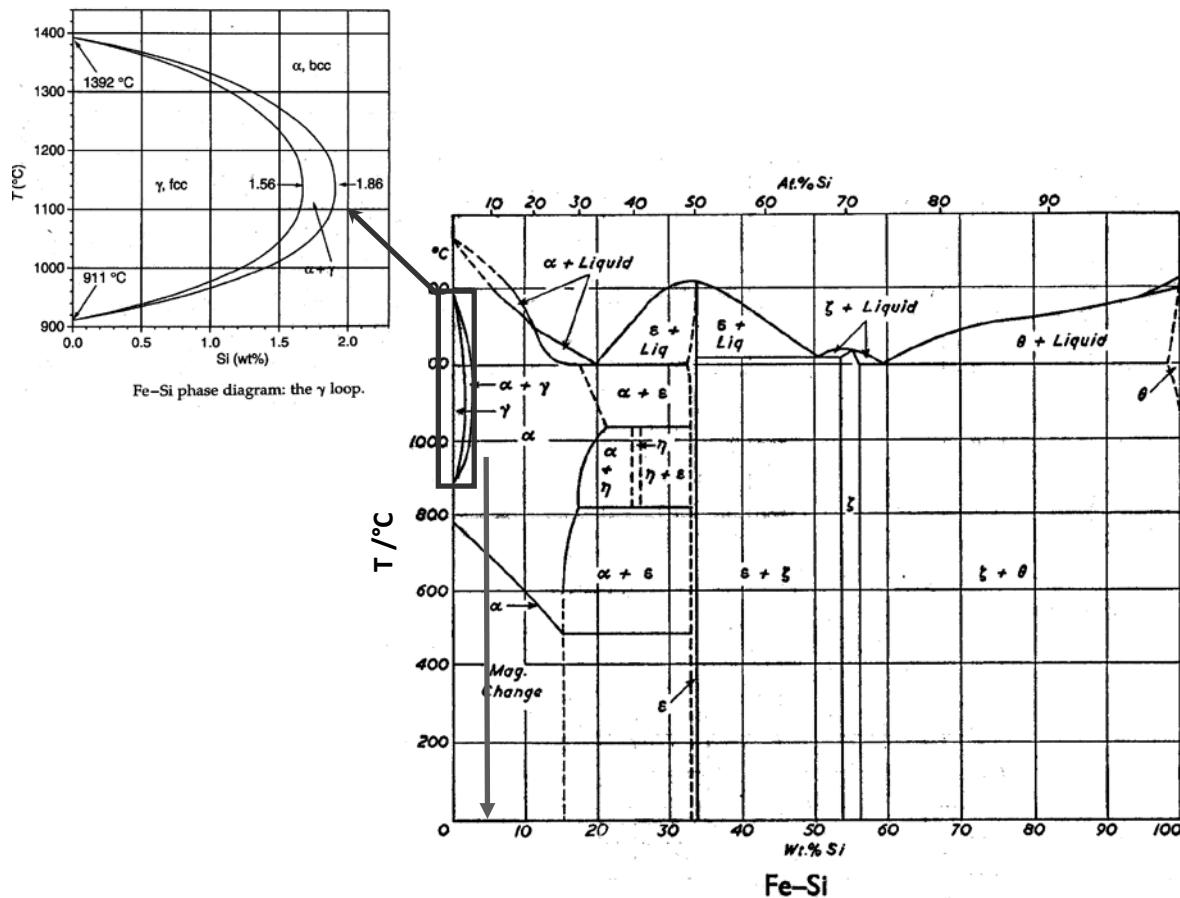


Fig. 4: Phase diagram of the Fe-Si system (from Ref. [18]) and detail of the austenitic ‘ γ loop’ (from Ref. [14]). Si content of 3.5% (vertical arrow) represents an upper industrial limit (limited ductility for higher contents).

2.6 Oriented silicon steels

Iron single crystals exhibit minimum coercivity and maximum permeability when magnetized along one of the $<001>$ axes. Fe-Si is also most easily magnetized in this direction. The so-called ‘Goss’ texture $(110)[001]$ (Fig. 5) can be developed in silicon steels by a controlled sequence of cold rolling and annealing steps.

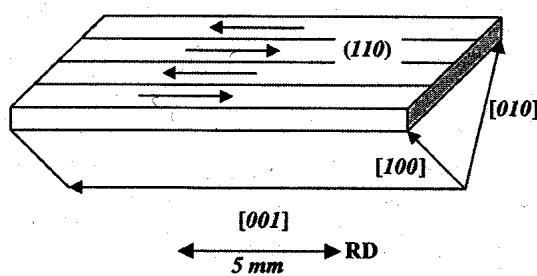


Fig. 5: Prevalent orientation of grain crystal axes with respect to rolling direction (RD) in grain oriented sheets of silicon steels (from Ref. [14])

Control of the texture, achievement of a large grain size and low impurity content will allow coercive fields as low as 4 A/m to 10 A/m and maximum permeability around 5×10^4 to be achieved in grain-oriented (GO) alloys. These figures are approximately 10 times higher than in non-grain-oriented (NGO) silicon steels. Conventional GO (7° dispersion of the [001] axes around RD) sheets represent 80% of the market, 1 M ton/y and some 1500 M EUR/y. High permeability GO (HGO, 3° dispersion) steels can achieve even higher properties. Since very large domains are detrimental, scribing the sheet surface (a series of parallel lines arrayed perpendicular to RD, spaced a few mm apart, obtained through mechanical scratching or laser irradiation) allows multiplication of domains oriented along the [001] axis in HGO sheets [14]. Surface coatings capable of exerting a tensile stress of 2 MPa to 10 MPa improve magnetic performance further.

The industrial processing of oriented silicon steels includes several steps [23]: starting from hot-rolled strip, surface descaling is performed in shot-blasting and pickling lines. Heavy cold rolling (CR) is followed by an intermediate annealing and final cold rolling (strips may be cold rolled twice on special cold rolling mills with an intermediate annealing in a continuous annealing furnace). Total CR exceeds 50%. A decarburization step is performed on strips that are coated with an annealing separator (magnesium oxide) to prevent the windings of the coiled strips from adhering to each other during subsequent high-temperature annealing.

A further high-temperature box annealing at a T up to 1200°C for a treatment lasting 5 d to 7 d under a protective atmosphere, allows Goss-texture to be developed. Individual grains up to 5 mm and 20 mm can be grown. Material, further refined by diffusion annealing, is stress relief annealed and coated with insulation and thermally flattened. Stress relief annealing is performed in continuous annealing furnaces. Final steps are side trimming and slitting.

Power losses due to eddy currents can be minimized by reducing the sheet thickness d . Indeed, the ‘classical’ eddy current power losses are proportional to d^2 for a constant permeability and complete flux penetration. Nevertheless, for very thin sheets the domain wall spacing is reduced. The presence of concentrated electric fields at domain boundaries and of supplementary domain structure in thin sheets, limit the beneficial effect of reducing sheet thickness under a few tenths of mm [24,25]. For this reason, the industrial thinner sheets usually have $d = 0.23$ mm.

2.7 Fe-Ni alloys

This family of alloys, before their use as magnetic materials, was already known for the low thermal expansion coefficient of Invar, an Fe-36%Ni alloy, discovered by C. E. Guillaume in 1896 [26]. The ferromagnetic γ phase can be retained by a suitable choice of annealing T , cooling rates, addition of other alloying elements such as Mo, Cu, Cr. For Ni contents above 35%, the $\gamma \rightarrow \alpha$ transition still exists, but it occurs at $T < 500$ °C and is therefore limited because of low diffusion rates (Fig. 6). Two main families of alloys for magnetic applications have been developed: the ‘low-Ni’ FeNi alloys, containing 47% to 50% of Ni, featuring higher saturation fields (1.6 T) and maximum permeabilities up to 60 000, and the ‘high Ni’ alloys including mumetal and containing approximately 80% of Ni, with lower saturation fields (0.8 T) but higher maximum permeability (up to 800 000).

High-Ni alloys were the first to find commercial applications: Elmen’s work in 1913 was aimed at finding a material superior to Si-steels for ‘use in telephone apparatus operating at less than a few hundred gausses’. Low-Ni alloys are prepared as strongly textured (110)[001] sheets by means of severe CR and annealing at $T \approx 1000$ °C. Magnetic annealing is usually applied. These alloys, and in particular 45 Permalloy, also found early applications in telephone apparatus, since they featured higher saturation than any of the other permalloys and could be operated at higher induction.

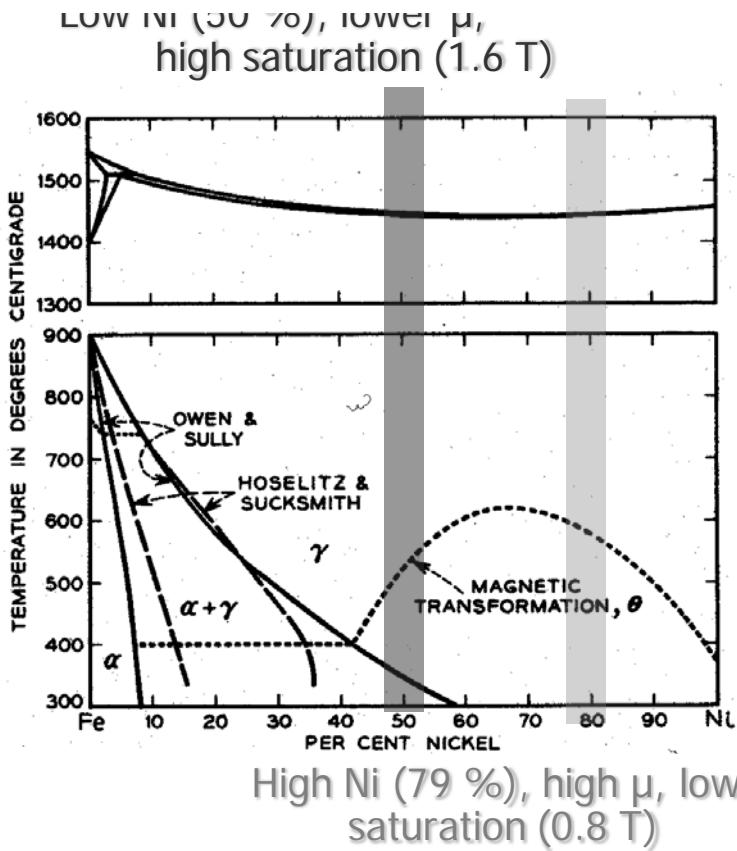


Fig. 6: Fe-Ni phase diagram (from Bozorth [3]), showing the two main compositions of alloys of this family for magnetic applications

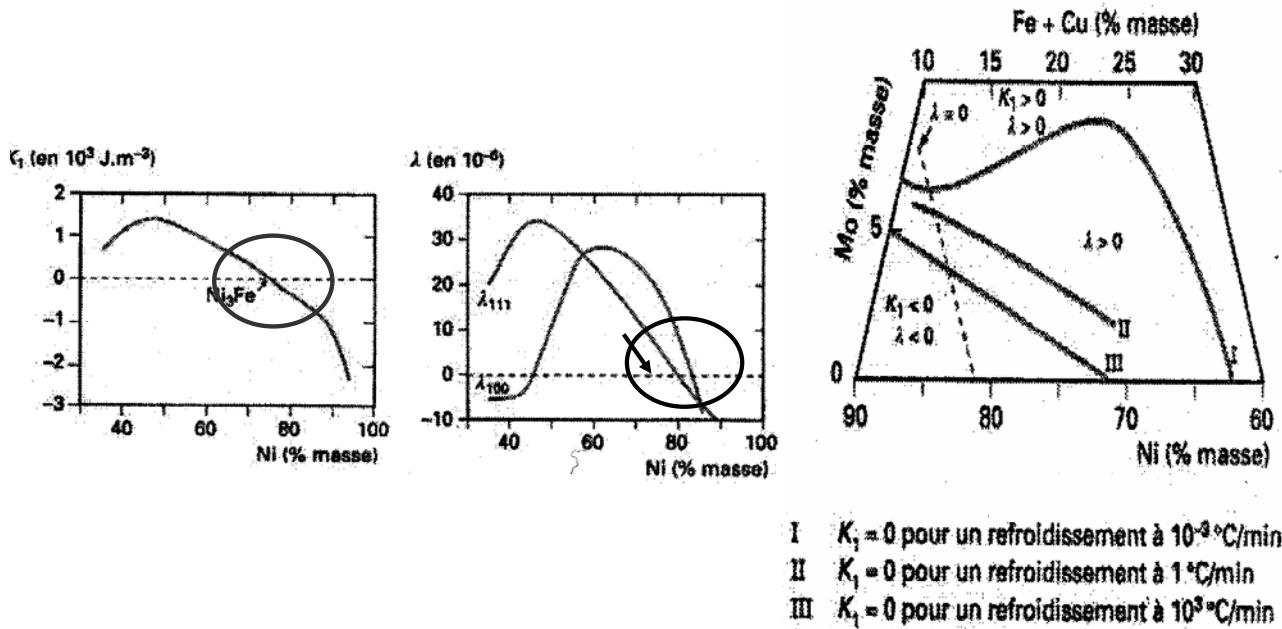


Fig. 7: Between 65% and 85% Ni crystalline anisotropy K_1 (the relevant coefficient of the first order expansion of the crystal anisotropy energy for a given angle of misorientation [6]) vanishes and magnetostriction λ as well, but not simultaneously. Addition of other elements (Mo, Cu, Cr, etc.) allows the two parameters to be reduced simultaneously and a very high permeability to be achieved (from Ref. [27]).

As already mentioned, the highest permeability in this alloy family is featured by alloys containing approximately 80% of Ni. Indeed, between 65% and 85% Ni crystalline anisotropy vanishes and magnetostriction as well (but not simultaneously, see Fig. 7). The addition of other elements (Mo, Cu, Cr, etc.) reduces simultaneously the two parameters, while increasing the resistivity (relevant for ac applications).

Vacuum chambers for the circulating beams in the LHC injection and extraction septa [28] were manufactured from a 77Ni-5Cu-4Mo-Fe mumetal [29] produced by Imphy /FR. Because of very limited solubility of N in mumetal, mumetal tubes should be welded to AISI 304L or AISI 316L flanges depending on the application, in order to avoid gross porosity unavoidable in mumetal/AISI 316LN autogeneous welds [30].

Internal stresses, plastic deformation induced during forming and welding all degrade the magnetic performance of mumetal (Fig. 8). A high-temperature annealing treatment allows restoration of magnetic properties. Nevertheless, in order to achieve the best magnetic properties of mumetal (max. permeability as high as 100 000, coercivity as low as 0.05 Oe), the final annealing treatment should be performed under hydrogen at 1120°C.

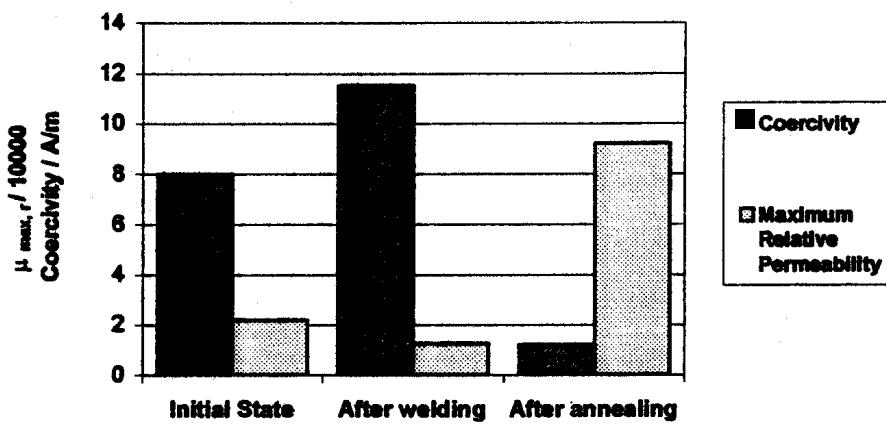


Fig. 8: Evolution of coercivity and maximum permeability of mumetal with fabrication steps. Sample machined from the as-delivered material ('initial state'), sample containing TIG welds ('after welding'), idem after annealing under vacuum at 1070°C for 1 h ('after annealing'). Coercivity, which was 8.0 A/m in the initial state, increased due to the presence of the welds (11.5 A/m) and was considerably reduced by annealing (1.2 A/m). Permeability, approximately 21 800 in the as delivered state, decreased about 60% due to the presence of the welds, to a value of around 12 570. After the annealing treatment, the permeability reached 92 000 [30]. Material delivered by Telcon Ltd. /UK.

For some Fe-Ni alloys, particular care has to be paid to cooling conditions from an 'ordering temperature' ($\approx 500^\circ\text{C}$). Heat treatment in a magnetic field, discovered by Kelsall in 1934 [31] can cause a large increase in permeability ($\times 10$) of some Fe-Ni alloys. Taking as an example 65 Permalloy, the presence of the field during cooling from T_c to 400°C is essential. Magnetic annealing is based on non-random diffusion of atoms and preferential alignment of like-atom pairs. It is explained on the basis of a 'directional-order' theory [32–34]. At a temperature $T \leq T_c$, but high enough for diffusion to occur, like-atom pairs tend to be aligned in the direction of the local magnetization. As temperature is lowered during cooling, since diffusion constants become too low for further diffusion to occur, the 'freezing' in place of like-atom pairs produces a uniaxial anisotropy in the material [35].

2.8 Compressed powdered iron and iron alloys

Iron-based products produced by powder metallurgy (PM) techniques have the advantage of featuring isotropic 3-D properties. For application in the medium frequency (kHz) range, cores of compacted iron are often used. PM products can be near-net shaped to close dimensional tolerances and show satisfactory T stability due to limited internal stresses. Designers can exploit 3-D flux paths [36]. For some components (claw pole, brushless motor) use of laminations would not be applicable.

Coated iron powders (of a typical size 50 μm to 100 μm) are mixed with some 1% of binding material, compressed as cores of the desired shape and then sintered. The cores are coated by protective painting. Fe, Fe-Si and other Fe powders can be compacted. The maximum achievable permeability is controlled by grain size, sintering T and degree of porosity. The dc properties of hot-pressed high-purity Fe are considered as good as or better than conventional Fe. Electromagnetic actuators of complex core shapes for use in transport, electrical rotating machines are produced through this technology. The usual field of compressed powdered irons are dc and medium frequency applications (ferrites dominate in very high frequencies). Two components of core losses are identified in powder metallurgy irons: a classical hysteresis contribution and an eddy current one [37]. Powder metallurgy cores exhibit larger hysteresis contributions than steel sheets, but (particle to particle) eddy current contributions are much smaller already at 60 Hz (estimated at 5% of the total in the example of Fig. 8). Eddy currents become dominant at higher frequencies, where, due to lower eddy current contributions, pressed materials are competitive in performance with laminated steels (Fig. 9). In powder materials, two kinds of eddy current are identified, circulating within the insulating particles and around clusters of particles. This explains size effect contributions to the total losses [38], and the advantage of the use of insulating coatings on powder particles, respectively.

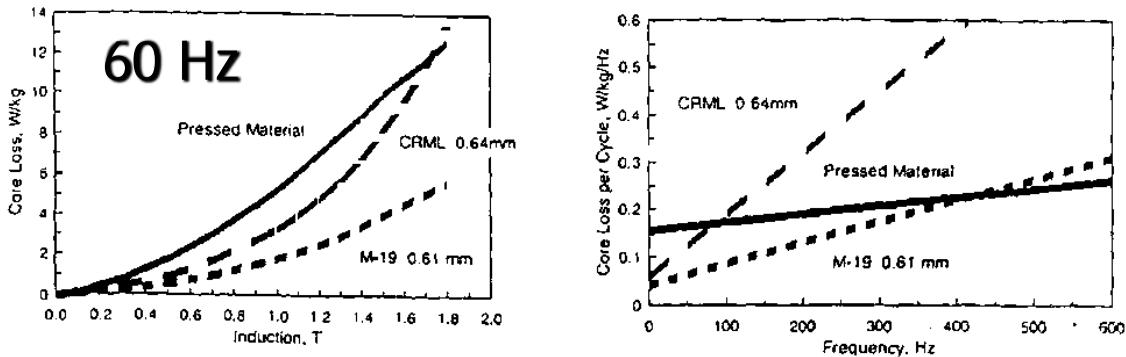


Fig. 9: Comparison of a pressed material with 0.64 mm thick ‘Cold Rolled Motor Lamination’ (CRML) steel and 0.061 mm thick ‘M-19’ NGO silicon steel. At 60 Hz, the total core loss of the pressed material is dominated by hysteresis contributions, while the eddy current contribution is estimated at only 5% of the total losses. The pressed material has lower performance than conventional steels. At higher frequencies, the total loss for the pressed material is less than for laminated steels (from Ref. [37]).

2.9 Soft spinel ferrites

Ferrites are largely applied in the high-frequency range up to few hundred MHz. Their composition is $\text{MO} \bullet \text{Fe}_2\text{O}_3$ where M is a divalent metal ($\text{M} = \text{Fe}^{2+}, \text{Mg}^{2+}, \text{Mn}^{2+}, \text{Ni}^{2+}, \text{Zn}^{2+}$; lodestone $\text{FeO} \bullet \text{Fe}_2\text{O}_3 = \text{Fe}_3\text{O}_4$ is a particular case of a ferrite). They are ceramics featuring very high resistivity, between $10^6 \Omega \text{ m}$ and $10^{12} \Omega \text{ m}$. They are used at frequencies where eddy current losses for metals become excessive. At very high frequencies, they are ideal soft magnetic materials. Disadvantages are low magnetic saturation (typical range 0.15 T to 0.6 T), low T_c (330°C to 585°C) poor mechanical

properties, hardness and brittleness. Since they are practically unmachinable, close dimensional tolerances are achieved by grinding.

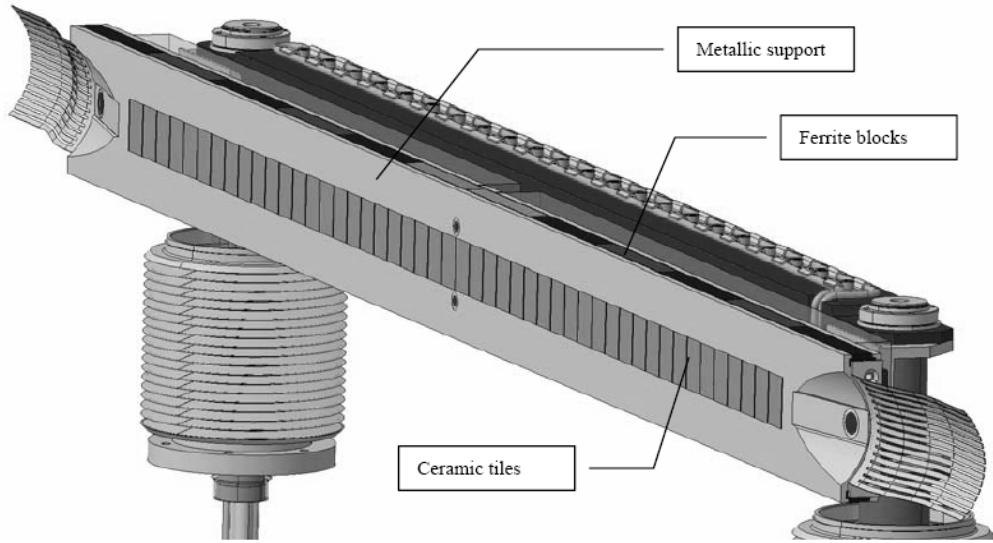


Fig. 10: Collimator jaw equipped with BPM buttons and ferrite blocks (courtesy of A. Dallocchio)

Ferrites are used as cores for electromagnetic interference suppression, to control transmission or adsorption of electromagnetic waves. One example of the application of ferrites in accelerators is the LHC collimators, where ferrite blocks are used to damp trapped modes (Fig. 10).

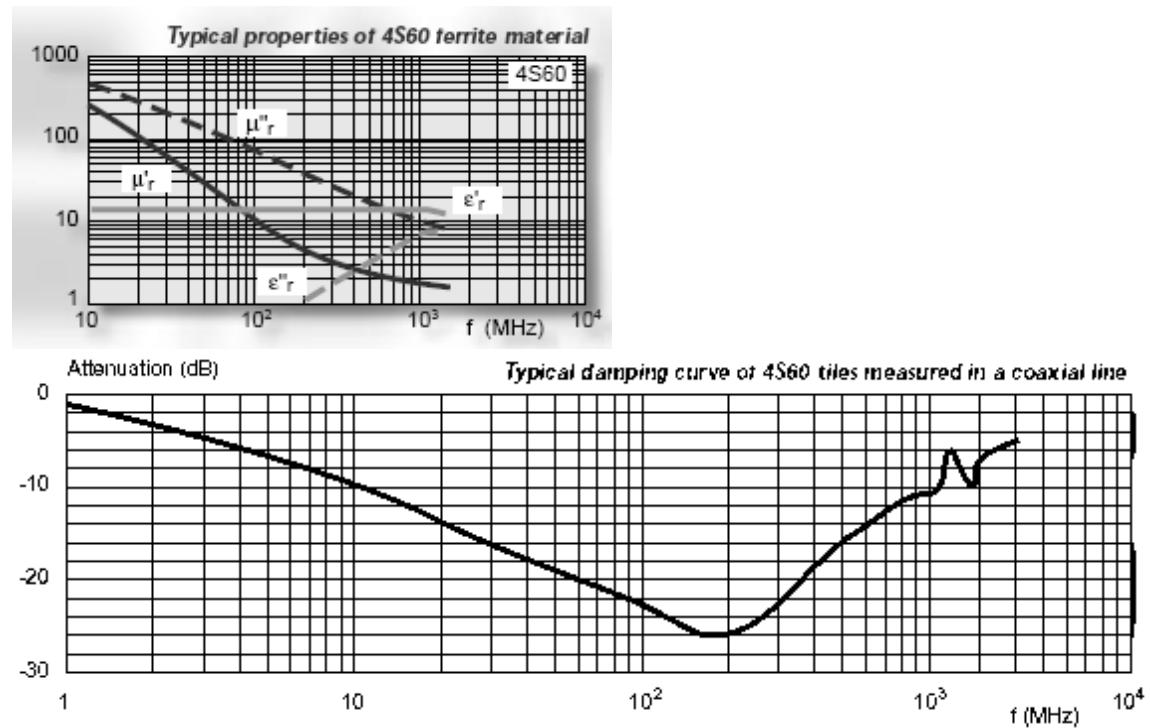


Fig. 11: Properties of FERROXCUBE 4S60 ferrite. Maximum absorption occurs where μ_r matches the dielectric constant ϵ_r (from Ref. [39])

The ferrite FERROXCUBE 4S60 is used for the TCLIA collimators in the LHC at CERN. This ‘NiZn’ ferrite is of a mixed type. The proportions of Ni and Zn are tailored for the specific application. While MnZn ferrites show the highest permeability, NiZn, shows a broadband operation up to 1000 MHz and features higher resistivity (Fig. 11).

Ferrites are prepared through a powder metallurgy process. After mixing and weighing the base oxides in the form of fine powders, the powder is heated to between 900°C and 1200°C (pre-firing in air) to produce flakes of a few cm³. During this stage, the spinel structure is formed by the reaction of Fe₂O₃ with MO. This material is ball milled in water and mixed with a binder. A spray drying step results in balls of few mm³. The powder is then mechanically pressed in a mould by die punching or hydrostatic pressing. The green (filling factor 50% to 60%) is batch sintered in a kiln at T between 1200°C and 1400°C, with or without external pressure, in an oxidizing atmosphere (filling factor 95% to 98%). Shape is conferred by a final grinding [14,40].

2.10 Innovative materials: amorphous alloys

Amorphous alloys (metallic glasses) are materials devoid of long-range atomic order. They are produced by rapid solidification from the liquid or gaseous state. Generally they are in ribbon form and obtained by rapid solidification from the liquid state. Metastable sputtered amorphous thin-films of Co-Au were produced by Mader and Nowick in 1965 [41]. Today, ribbons up to 100 mm or 200 mm wide of a thickness of 10 µm to 40 µm can be produced. In order to develop the amorphous structure, cooling rates of 10⁵°C/s to 10⁶°C/s have to be obtained. Amorphous wires are also produced. The preparation of amorphous magnetic materials by rapid quenching from the melt is generally performed through planar flow casting on metallic wheels or drums in air, shielding with atmosphere or vacuum (Fig. 12).

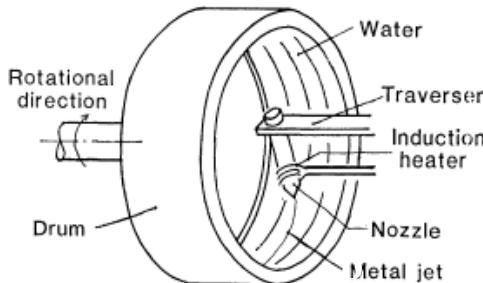


Fig. 12: Rotating drum. A jet of molten metal is ejected into a rotating water layer (from Ref. [42]). In order to guarantee rapid quenching, drum velocities are as high as 3 m/s to 10 m/s. Reported range of peripheral velocities of metallic wheels for the production is 10 m/s to 40 m/s [14].

Amorphous alloys are generally of the type T₇₀₋₈₀M₃₀₋₂₀, where T is a transition metal (Fe, Co, Ni) and M is a combination of metalloids (B, Si, P, C, etc.). Amorphous ribbons of 25 µm thickness in Fe₇₈B₁₃Si₉, compared to a conventional GO Fe-3Si of 0.23 mm of thickness, show a coercive field of 2 A/m after annealing (5 A/m for the latter) and a maximum relative permeability of 2 10⁵ (8·10⁴, idem). Annealing allows for a local crystallization. These materials are hard and show ductility limited to some 2.5% elongation at failure. Alloy glass ribbons have an excellent magnetic softness. Magnetization curves have an ideal loop with a precise ‘cut’ at a specific value of B (see the example of the loop of a commercial Vitrovac material in Fig. 13a). This behaviour, combined with a high magnetomechanical coupling, makes them ideal materials for sensors for electronic article surveillance in the so-called ‘harmonic-electromagnetic systems’: glass ribbons are used for security tags in libraries and stores [43, 44]. In order to be conveniently activable and deactivable and not to be demagnetized by the fields in the interrogation zone, materials should show H_c between 1600 A/m

and 8000 A/m [43]. Permalloys could also be used for the purpose illustrated in Fig. 13, but their ductility would imply a risk of degradation of their magnetic properties by handling.

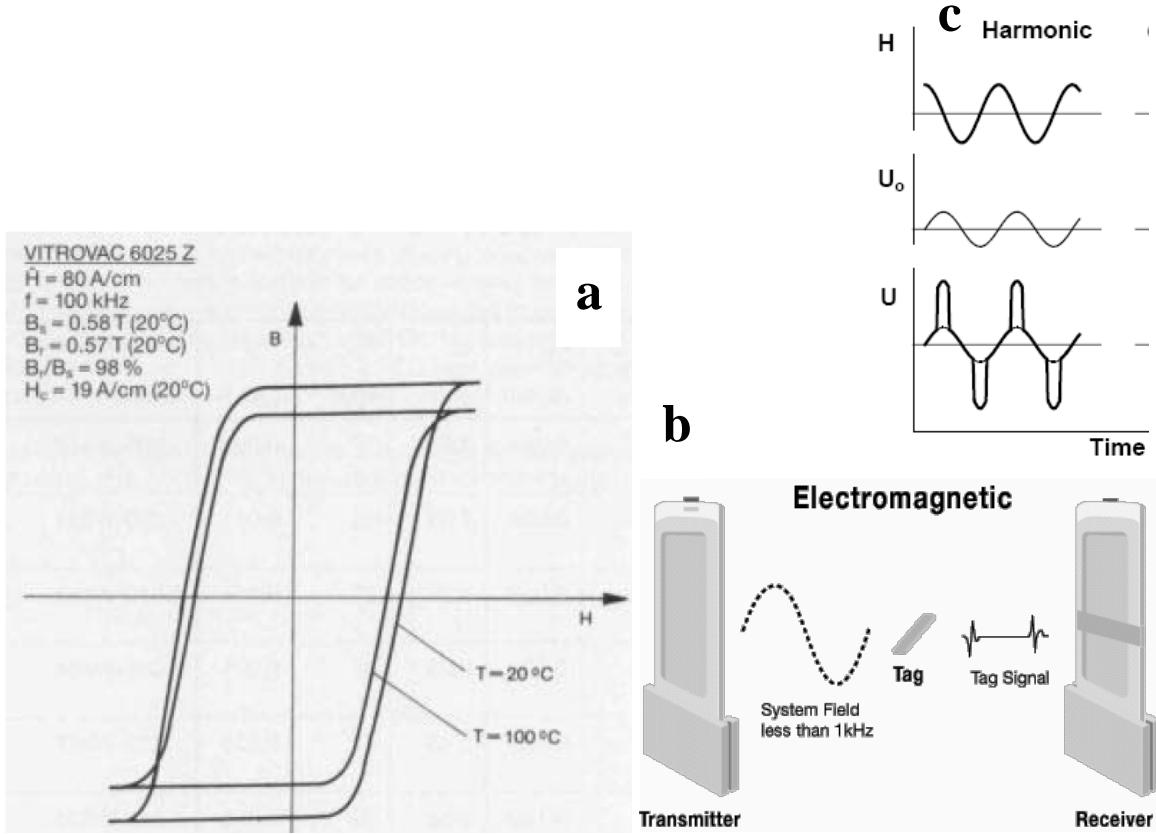


Fig. 13: a) Ideal loop of a commercial amorphous material (from Vacuumschmelze [45]). When applied in security tags, a length of amorphous alloy is packaged together with a hard magnet strip. The tag is activated by unmagnetizing the permanent magnet, and deactivated at the cash register by magnetizing (saturating) the strip. b) Activated labels respond to an electromagnetic field generated by a pedestal at the store exit with specific frequencies detected by the receiver, since the material is periodically driven into saturation. c) Response to a periodic excitation H : U_0 , unactivated label, U activated label: high permeability at fields below saturation induces high attenuation when the label is activated, while above saturation spikes are present during the transmission time (from G. Herzer [43]).

2.11 Innovative materials: nanocrystalline alloys

Nanocrystalline alloys consist of small ferromagnetic crystallites of bcc FeSi with grains of 10 nm to 15 nm embedded in an amorphous matrix, coupled to each other by exchange interaction. Crystallites are separated by 1–2 nm for interaction. These materials show very low coercivity (0.4 A/m to 8 A/m), high initial permeability (up to 150 000), low losses and magnetostriction and high saturation up to 1.3 T. They are available in ribbons of few tens of μm thickness. Compared to conventional Fe-Si steels and amorphous materials, they show exceptionally low core losses, specially at very high magnetization rates (Fig. 14). For this reason, they are envisaged for application to heavy-ion inertial fusion-energy, based on induction accelerators where some 30 000 tons of magnetic material are necessary for induction cores [46]. For this application, nanocrystalline materials would show the best performance in terms of core losses at the required magnetization rates between 10^5 T/s and 10^7 T/s .

They are commercially applied in low-loss high-frequency transformers, such as the one foreseen by the Linac group of the Spallation Neutron Source (SNS) of the Los Alamos National Laboratory for application in high-frequency polyphase resonant converters for the ILC (International Linear Collider) [47]. These materials are interesting for both high-energy physics and accelerator applications, but a drawback is their cost (between \$20 and \$150 per kilogram).

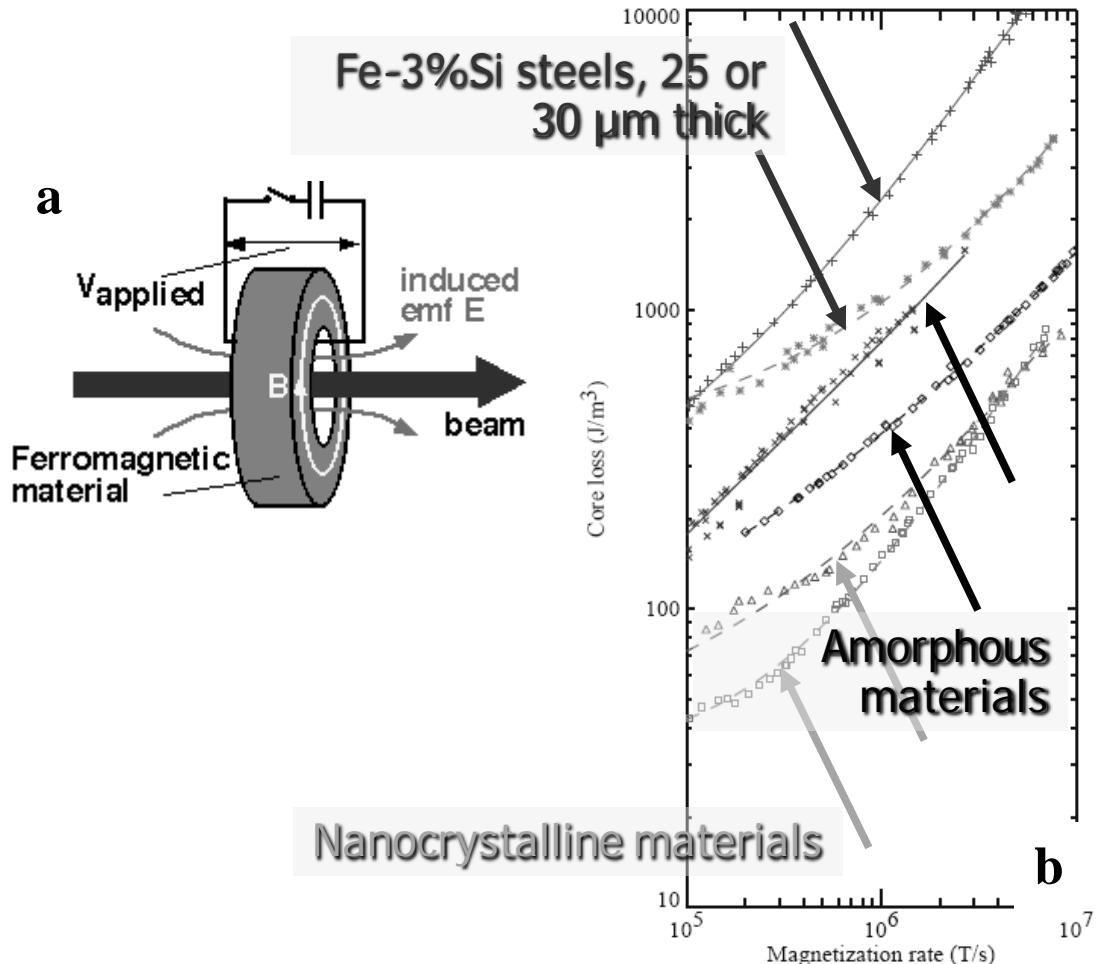


Fig. 14: a) In heavy-ion inertial fusion-energy accelerators some 30 000 tons of magnetic material are necessary for induction cores in order to accelerate heavy-ion energies in the GeV range and deliver several MJ per pulse to a target (from Ref. [46]). b) Nanocrystalline materials are the highest performing ferromagnetic materials as induction core alloys. They form the lowest loss group at the very high magnetization rates required for this application (from Ref. [48]).

3 Methods of measurement

3.1 Characterization of soft magnetic materials

Characterization of the magnetic properties of soft magnetic materials is generally based on the measurement of a transient voltage induced on a secondary winding by a step-like field variation applied on a primary winding. The two coils are generally wound together on a toroid sample. The signal is integrated over a time interval for complete decay of eddy currents, since every recorded

point should correspond to a stable microscopic configuration of the system. Since flux variations are measured, providing a discrete sequence of field values, a reference condition is needed, that is generally the saturated or demagnetized state. A ‘split-coil’ permeameter (Fig. 15) is available at CERN, facilitating the exchange of samples, and not requiring a prior winding contrary to the wound toroid methods [49]. For a careful assessment of the magnetic properties through these methods, a precise measure of the cross sectional areas of the samples is necessary, which should not be estimated from sample mass and density.

The technique of the split-coil permeameter has recently been used for the electromagnetic characterization of the steels used for the OPERA magnets [50].

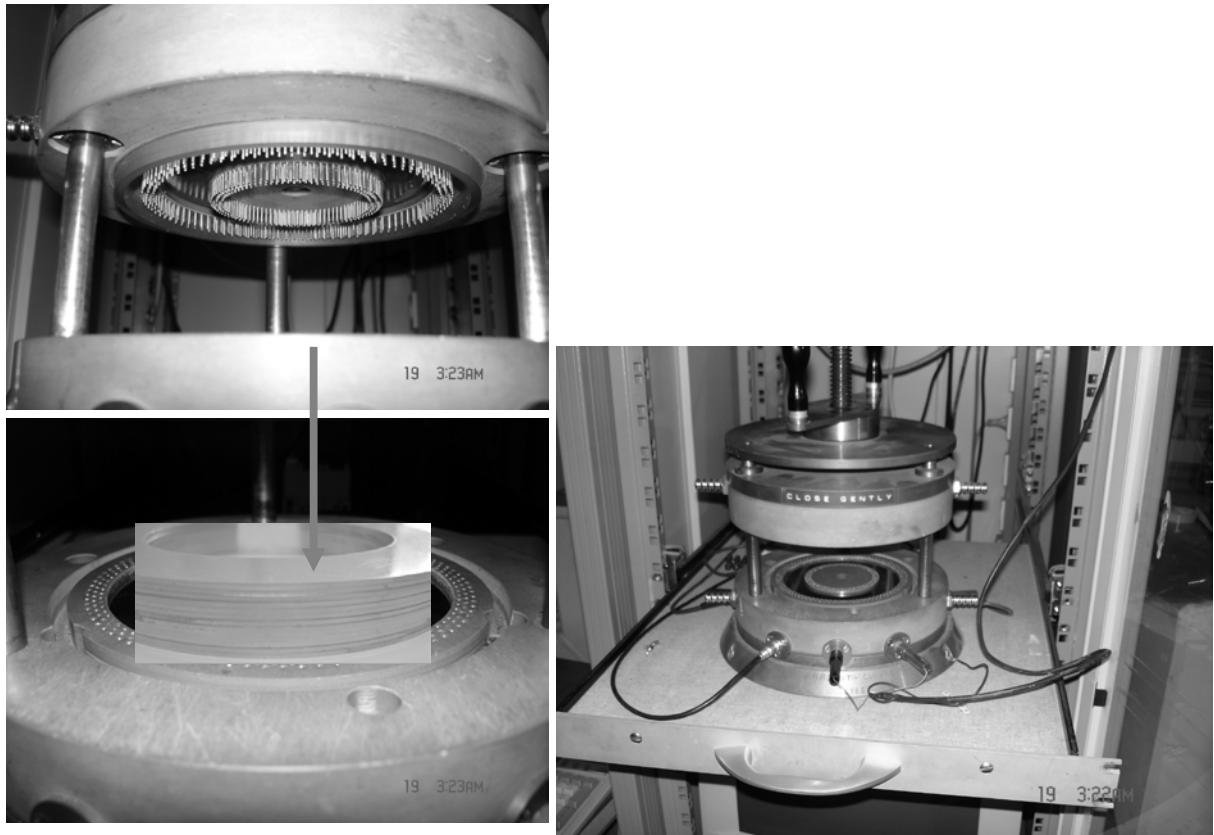


Fig. 15: Split-coil permeameter. The flux measuring and excitation coils are not wound directly onto the samples. The advantage of the system is the rapid exchange of samples (rings, or pile up of rings), an automatic measurement of the relevant parameters of the hysteresis curve, making it adapted to the evaluation of large series. A drawback is the cumulative contact resistance between the two split parts of the coils (two contacts per turn), making this method inadequate for measurements in liquid He (excessive power dissipation).

CERN owns special coercimeters aimed at measuring the coercivity directly on steel sheets without having to cut samples. The instrument was used to perform the coercivity measurements needed during the production of the 11000 t of steel sheets for the LEP dipole magnets and to measure their permeability. The coercimeter is based on a main excitation and detector coil, and auxiliary coils used to estimate the air gap of the yoke contacts (Fig. 16). The yokes are pressed against the sheets that are introduced through a system of rollers in the coercimeter. The sequence of measurements is automatic. After demagnetizing the yokes, a stable hysteresis cycle is assessed. Four flux variations $\Delta\varphi_i$ ($i = 1$ to 4) are measured along the cycle and the remanent flux φ_r is calculated as $\varphi_r = \frac{1}{4} [|\Delta\varphi_1| - |\Delta\varphi_2| + |\Delta\varphi_3| - |\Delta\varphi_4|]$. The current I_c , necessary to cancel φ_r is measured on both sides of

the cycle, allowing the coercive field to be calculated from the mean absolute value of I_c taking also into account the coercive field of the yoke (from Ref. [51]).

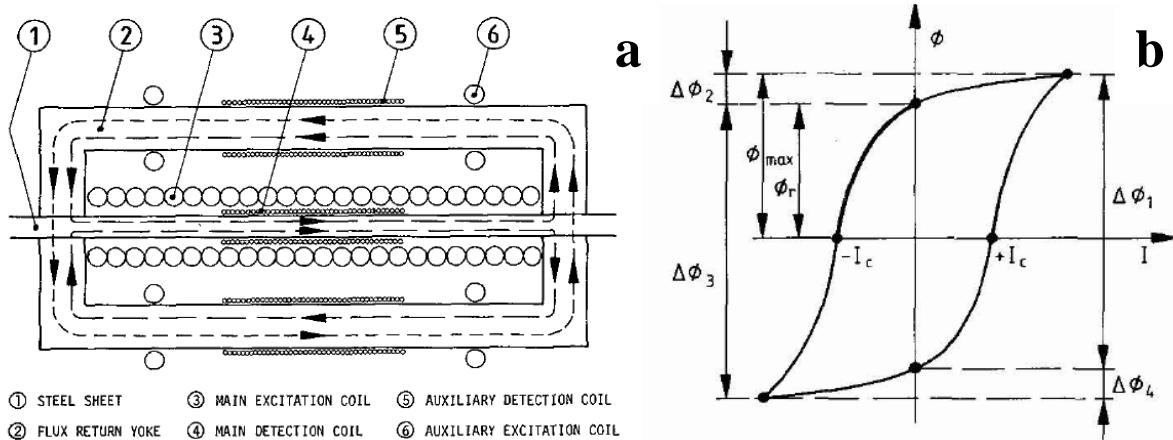


Fig. 16: a) The coercimeter is based on a main excitation and detector coil. b) hysteresis cycle (from Ref. [51])

3.2 Measurement of feebly magnetic materials

Measurement of permeability of feebly magnetic materials (materials with permeability in the range $\mu_r = 1.00001$ to 2 or 4) is possible through several techniques. Portable magnetometers are adapted to non-destructive measurement of materials of complex shape but having a radius of curvature not less than 40 mm or a flat area not less than 20 mm in diameter. The material should be thicker than 8 mm. Measurements of materials thinner than 8 mm can be performed by stacking two pieces or applying corrections. The air gap between two pieces should be as small as possible; otherwise permeability less than actual value will be estimated.

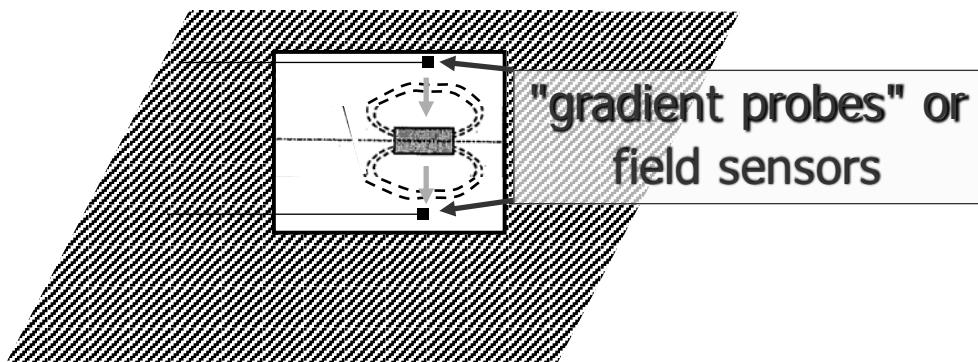


Fig. 17: Gradient probes allow measurement of the displacement of the zero field line of the permanent magnet towards the feebly magnetic material to be measured. The field strength of the probes generally used to assess permeability of austenitic stainless steels is approximately 80 kA/m.

Probes of portable magnetometers are based on a permanent cylindrical magnet (Fig. 17). A permanently built-in field sensor, sensitive to the field emanated from the sample, is placed on either side of the cylindrical magnet in the plane perpendicular to the cylinder axis at the centre of the

permanent magnet. Since the permeability of the feebly magnetic material to be measured is larger than 1, the zero field line of the cylindrical magnet will be displaced towards the sample. This displacement allows the permeability of the material to be assessed [52].

Magnetic balance measurements are suitable for measurement of materials with $\mu < 1.05$. This destructive test method is applicable to the evaluation of semifinished products or welds before fabrication of parts. Samples measurable at CERN are cylinders with a diameter of 3 mm and a height of 2 mm. Larger samples are measurable according to relevant standards. After several reverses of the current to delete the effects of the hysteresis in the core, an increasing current is turned on in an electromagnet, producing step-by-step increasing field strength up to more than 450 kA/m. The sample should not be overheated. The sample is suspended from the balance and positioned above the centreline of the air gap of the electromagnet (Fig. 18). The permeability is calculated from the apparent change in mass of the specimen [53].

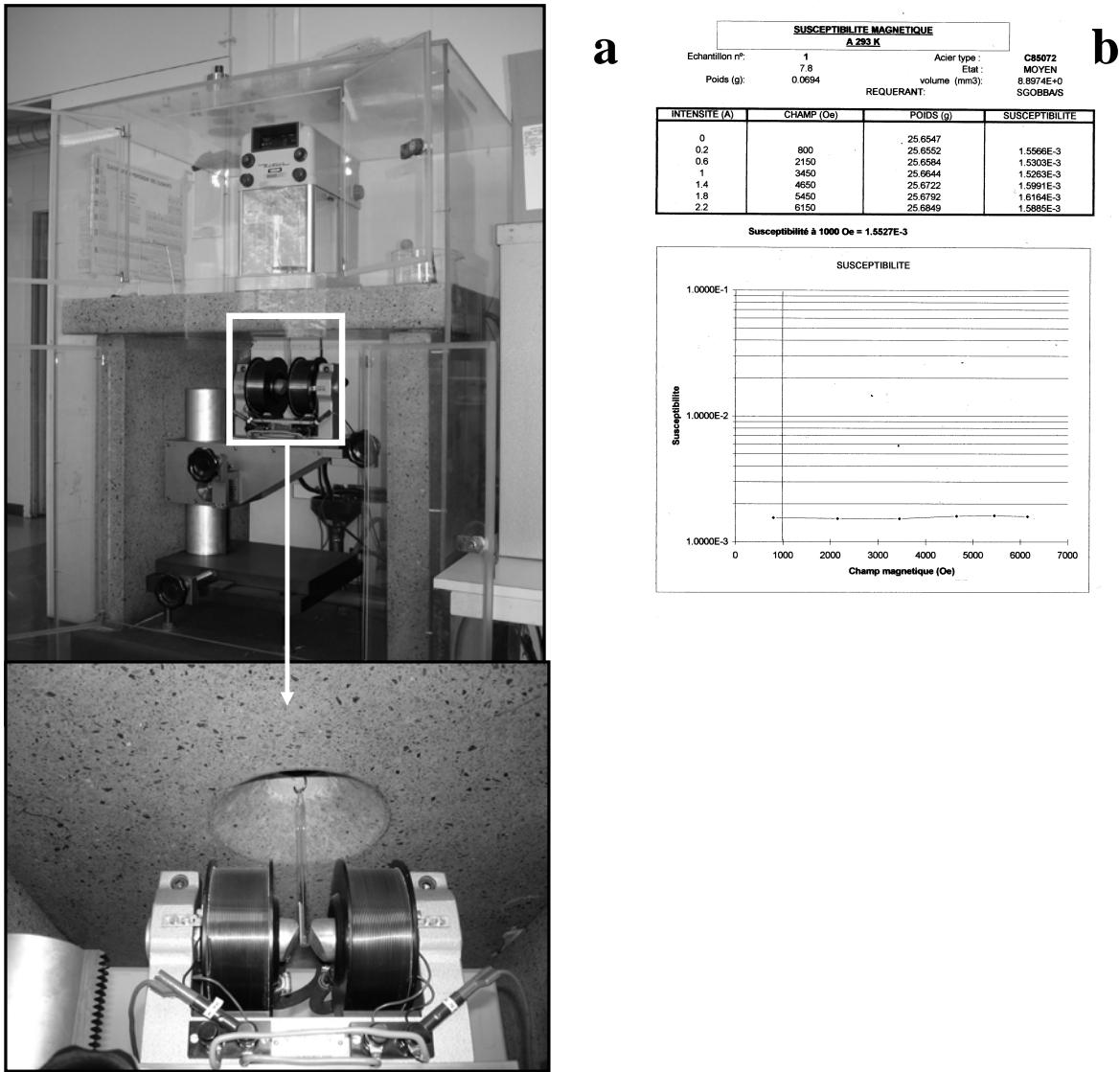


Fig. 18: a) Magnetic balance built up at CERN. b) Magnetic permeability is assessed for different fields from the apparent change of mass of the sample.

Magnetic Field Microscopy (MFM) techniques allow identification of a magnetic contrast between magnetic and non-magnetic phases on a very local scale (Fig. 19). More recently, MOKE (Magneto-Optical Kerr Effect) techniques have been associated to MFM and used to quantify locally the magnetic properties of phases in the samples, including measurement of hysteresis loops [54].

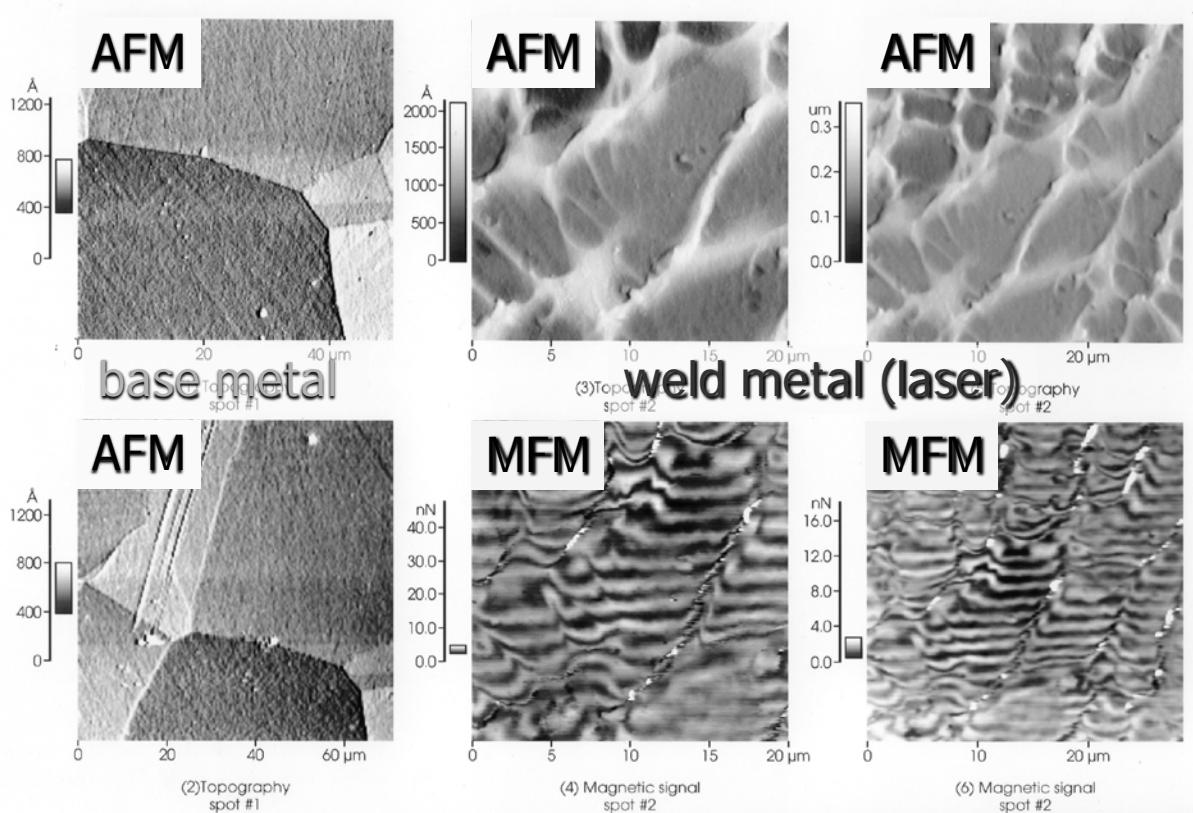


Fig. 19: Atomic Force Microscopy (AFM) and MFM measurements of a sample of an austenitic stainless steel including a longitudinal laser weld, after tensile testing at 4.2 K. While AFM images are only sensitive to the topographical contrast of the surface (grain boundaries and dendrite boundaries are visible in AFM images of the base metal and the weld, respectively), MFM allows qualitative identification of the presence of a magnetic contrast between magnetic and non-magnetic phases possibly present in the weld.

4 Magnetic lag

The two main sources of magnetic lag are discussed: lag due to eddy currents, and the so-called magnetic ‘after-effect’ (Nachwirkung), which is material dependent.

4.1 Lag due to eddy currents

When one applies a magnetizing current to a bar, eddy currents develop whose direction is opposite to that of the applied current. This current flow has both an effect in DC (the field cannot penetrate immediately into the interior of the material) and in AC (set-up of field gradient between surface and interior). Lag due to eddy currents should be taken into account when measuring the magnetization curves of soft magnetic materials (see Section 3.1). Indeed, the sudden application of a field to a cylinder of field-dependent permeability of diametre d requires a time τ to reach a field B :

$$\tau = 0.55 \frac{d^2}{\rho} \int_b^1 \frac{dB}{dH} \frac{db}{b} \quad (5)$$

where $b = 1 - B/B_0$, B_0 is the ultimate field and ρ is the resistivity [3]. For a constant permeability, a variation of field applied to a lamination of relative permeability μ_r , conductivity σ , thickness d implies a decay time of eddy-current-generated counterfield $\tau \propto \mu_0 \mu_r \sigma d^2$ [14]. For a 10 mm sheet of a 1010 steel, τ is approximately 3 s [55].

Undesirable effects due to lag induced by eddy currents were identified in CERN SPS magnets, when measuring in multicycles the influence of the 450 GeV proton cycle on the following positron cycle [56].

4.2 Magnetic ‘after-effect’

The application of a magnetic field requires a given time to reach the final induction value in a magnetic material. Apart from the contribution to the lag due to eddy currents discussed in Section 4.1, additional delayed effects have a metallurgical origin (impurities such as C and N in Fe, dislocations, etc.) and are mainly due to a time-dependent microstructural redistribution associated with strain induced by magnetostrictive effects. The commonality of phenomena between magnetic (Fig. 20a) and anelastic (Fig. 20b) after-effect suggests a common origin. The time constant of magnetic after-effects can be appreciable. The effect is strongly dependent on temperature (faster for higher temperatures) and on material purity. A detailed discussion of the different contributions to magnetic after-effect is found in Ref. [57].

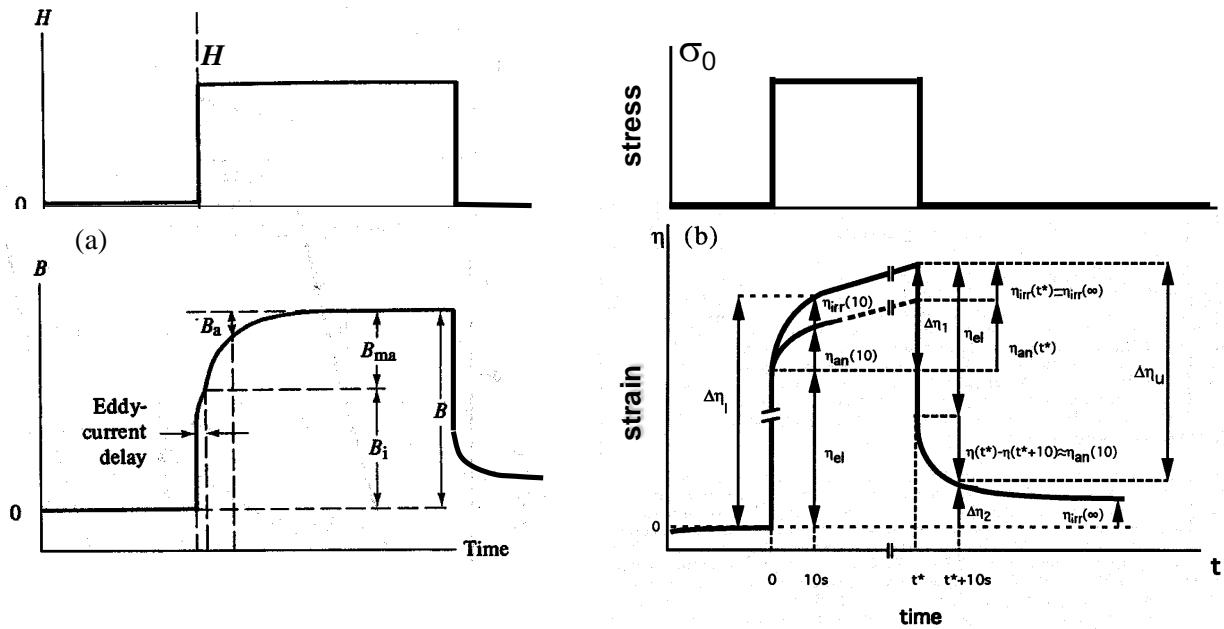


Fig. 20: Analogy between magnetic (a, from Ref. [6]) and anelastic (b, from Refs. [58, 59]) after effect. In Fig. 20a, the field strength in a demagnetized sample is increased suddenly from zero to a constant value H . The induction B , apart from the delay due to eddy currents, rises immediately to a value B_i , and then at a finite rate to the final value B that is associated to H in the magnetization curve. In Fig. 20b, the sudden application of a stress σ_0 to a sample suddenly induces an elastic strain η_{el} , followed by an elastic after-effect (time-delayed deformation), composed of reversible anelastic (η_{an}) and irreversible viscoelastic (η_{irr}) contributions. Delayed effects are observed as well on removal of the field (magnetic or elastic).

5 Conclusions

Magnetic materials are key elements of magnet technology. They should be procured on the basis of careful selection and adapted specifications, since their primary and secondary metallurgy, chemical composition, purity, applied thermal treatments, and microstructure will have a significant influence on their final properties. Low-carbon steel laminations, but also general-purpose constructional steels, such as type 1010, generally used for applications that require ‘less than superior’ magnetic properties [4], are often applied as yoke materials for accelerator and experiment magnets. They are not always purchased to magnetic specifications. Soft ferromagnetic materials of better controlled composition and impurity limits, properties and metallurgy might be considered for specific applications, such as fast magnet systems. On the other hand, innovative materials such as nanocrystalline and amorphous alloys are being considered or are already used for an increasing number of devices, including for high-energy physics and fusion-related applications. Examples are high-frequency transformers for the International Linear Collider and induction cores of heavy-ion inertial fusion-energy accelerators, respectively. In 2007 nanocrystalline materials represented a production of 1000 t. The importance and use of powder metallurgy is also increasing for application to structural components of magnets, soft magnetic materials, and materials for permanent magnets. In 2003, powder-based soft ferrites represented 5% of the world market of magnetic materials, including semihard and hard materials, compared to 27% covered by conventional steels [60].

Acknowledgement

The author wishes to thank L. Walckiers and D. Brandt for their invitation to contribute to the special CERN Accelerator School, D. Tommasini for fruitful exchanges, G. Peiro and J. Garcia Perez for helpful discussion on testing facilities available at CERN, F. Caspers for the expertise provided on soft ferrites, and R. Veness for the final spell checking of the manuscript.

References

- [1] Online Etymology Dictionary, <http://www.etymonline.com/>.
- [2] P.F. Mottelay, Bibliographical History of Electricity and Magnetism (Charles Griffin, London, 1922).
- [3] R.M. Bozorth, Ferromagnetism (Van Nostrand, New York, 1951).
- [4] ASM Metals Handbook, 9th ed., vol. 3: Properties and Selection: Stainless Steels, Tool Materials and Special-Purpose Metals (American Society for Metals, Metals Park, OH, 1980).
- [5] R. C. O’Handley, Modern Magnetic Materials: Principles and Applications (Wiley, New York, 1999).
- [6] B.D. Cullity and C.D. Graham, Introduction to Magnetic Materials, 2nd ed. (Wiley, Hoboken, NJ, 2009).
- [7] Handbook of Magnetism and Advanced Magnetic Materials, vol. 4 – Novel Materials, H. Kronmüller and S. Parkin (eds.) (Wiley, Hoboken, NJ, 2007).
- [8] Fletcher et al., Low gravity phase separator, United States Patent 4027494 (1977).
- [9] Alpha Magnetic Spectrometer (AMS) Experiment Utilization Plan for Exploration Magnet Development, NASA AMS Project Office, Houston, Texas, Preliminary Draft March 24, 2005.
- [10] F.R. Fickett and R.B. Goldfarb, Magnetic properties, *in*: Richard P. Reed and Alan F. Clark (eds.), Materials at Low Temperatures (American Society for Metals, Metals Park, OH, 1983), p. 209.

- [11] P.A. Serena and N. García, Ferromagnetism in FeCu metastable alloys, Phys. Rev. B 50 (1994) 944–948.
- [12] L.A.A. Warnes and H.W. King, Cryogenics, 16 (1976) 659–667.
- [13] S. Sgobba and G. Hochoertler, A new non-magnetic stainless steel for very low temperature applications, Proceedings of the International Congress Stainless Steel 1999: Science and Market, Chia Laguna, Italy, 1999 (Associazione Italiana di Metallurgica, Milan, 1999), vol. 2, pp. 391–401.
- [14] F. Fiorillo, Measurement and Characterization of Magnetic Materials (Elsevier, Amsterdam, 2004).
- [15] W. Fulkerson, J.P. Moore, and D.L. McElroy, Comparison of the thermal conductivity, electrical resistivity, and Seebeck coefficient of a high-purity iron and an Armco iron to 1000°C, J. Appl. Phys. 37 (1966) 2639–2653.
- [16] A.J. Otter, Annealing cycles for magnet steel and variation in *B-H* characteristics, Tri-University Meson Facility, Design Note n° TRI-DN-69-3, University of British Columbia, Vancouver, B.C. (1969).
- [17] Armco Magnetic Ingot Iron (The American Boiling Mill Co., Middletown, OH, 1945).
- [18] C.J. Smithells (ed.), Metals Reference Book, 4th ed., vol. 2 (Butterworths, London, 1967).
- [19] ThyssenKrupp Steel catalogue, Non Grain Oriented Electrical Steel Powercore® (2008).
- [20] IEC 60404-1-1 Magnetic materials Part 1-1: Classification surface insulations of electrical steel sheet, strip and laminations (2004).
- [21] S. Bates, D. Willis, and K. Sinram, IEEE Trans. Magn. 28 (1992) 964.
- [22] EN 10106:1996, Cold rolled non-oriented electrical steel sheet and strip delivered in the fully processed state.
- [23] PowerCore® grain oriented electrical steel, ThyssenKrupp Steel catalogue (2008).
- [24] T. Yamaguchi and K. Takeda, Effects of crystallographic misorientation on power loss in (110)[001] three percent Si-Fe sheets with low thickness, IEEE Trans. Magn. 21 (1985) 41–44.
- [25] S.D. Washko and R.F. Miller, Sheet thickness effects on energy loss in 3% silicon-iron, J. Magn. Mater. 19 (1980) 361–364.
- [26] G. Béranger, F. Duffaut, J. Morlet, and J.F. Tiers, The Iron-Nickel Alloys: A Hundred Years After the Discovery of Invar (Lavoisier Pub., Secaucus, NJ, 1996), translated from the French.
- [27] G. Couderchon, Techniques de l'Ingénieur, Alliages fer-nickel et fer-cobalt – Propriétés magnétiques (1994).
- [28] J.M. Jimenez and B. Henrist, Vacuum chambers for the circulating beams in the LHC injection and extraction septa, functional specification, CERN EDMS document n° 405734, 14/05/2004.
- [29] S. Sgobba, Tubes en Mumetal® 77Ni-5Cu-4Mo-Fe pour chambres spéciales LSS, fiche technique CERN n° 555 – rev. 0, 08/01/2004.
- [30] S. Sequeira Tavares and S. Sgobba, CERN Metallurgy report n° 98/10/06 (1998).
- [31] G.A. Kelsall, Physics, 5 (1934) 169–172.
- [32] L. Néel, J. Phys. Radium, 15 (1954) 225.
- [33] S. Taniguchi, Sci. Rep. Res. Inst. Tcihku Univ. A7 (1955) 269.
- [34] S. Chikazumi and T. Oomura, J. Phys. Soc. Jpn. 10 (1955) 842.

- [35] D.I. Gordon and R.S. Sery, A new magnetic-field electron-irradiation treatment for producing low-coercive-force rectangular hysteresis loops in Supermalloy, *IEEE Trans. Nucl. Sci.* NS-11 (1964) 105–110.
- [36] F.G. Hanejko, H.G. Rutz, and C.G. Oliver, Effects of processing and materials on soft magnetic performance of powder metallurgy parts, paper presented at the 1992 Powder Metallurgy World Congress, San Francisco, CA, 21–26 June 1992.
- [37] J.H. Bularzik, R.F. Krause, and H.R. Kokal, Properties of a new soft magnetic material for AC and DC motor applications, *J. Phys. (Paris)* IV, 8 (1998) Pr2-747–Pr2-753.
- [38] G. Nord, L.O. Pennander, and A.G. Jack, Loss calculation for soft magnetic composites, 16th ICEM Conference, Cracow, Sept. 2004.
- [39] Ferrite Absorber Tiles, FERROXCUBE catalogue (2003).
- [40] Ferrite Production, unpublished FERROXCUBE document.
- [41] S. Mader and A.S. Novick, Metastable CoAu alloys: example of an amorphous ferromagnet, *Appl. Phys. Lett.* 7 (1965) 57.
- [42] Y. Abe, K. Miyazawa, M. Nakamura, and T. Ohashi, The behavior of metal jet in the in-rotating-water spinning method, *Trans. ISIJ*, 27 (1987) 929.
- [43] G. Herzer, *J. Magn. Magn. Mater.* 254/255 (2003) 598–602.
- [44] T. Meydan, Amorphous materials in sensor applications, magnetic materials for sensors and actuators, *in:* Technical Digest, IEE Colloquium on New Magnetic Materials, London, UK, 11 October 1994.
- [45] Tape-Wound Cores for Magnetic Amplifier Chokes VITROVAC 6025 Z, Vacuumschmelze document n° PK-002 (1998).
- [46] W. Waldron, The Heavy Ion Fusion Virtual National Laboratory, hif.lbl.gov.
- [47] W.A. Reass, D.M. Baca, and D.E. Rees, Possibilities of high frequency polyphase resonant converters for ILC application, paper presented at the 2004 Power Modulator Conference, San Francisco, CA, 2004.
- [48] A.W. Molvik and A. Faltens, Induction core alloys for heavy-ion inertial fusion-energy accelerators, *Phys. Rev. Spec. Top. Accel. Beams*, 5 (2002) 080401.
- [49] P. Haren, Magnetic permeability measurements, automatic testing of ring samples at cryogenic and ambient temperatures, CERN ISR-BOM/78-4A.
- [50] A. Cazes, A. Cecchetti, B. Dulach, F. Jungo, M. Incurvati, D. Orecchini, G. Peiro, C. Sanelli, F. Terranova, and M. Ventura, Electromagnetic characterization of the 990 ton gapless magnets for the OPERA experiment, technical report published by Institute of Physics Publishing and SISSA (2007).
- [51] J. Billan, Coercimeter for non-destructive measurement of the coercivity of steel sheets, *J. Phys. Colloques*, 45, C1 (1984) C1-965–C1-968.
- [52] Foerster Instruments Incorporated, documentation on portable MAGNETOSCOP® 1.069, (2008).
- [53] ASTM A342/A342M – 04 Standard Test Methods for Permeability of Feebly Magnetic Materials.
- [54] J. Sort, A. Concristell, E. Menéndez, S. Suriñach, M. D. Baró, J. Farran, and J. Nogués, Selective generation of local ferromagnetism in austenitic stainless steel using nanoindentation, *Appl. Phys. Lett.* 89 (2006) 032509.

- [55] A. Chao and M. Tigner, *Handbook of Accelerator Physics and Engineering* (World Scientific, Singapore, 1999).
- [56] C. Arimatea and D. Jacquet, Mesure de l'influence magnétique du cycle proton 450 GeV sur le cycle positron, SL-Note-2001-013 MD 20, février 2001.
- [57] B. Hohler and H. Kronmüller, Magnetic after-effects in ferromagnetic materials, *J. Magn. Magn. Mater.* 19 (1980) 267–274.
- [58] S. Sgobba, PhD thesis EPFL n° 1215, 1994.
- [59] S. Sgobba, H.U. Künzi, and B. Ilschner, *Acta Metall. Mater.* 43 (1995) 1171–1181.
- [60] M. Vazquez, *Materiales Magnéticos* (COTEC Foundation, Madrid, 2003).

Basic design and engineering of normal-conducting, iron-dominated electromagnets

Th. Zickler

CERN, Geneva, Switzerland

Abstract

The intention of this course is to provide guidance and tools necessary to carry out an analytical design of a simple accelerator magnet. Basic concepts and magnet types will be explained as well as important aspects which should be considered before starting the actual design phase. The central part of this course is dedicated to describing how to develop a basic magnet design. Subjects like the layout of the magnetic circuit, the excitation coils, and the cooling circuits will be discussed. A short introduction to materials for the yoke and coil construction and a brief summary about cost estimates for magnets will complete this topic.

1 Introduction

The scope of these lectures is to give an overview of electromagnetic technology as used in and around particle accelerators considering *normal-conducting, iron-dominated* electromagnets generally restricted to direct current situations where we assume that the voltages generated by the change of flux and possible resulting eddy currents are negligible. Permanent and superconducting magnet technologies as well as special magnets like kickers and septa are not covered in this paper; they were part of dedicated special lectures.

It is clear that it is difficult to give a complete and exhaustive summary of magnet design since there are many different magnet types and designs; in principle the design of a magnet is limited only by the laws of physics and the imagination of the magnet designer. Furthermore, each laboratory and each magnet designer or engineer has his own style of approaching a particular magnet design. Nevertheless, I have tried to gather general and common principles and design approaches.

I have deliberately focused on applied and practical design aspects with the main goal of providing a guide-book with practical instructions on how to start with the design of a standard accelerator magnet. As far as I know, there is no such manual that provides step-by-step instructions allowing the setting up of a first, rough analytic design before going into a more detailed numerical design with field computation codes like ROXIE, OPERA, ANSYS, or POISSON. This guide-book should also help to assess and validate the feasibility of a design proposal and to draft a list of the key parameters (with just pencil and paper) without spending time on complex computer programs.

Please keep in mind that these lectures are meant for students of magnet design and engineering working in the field of accelerator science — not for advanced experts.

For the sake of brevity and simplicity I have refrained from deriving once again Maxwell's equations — they have been extensively treated by experts in other lectures. You will also find mathematics reduced to a bare minimum. The derivation of formulas in this text might sometimes appear condensed, but in case you want to learn more, you should always be able to find the sources with the help of the bibliography cited at the end. To guarantee consistency throughout, SI (MKSA) units are used systematically.

The paper starts with a short introduction to basic concepts and magnet types, followed by a section dedicated to collecting information and defining the requirements and constraints before starting the actual design. The main part gives an introduction to basic analytic magnet design covering topics such as yoke design, coil dimensioning, cooling layout, material selection, and cost estimation. Although the lectures presented during the course included a section introducing numerical design methods, it has been omitted from the proceedings since it was found to be too exhaustive. The bibliography recommends literature for further reading for those who wish to go more deeply into this subject.

2 Basic concepts and magnet types

We introduce basic concepts and classical normal-conducting magnet types, highlight their main characteristics, and explain very briefly their function and purpose in a particle accelerator.

2.1 Dipoles

In a circular particle accelerator or in a curved beam transfer line, dipoles are the most commonly used elements. A dipole provides a uniform field between its two poles which is excited by a current circulating in the coils. The system follows the right-hand convention, i.e., a current circulating clockwise around the poles produces a magnetic field pointing downwards.

Their purpose is to bend or steer a charged particle beam. Applying again the right-hand rule, when a beam of positively charged particles directed into the plane of the paper sees a field pointing downwards, it is deflected to the left, as shown in Fig. 1 (a).

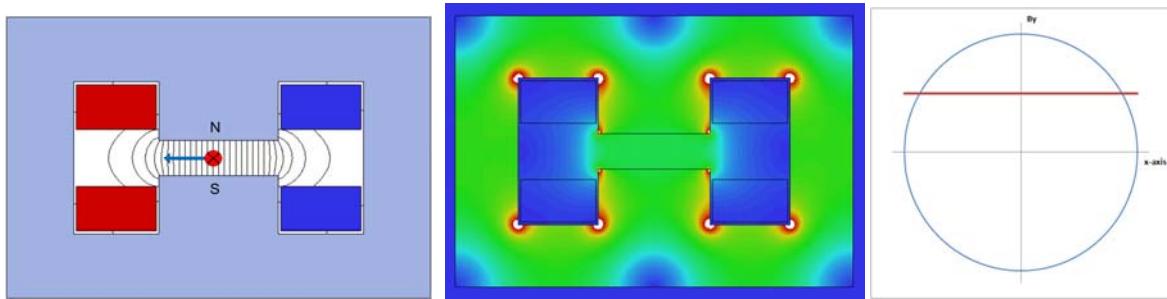


Fig. 1: Dipole: cross-section (a), 2D-field distribution (b), and field distribution on the x -axis (c)

The equation describing a normal ideal (infinite) pole is:

$$y = \pm r$$

where r is the half-gap height. The magnetic flux density between these two poles is ideally constant and has only a component in the y -direction, as one can see from Fig. 1 (a)–(c):

$$B_y = a_1 = B_0 = \text{const.}$$

In an ideal dipole only harmonics of: $n = 1, 3, 5, 7\dots$ ($= 2n$ pole errors) can appear. These are called the ‘allowed’ harmonics.

2.2 Quadrupoles

The second most commonly used magnetic elements are quadrupoles. Their purpose is to focus the beam. Note that a horizontally focused beam is at the same time vertically defocused. A quadrupole has four iron poles with hyperbolic contour which can be ideally described for a normal (non-skew) quadrupole by

$$2xy = \pm r^2 ,$$

where r is the aperture radius.

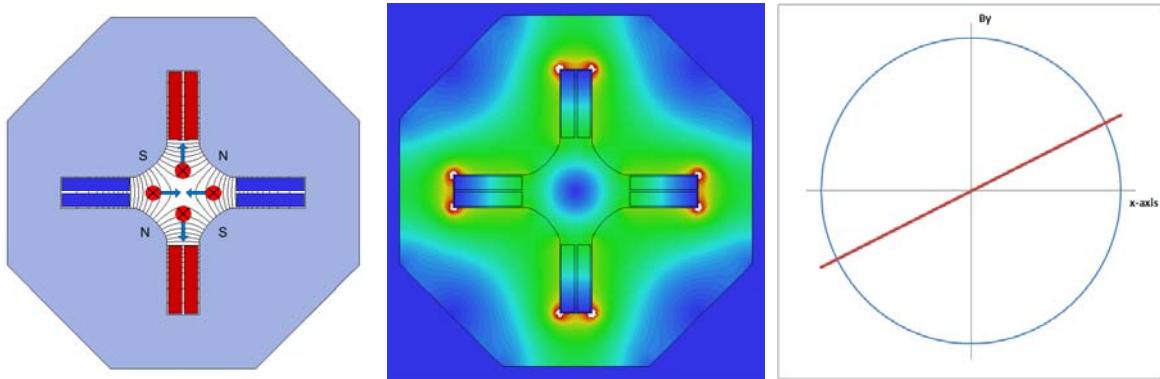


Fig. 2: Normal quadrupole: cross-section (a), 2D-field distribution (b), and field distribution on the x -axis (c)

A quadrupole provides a field which is zero at the centre and increases linearly with distance from the centre as shown in Fig. 2 (c). The equipotential lines are hyperbolas ($xy = \text{const.}$) and the field lines are perpendicular to them. Dipoles and quadrupoles are linear elements, which means that the horizontal and the vertical betatron oscillations are completely decoupled. The Cartesian components of the flux density in an ideal quadrupole are not coupled; the x -component in a certain point only depends on the y -coordinate and the y -component only depends on the x -coordinate following the relation

$$B_y = a_2 x \quad \text{and} \quad B_x = a_2 y .$$

With the polarity shown in Fig. 2 (a), the horizontal component of the Lorentz force on a positively charged particle moving into the plane of the drawing, is directed towards the axis; the vertical component is directed away from the axis. This case thus exhibits horizontal focusing and vertical defocusing.

The ‘allowed’ harmonics in an ideal quadrupole are: $n = 2, 6, 10, 14, \dots$ ($= 2n$ pole errors).

2.3 Sextupoles

Sextupoles can be found in circular accelerators and less often in transfer lines. They have six poles of round or flat shape. Their main purpose is to correct chromatic aberrations: particles which are off-momentum will be incorrectly focused in the quadrupoles, which means that high-momentum particles with stronger beam rigidity will be under-focused, so that betatron oscillation frequencies will be modified. A positive sextupole field can correct this effect and can reduce the chromaticity to zero, because off-momentum particles circulate with a radial displacement with respect to the ideal trajectory and see therefore a correcting field in the sextupole as shown in Fig. 3 (a). We have also seen that the first ‘allowed’ harmonic in a dipole is the sextupole component, which leads to a resulting negative chromaticity requiring compensation by distinct sextupole elements.

The equation for a normal (non-skew) sextupole with ideal poles is

$$3x^2y - y^3 = \pm r^3$$

where r is again the aperture radius. The magnetic field varies quadratically with the distance from the magnet centre as one can see in Fig. 3 (c). Sextupoles are non-linear elements, which means that the y -component of the flux density at a certain point in the aperture depends on both the x - and y -coordinate, and is described by

$$B_y = a_3(x^2 - y^2)$$

The ‘allowed’ harmonics in an ideal sextupole are: $n = 3, 9, 15, 21\dots$ ($= 2n$ pole errors).

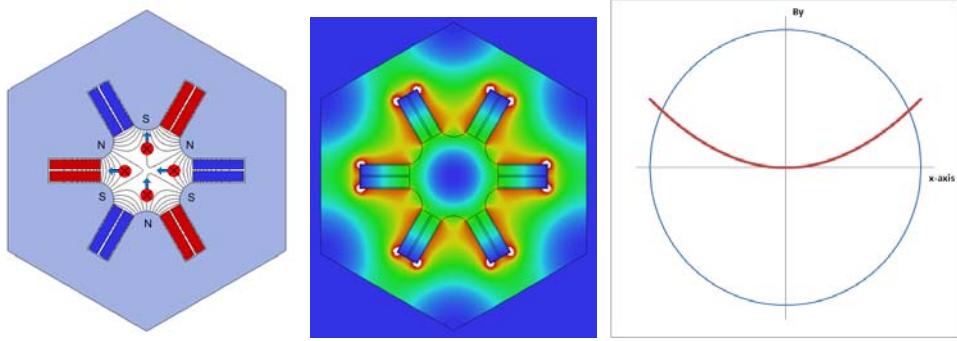


Fig. 3: Normal sextupole: cross-section (a), 2D-field distribution (b), and field distribution on the x -axis (c)

2.4 Octupoles

Octupoles are quite rarely used and can be mainly found in colliders and storage rings. Amongst other purposes, they are used for ‘Landau’ damping, to introduce a tune-spread as a function of the betatron amplitude, to de-cohere the betatron oscillations, and to reduce non-linear coupling. The eight poles of a normal (non-skew) ideal octupole as shown in Fig. 4 (a) and (b) follow the equation

$$4(x^3y - xy^3) = \pm r^4$$

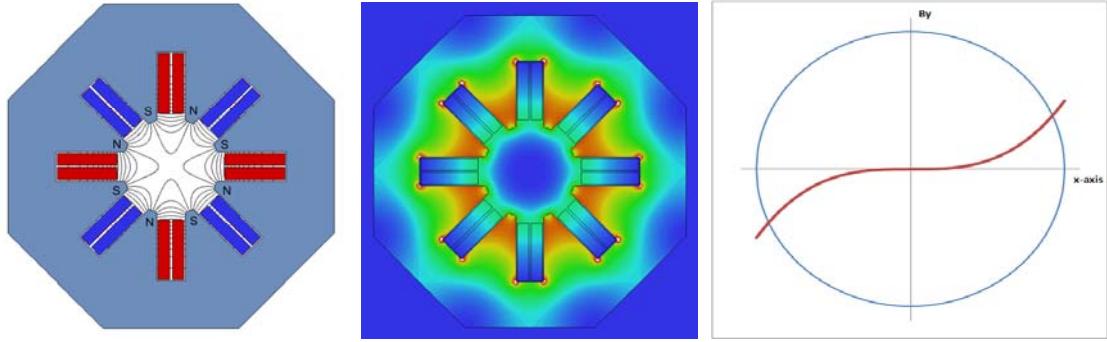


Fig. 4: Normal octupole: cross-section (a), 2D-field distribution (b), and field distribution on the x -axis (c)

The y -component of the magnetic flux density in any point of the aperture can be described by the following relation

$$B_y = a_4(x^3 - 3xy^2)$$

The ‘allowed’ harmonics in an ideal octupole are: $n = 4, 12, 20, 28\dots$ ($= 2n$ pole errors).

2.5 Skew magnets

Skew versions exist for all of the above-described magnet types. Skew means a rotation of the magnet along the longitudinal axis by $90^\circ/n$, where n is the index of the main field component (i.e., $n = 1$ for dipole, $n = 2$ for quadrupole, $n = 3$ for sextupole). Rotating linear magnetic elements leads to loss of the betatron decoupling. Fig. 5 shows a skew quadrupole, the purpose of which is to control the coupling of horizontal and vertical betatron oscillations. In a skew quadrupole, a beam that is displaced in the horizontal plane is deflected vertically, and a beam that is displaced in the vertical plane is deflected horizontally.

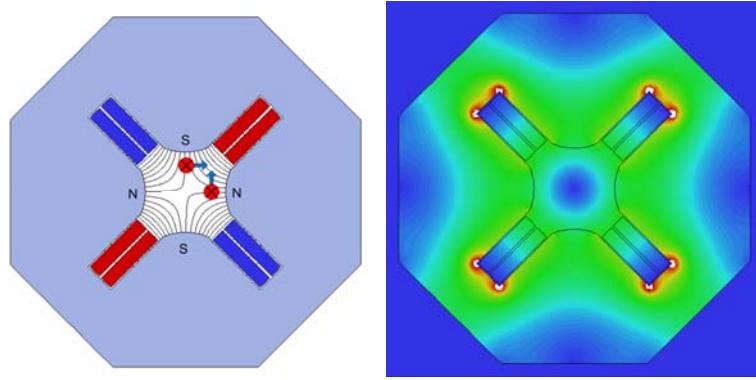


Fig. 5: Skew quadrupole: cross-section (a), 2D-field distribution (b)

2.6 Combined-function magnets

Combined-function magnets unite several main field components in one magnet, e.g., a dipole and a quadrupole. We can distinguish between two types of combined-function magnet.

These are magnets where the different functions are generated by the sum of scalar potentials and the shape of the pole, and magnets where the different functions are generated by separate coils individually powered.

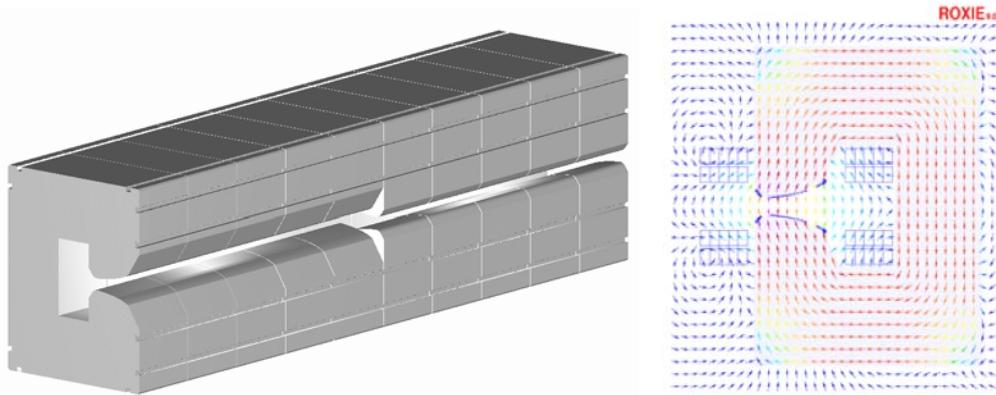


Fig. 6: Combined-function magnet yoke of the CERN Proton Synchrotron

The second type is of minor importance and sometimes used when limited space in the machine demands special solutions. An example of a quadrupole with integrated steering coils is illustrated in Fig. 7. Other types combine sextupoles with steering functions or quadrupoles with sextupoles. The advantage here is that the amplitudes of both field components can be adjusted independently, but often the field quality of one function is significantly reduced. In the example shown, the dipole field

suffers from a strong sextupole component because the yoke geometry has been tailored to develop a quadrupole field.

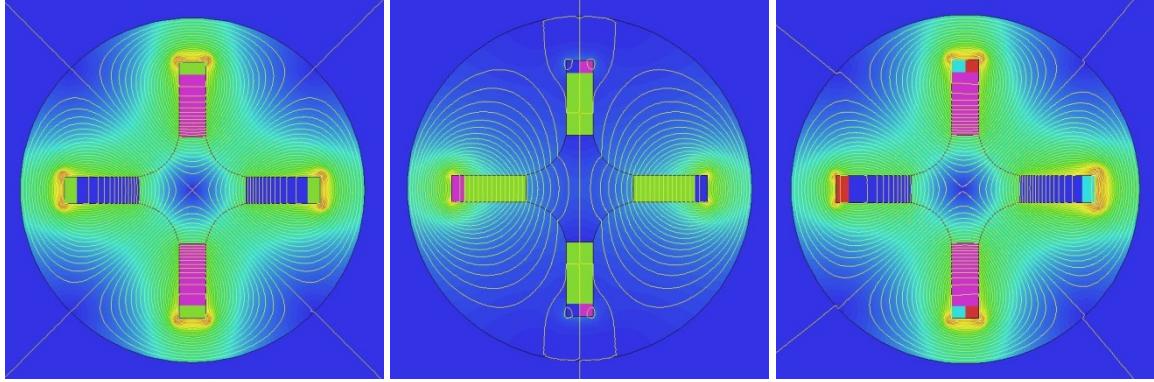


Fig. 7: Quadrupole with integrated steering coils: quadrupole field only (a), dipole field only (b), dipole field superimposed on quadrupole (c)

A solution to circumvent this field quality problem is shown in Fig. 8, where several functions (horizontal dipole, vertical dipole, quadrupole, skew quadrupole, and sextupole) are incorporated in one magnet. The field distribution in this case is solely determined by the conductor geometry and not by iron poles. In Fig. 8 (a) only the coils providing the horizontal dipole field are powered, while Fig. 8 (b) illustrates the field distribution when all magnetic functions are excited.

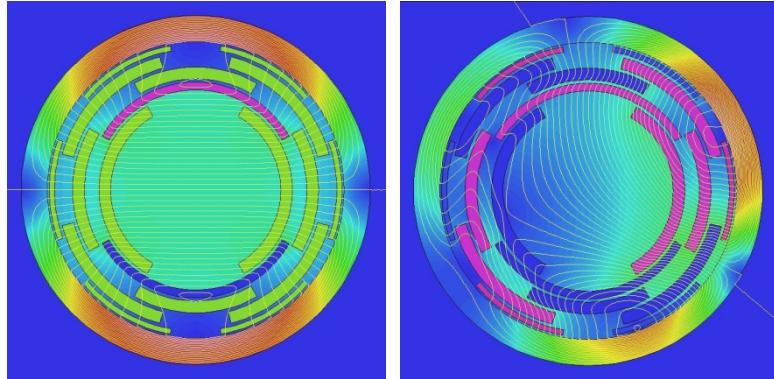


Fig. 8: Nested combined-function corrector: vertical dipole (a) and combined field distribution (b)

2.7 Solenoids

Although solenoids are strictly speaking not iron-dominated magnets, they will be briefly introduced here for the sake of completeness. A lecture entirely dedicated to solenoids can be found in these proceedings.

Solenoids are relatively simple lenses with a field created by a rotationally symmetric coil. From Maxwell's equation $\text{div}B = 0$, the magnetic field, which is purely longitudinal in the inner part of the coil, must contain radial components at the entrance and at the exit. While particles moving exactly on the axis do not experience any force, the others suffer an azimuthal acceleration due to the radial component while entering and leaving the lens. Because of the azimuthal motion there is a radial force in the longitudinal field. This force is proportional to the radial distance from the axis. To increase the field close to the axis and to capture and limit the stray field, solenoid coils are usually surrounded by an iron yoke.

3 What do we need to know before starting?

Before one can enter into the design of a magnet all relevant information which will have an influence on the design, construction, installation, and operation of the future magnet has to be put together. What the term ‘relevant’ means is explained in this section preceded by a brief discussion about goals in magnet design and magnet life cycles.

3.1 Goals in magnet design

We should always keep in mind that the goal in magnet design is to produce a device which is just *good enough* to perform *reliably* with a sufficient *safety factor* at the *lowest cost* and *on time* [1].

What ‘on time’ means should be obvious: in particular in commercial projects, a delay in the start-up of the operation will result in financial losses. The meaning of ‘lowest cost’ should also be clear. We will see later how the costs can be optimized. But what does ‘good enough’ mean? On each project, the obvious parameters such as magnetic field, magnet aperture, magnet dimensions, power consumption, etc. are more or less clearly specified, but it is the tolerances on these parameters that are very often challenging to define. They are a function of the expected machine performance and acceptable deviations from an ideal machine. In this context, orbit distortions, dynamic aperture, tune width, and transfer efficiency could be mentioned, which can be calculated analytically, but nowadays this is usually done numerically. Nevertheless, the interpretation of the results is not straightforward and in many cases the tolerances which are requested by the accelerator physicists tend to be unnecessarily tight. Overly tight tolerances lead to increased costs. An enhanced communication between the magnet designer and the accelerator physicist and mutual understanding can help to solve this problem.

The term ‘reliably’ basically means to get the Mean Time Between Failures (MTBF) and the Mean Time To Repair (MTTR) to a reasonably low level. Probability theory and risk analysis are well established for industrial engineering and more and more applied now by physicists working in an experimental environment. But for a new design the reliability is usually unknown so one counts on the experience of the magnet engineer to search for a compromise between extreme caution and extreme risk. A detailed design analysis in the framework of an expert review can be helpful in finding this well balanced compromise before proceeding with magnet manufacture.

The last term to be considered is the ‘safety factor’. In many projects, the initial design parameters were raised after a few years of operation. Applying a safety factor allows operating a device under more demanding conditions than those initially foreseen but it also permits operating under nominal conditions with less wear, and design flaws are less critical. Since safety factors are typically linked to a rise of production costs, they need to be negotiated between the project engineer and the management. However, the pileup of arbitrary and redundant safety factors at multiple project levels has to be avoided because it leads to an unnecessary increase of costs.

3.2 Magnet life cycle

The flow diagram in Fig. 9 shows the typical life cycle of a magnet from the design and construction to the installation and operation and to its final disposal or destruction. We will concentrate mainly on the part which is related to design and calculation. This phase can be split up into different steps which are followed more or less sequentially with possible feedback loops at certain stages. At the beginning of each project the requirements, constraints, and boundaries have to be defined. From this set of parameters a first analytic design should be derived followed by a basic numerical design. After each of the sequential steps (electrical design, mechanical design, integration assessment, and cost estimation) one or more re-iterations of the analytic design might become necessary. Once these steps deliver satisfactory results, an advanced numerical design including field optimizations can be launched.

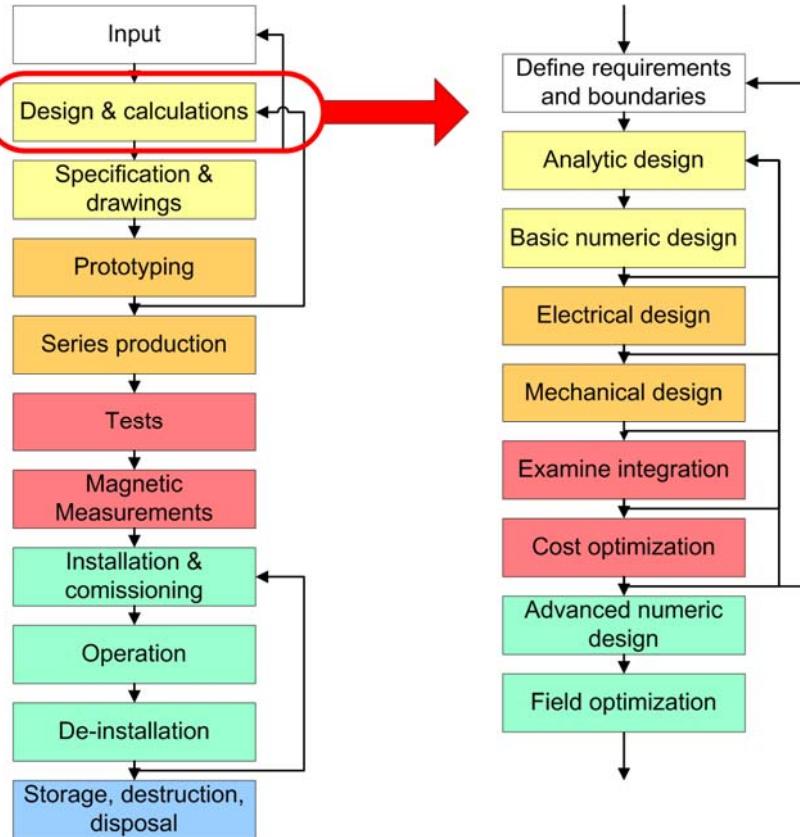


Fig. 9: Magnet life cycle

3.3 Input parameters

It is essential to realize that a magnet is not a stand-alone device. Throughout its life, a magnet has various interactions with other devices and services. These interactions have to be fully considered in the design phase and the magnet designer calls on his experience to ensure that nothing is forgotten. Ignoring one of the key aspects may result in implementing difficult modifications on the finished product. The main interaction partners are summarized in Fig. 10. Some of them like beam optics, power converters, and cooling are obvious and therefore always taken into account from the beginning. Others, such as vacuum, survey, and integration are often considered in a later stage of the project but sometimes too late, thus complicating the life of the involved parties unnecessarily. Examples of partners which are most likely to be forgotten are safety and transport, with the result that substantial and expensive engineering modification might become necessary in order to install or operate a magnet safely. A good and regular communication with all potential partners from the very start of the project and a clear definition of the interfaces can help to avoid such issues.

It is good practice to contact the responsible partners, collect all necessary information, understand the requirements, constraints and interfaces, and summarize them in a functional specification to be finally approved by each of the involved parties before starting the actual design work.

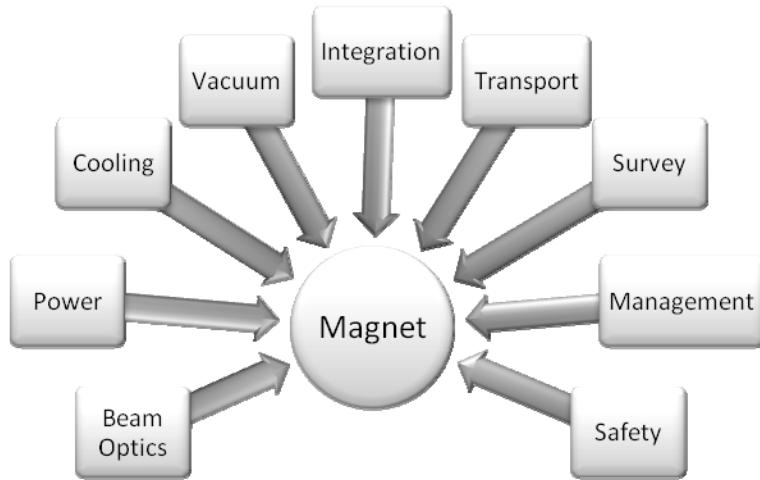


Fig. 10: Magnet interaction partners

The following paragraph should give a general idea of possible requirements and help to set up a check-list of information a magnet designer might need to bring together for creating a comprehensive magnet design.

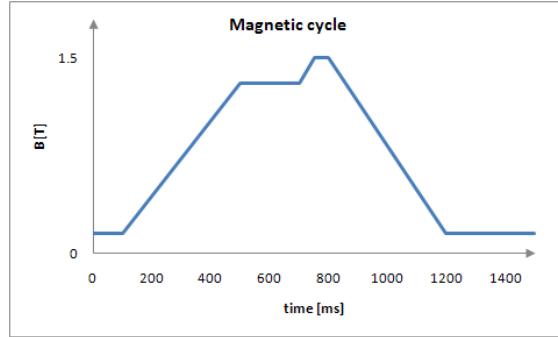
3.3.1 General requirements

First of all the magnet type (dipole, quadrupole, sextupole, octupole, combined-function, solenoid, etc.) and its main purpose (bending, steering, charge stage separation, etc.) need to be defined. It is also important to know where the magnet is foreseen to be installed. It makes a difference for the performance of a magnet whether it will be installed in a storage ring, a synchrotron light source, a collider, a pre-accelerator, or in a beam transport line. The tolerances on storage ring magnets are generally more demanding than on accelerators, because the phase spaces of the beams have to be maintained for many revolutions. It also makes sense to discuss the spares policy with the project management at this stage. To foresee a certain number (typically 10%) of spare units and manufacture them together with the units to be installed helps to reduce or avoid down-time in the case of a magnet failure and to save money at the same time. Spare magnets which are produced afterwards or in a case of urgency are inevitable much more expensive.

3.3.2 Performance requirements

To start with the initial design work, a magnet designer needs to know at least the basic performance parameters, which are typically provided by the accelerator physicists:

- Beam parameters: type of beam (mass and charge state), energy range and deflection angle (k -value in case of quadrupoles).
- Magnetic field: integrated field (or integrated gradient in case of quadrupoles); alternatively the local field (gradient) and magnetic length can be defined.
- Aperture: physical (mechanical) aperture and useful magnetic aperture ('good field region').
- Operation mode: continuous operation, pulsed-to pulse modulation, fast pulsed, definition of the magnetic cycle (an example is shown in Fig. 11) and ramp rates.

**Fig. 11:** Magnetic cycle

- Field quality: requirements on field homogeneity (uniformity), the allowed harmonic content, requirements on stability and reproducibility, maximum settling time (time constant) for transient effects generated by eddy currents. A simple but lucid method to judge the field quality of a magnet is to plot the homogeneity of the field or the gradient along the boundary of the defined good field region. Achieving the following homogeneity values is reasonable but nevertheless challenging.

Dipole:
$$\frac{\Delta B}{B_0} = \frac{B_y(x,y) - B_y(0,0)}{B_y(0,0)} \leq 0.01\%$$

Quadrupole:
$$\frac{\Delta B'}{B'_0} = \frac{B'_y(x,y) - B'_y(0,0)}{B'_y(0,0)} \leq 0.1\%$$

Sextupole:
$$\frac{\Delta B''}{B''_0} = \frac{B''_y(x,y) - B''_y(0,0)}{B''_y(0,0)} \leq 1\%$$

3.3.3 Physical requirements

Important for the mechanical layout — which is of course always linked to the magnetic layout — is to identify whether geometric boundaries or constraints exist. This can be either a limit in the available space in the accelerator or the beam line, a transport limitation like the maximum allowed charge of an existing crane or a weight limitation of the supporting ground. In particular, the accessibility of the installed magnet should be mentioned here. A magnet designer has not only to assure that the magnet can be transported to its position in the machine, but he has also to take care that sufficient space around the magnet is available to handle the electrical and hydraulic connections and to allow unrestricted access to the reference targets so that the survey group can align the magnet accurately in its final position.

3.3.4 Interfaces

The interaction of the magnet with other equipment like power converters, cooling infrastructure, and vacuum systems is quite obvious, but is nevertheless repeated here, since a clear communication and a mutual understanding between the involved groups is essential to avoid any misinterpretation or oversight. It is important to make contact with these groups in an early phase of the design process to clearly define the interfaces and to avoid developing equipment in diverging directions. In this context an example would be a fast-pulsed power converter that cannot be matched to the inductance of the magnet. Another example would be a UH vacuum system requiring in situ bake-out, but the magnet aperture does not then allow installation of such equipment.

The interfaces to the following equipment must be unambiguously defined:

- Power converter: maximum current, maximum voltage, operation mode (pulsed or dc), maximum RMS current, requirements on stability and reproducibility (minimum current), possible control strategies (feed-back, feed-forward), maximum current ramp rates.
- Cooling: available cooling power, maximum flow rate, maximum available pressure, water quality (see Section 4.6.6), circuit type (aluminium or copper), water inlet temperature, temperature stability.
- Vacuum system: size of vacuum chamber, wall thickness and material properties to evaluate potential eddy current issues (only for pulsed magnets), required space for pumping ports and bake-out system, a captive vacuum chamber requiring opening the magnet.

Apart from the above-mentioned points, a magnet may interact with other equipment related to beam diagnostics and monitoring, injection and extraction, RF and control, to name just a few.

3.3.5 *Environmental aspects*

Since environmental aspects are not always evidently related to the performance of a magnet, they often don't get the deserved and necessary attention. Consequently, designers and engineers have to be explicitly asked to take care that these aspects are considered and respected in the magnet design. Neglecting such aspects can lead to serious performance problems with the magnet or surrounding equipment. Remedies for such problems are often complicated and costly. In some rare case where it is impossible to find a suitable solution this negligence can even put the whole project into jeopardy.

Since this field covers a wide spectrum and depends on many parameters, it is impossible to present a universal and exhaustive list of all potential risks, eventualities and dangers. The focus is put on the most common issues, but it has to be understood that it is the clear responsibility of a magnet designer or engineer to look beyond the issues stated here, to critically analyse the environment and to identify and point out all possible risks which could endanger the correct performance of the magnet or the surrounding environment. It can be helpful for such an analysis to bear in mind that the interactions are often bi-directional. This means that the magnet can have an influence on the environment, but the environment can also have an influence on the magnet. The following examples serve to illustrate this principle:

Temperature: elevated environmental temperatures can influence the dew-point such that condensation appears on the surface of water-cooled coils. On the other hand the amount of heat dissipated from the magnet into the tunnel can be so large that it exceeds the capability of the ventilation system and cannot be removed from the tunnel, causing the temperature to rise.

Ionizing radiation: is a specific subject that requires special attention (and which is beyond the scope of this lecture). I would just like to mention the need to select radiation-hard materials and components for accelerator magnets exposed to high radiation levels. The operation of magnets in such an environment also calls for a dedicated design allowing fast repair or replacement, in order to reduce the human intervention time to a minimum.

Electromagnetic compatibility: magnetic fringe fields emitted from the magnets can disturb nearby equipment, such as sensitive beam diagnostic devices, while surrounding equipment made of magnetic material can divert part of the magnetic flux and so locally perturb the field quality.

Safety aspects also have to be seen in this bi-directional way: covers protect the magnet from effects of the environment (dust, accidental water contact), but they also protect the environment from hazards potentially generated by the magnet (electrocution, burning from hot parts).

Table 1 briefly summarizes the aspects which have been discussed in detail above.

Table 1: Aspects to be considered by the magnet designer

General requirements	Magnet type and purpose Application Quantity
Performance requirements	Beam parameters Field requirements Magnet aperture Operation mode
Physical requirements	Geometric boundaries Accessibility
Interfaces	Power converter Cooling Vacuum
Environmental aspects	Temperature Ionizing radiation Electromagnetic compatibility Safety

4 Basic analytical design

Before entering into an extensive and detailed two- or three-dimensional magnetic field study using various available software packages allowing the calculation of field distribution and field quality of complex magnetic assemblies, a basic analytic and conceptual design is necessary. Such an approach will allow one to derive the most important characteristics and parameters of the future magnet with a relatively good accuracy and help one to find a reasonable starting point for the numerical design work (as well as the optics design) and thus reduce the number of design iterations.

A magnet is an assembly of different components. Fig. 12 shows a typical normal-conducting, iron-dominated electromagnet — in this case a quadrupole — and its main components: the magnetic circuit, the excitation coils, the cooling circuit, the alignment targets, the sensors and interlock devices, electrical and hydraulic connections, and the magnet support.

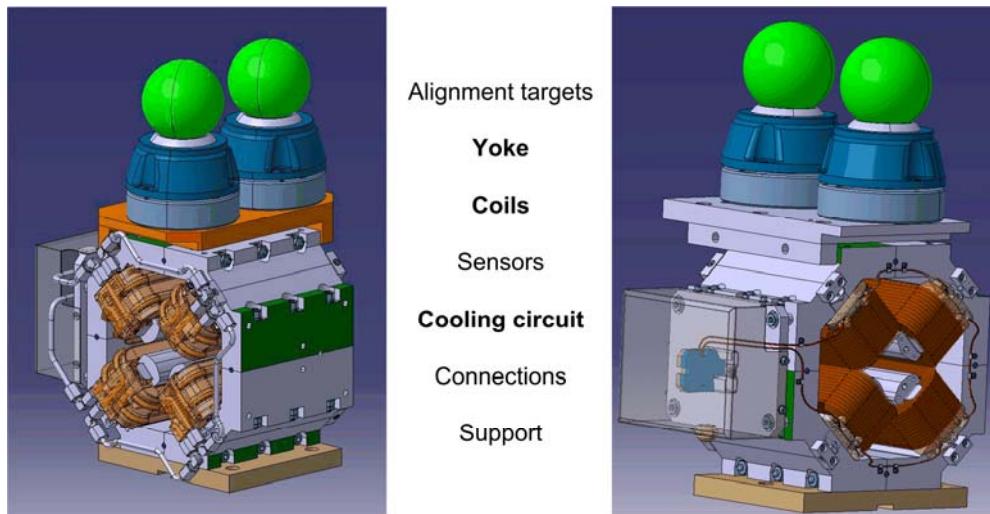


Fig. 12: Magnet main components

The following sections will explain how to design the magnetic circuit and the coils and how to dimension the cooling circuits. The other components will not be discussed here since they have no direct influence on the magnet performance and are rather part of the mechanical design.

4.1 Dipole yoke design

The first step is to derive the geometry of the magnetic circuit or magnet yoke. This means we have to translate the beam optics requirements into a magnetic design defining the yoke characteristics such as the magnetic induction, the aperture size, and the magnet excitation (ampere-turns).

4.1.1 Beam rigidity

A good starting point to define the necessary magnetic induction is to determine the beam rigidity as a function of the particle type and the envisaged beam energy. The beam rigidity $B\rho$ in [Tm] describing the stiffness of a beam can be seen as the resistance of a particle beam against a change of direction when applying a bending force and is defined as

$$B\rho = \frac{1}{qc} \sqrt{T^2 + TE_0} \quad (1)$$

where q is the particle charge number in [C, coulomb], c is the speed of light in [m/s], T is the kinetic beam energy in [eV], and E_0 is the particle rest mass energy in [eV] which is 0.51 MeV for electrons and 938 MeV for protons.

4.1.2 Magnetic induction

From the beam rigidity and the assumed bending radius of the magnets we can calculate the flux density or magnetic induction¹ B in [T] for a dipole magnet

$$B = \frac{B\rho}{r_M} \quad (2)$$

with r_M being the magnet bending radius in [m].

4.1.3 Aperture size

The aperture size of a magnet as presented in Fig. 13 is mostly determined by a central region around the theoretical beam trajectory. This region is referred to as ‘good field region’ and defines the region where the field quality has to be within certain tolerances. The good field region can be circular, rectangular, or elliptical, and takes into account the maximum beam size as well as a certain margin for closed orbit distortions (5–10 mm).

The maximum beam size can be calculated with the help of Eq. (3) which takes into account the lattice functions (beta functions β and dispersion D), the geometrical transverse emittances ε , which

are energy dependent and the momentum spread $(\frac{\Delta p}{p})$

¹ Generally speaking B has to be a vector. For our purpose it is sufficient and correct to assume that only the main field component in the y -direction is present, so B can be written as a scalar. Analogous considerations can be made for quadrupoles and sextupoles.

$$\sigma = \sqrt{\varepsilon\beta + \left(D \frac{\Delta p}{p}\right)^2} \quad (3)$$

For the beam envelope, a few sigma are typically assumed. The largest beam sizes can usually be expected at injection energy. The total required aperture size is the sum of the good field region, the vacuum chamber thickness (0.3–2 mm) and a margin for installation and alignment (0–5 mm).

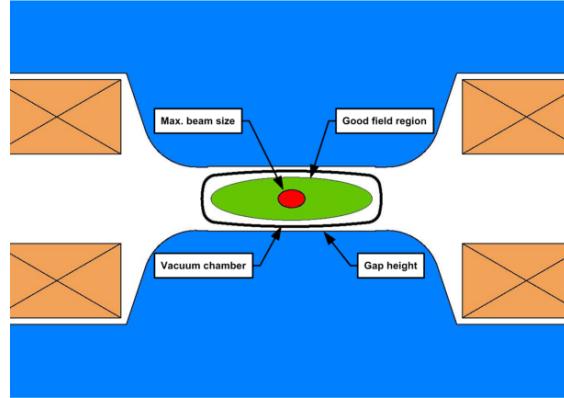


Fig. 13: Defining the aperture size

Please note that the numbers in parentheses are typical values for synchrotrons and are meant to give an indication for the order of magnitude. Depending on the individual case they can be significantly different from the quoted numbers.

4.1.4 Excitation current

Knowing the aperture of the magnet we can continue to calculate the excitation current in the coils required to drive the desired field strength.

Ampere's law

$$\oint \overrightarrow{H} \cdot d\overrightarrow{l} = NI$$

and

$$\overrightarrow{B} = \mu \overrightarrow{H}$$

and

$$\mu = \mu_0 \mu_r$$

$$NI = \oint \frac{\overrightarrow{B}}{\mu} \cdot d\overrightarrow{l} = \int_{gap} \frac{\overrightarrow{B}}{\mu_{air}} \cdot d\overrightarrow{l} + \int_{yoke} \frac{\overrightarrow{B}}{\mu_{iron}} \cdot d\overrightarrow{l} = \frac{Bh}{\mu_{air}} + \frac{B\lambda}{\mu_{iron}} \quad (4)$$

leads to

when we integrate B along a closed path as shown in Fig. 14 and assume that B remains constant along this path.

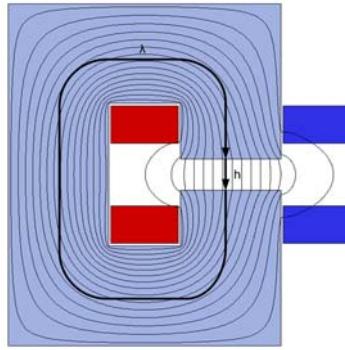


Fig. 14: Closed integration path in a dipole magnet

The gap height is indicated by h and the mean flux path in the iron circuit by λ . As long as the iron is not saturated we can further assume that

$$\frac{h}{\mu_{air}} \gg \frac{\lambda}{\mu_{iron}}$$

such that Eq. (4) can be simplified to

$$NI_{(per\ poles)} = \frac{Bh}{2\eta\mu_0} \quad (5)$$

where h is the magnet gap height in [m], H is the magnetic field vector in [A/m], η is the efficiency (typically 99%), μ_0 is the permeability of free space ($4\pi \cdot 10^{-7}$ [Vs/Am]), and μ_r is the relative permeability ($\mu_{air} = 1$ and $\mu_{iron} > 1000$ if not saturated). Note that Eq. (5) is only approximate and neglects fringe fields and iron saturation.

4.1.5 Reluctance and efficiency

In analogy to Ohm's law, one can define the 'resistance' of a magnetic circuit, called 'reluctance', as

$$R_M = \frac{NI}{\Phi} = \frac{l_M}{A_M \mu_r \mu_0} \quad (6)$$

with Φ indicating the magnetic flux in [Wb], l_M the flux path length in the iron part in [m] and A_M the iron cross-section perpendicular to the flux in [m^2].

The second term ($\frac{\lambda}{\mu_{iron}}$) in Eq. 4 is called 'normalized reluctance' of the yoke.

If we are not careful enough with our design we can create more or less saturated areas in the iron yoke. Saturation means a local decrease of the iron permeability (small μ_{iron}) which leads to inefficiencies of the magnetic circuit. It is good practice to keep the iron yoke reluctance smaller than

a few per cent of air reluctance ($\frac{h}{\mu_0}$) by providing a sufficiently large iron cross-section such that the magnetic flux in the iron remains smaller than 1.5 T. If the recommendation ($\frac{\lambda}{\mu_{iron}} < 0.01 \frac{h}{\mu_0}$) is followed diligently, the efficiency is better than 99%.

Efficiency:

$$\eta = \frac{R_{M,gap}}{R_{m,gap} + R_{M,yoke}} \approx 99\%$$

4.1.6 Magnetic length

To understand the concept of magnetic length, we have to imagine approaching the magnet with a measurement probe along the beam axis from infinity towards the magnet centre. What we will read on the instrument is a steady increase of the field when we move closer to the edge of the iron yoke passing through the stray field of the magnet. The field continues to rise even when we are entering the gap of the magnet and will reach its maximum value when we move the probe towards the centre of the magnet where it remains more or less stable until we move again away from the centre towards the other end of the magnet. We see from Fig. 15 that the field does not increase suddenly but steadily when we approach the edge of the iron yoke. Integrating the magnetic field along the longitudinal axis starting from far outside on one side and ending far outside from the magnet on the other side will give a higher value than simply multiplying the local magnetic field with the iron length of the magnet. Here we can introduce the term ‘magnetic length’ l_{mag} which is defined as

$$l_{mag} = \frac{\int_{-\infty}^{\infty} B(z) \cdot dz}{B_0} \quad (7)$$

We can conclude that the magnetic length is always larger than the actual iron length. To calculate exactly the magnetic length analytically can be quite difficult. Usually it is derived from numerical computations by integrating the field along the magnet as described above and dividing it by the local field in the centre of the magnet. Nevertheless, there is a way to approximate the magnet length, which works well in cases where the iron length of the magnet is much larger than its gap height. For a dipole we can estimate the magnet length with

$$l_{mag} = l_{iron} + 2hk \quad (8)$$

where k is a constant which is specific to the yoke geometry. The constant k gets smaller when the pole width is smaller than the gap height, when saturation occurs in the pole regions, or when the coil heads are close to the beam axis. Typical values of k are between 0.3 and 0.6. A precise determination of k is only possible with measurements or numerical calculations.

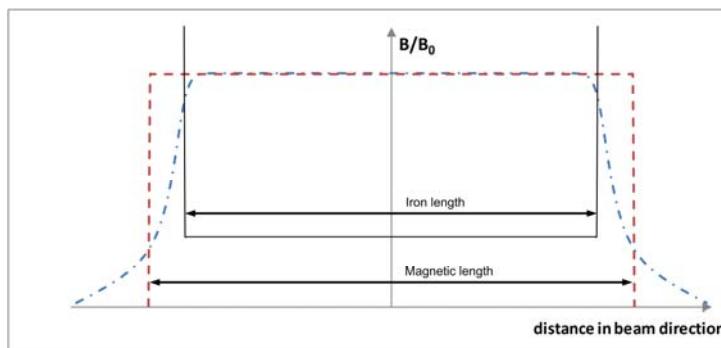


Fig. 15: Magnetic length – field distribution along the beam axis

4.1.7 Magnetic flux

The term magnetic flux Φ through a surface is defined as the integral of the normal component of the flux density over the cross-section area of this surface. In order to see whether there is any saturation issue in the iron we need to estimate the average flux density in the individual parts of the yoke.

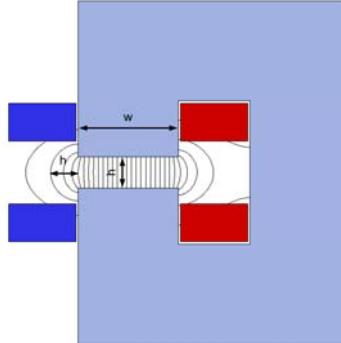


Fig. 16: Flux in the magnet aperture

This can be done by dividing the total magnetic flux by the cross-section area of the individual parts. As shown in Fig. 16 the flux entering the pole surfaces consists mainly of the useful flux in the gap, but for the correct calculation of the total flux in the return yoke we need to consider as well the stray flux entering on the sides of the poles. Again, a precise analytic calculation of the total flux is difficult, but for simple dipole geometry we can roughly estimate it by using the following relation.

Total flux in the return yoke is

$$\Phi = \int_A \mathbf{B} \cdot d\mathbf{a} \approx B_{gap}(w + 2h) l_{mag} \quad (9)$$

where h is the gap height and w the pole width.

4.1.8 Inductance

To size the power converters, we need to know the maximum current and the RMS current, and the dc power consumption, but also the voltage that the converter has to supply. The total required voltage is a function of the maximum current to excite the coils, the resistance and inductance of the coils, and the envisaged speed to reach the maximum field. The total voltage on a ramped magnet is given by

$$V_{tot} = RI + L \frac{dI}{dt} \quad (10)$$

where R is the total electrical resistance of the excitations coils in [Ω , ohm], L is the total inductance of the magnet in [H, henry], and $\frac{dI}{dt}$ the maximum current ramp rate in [A/s].

For a magnet cycled with $I = I_0 \sin(\omega t)$ the total voltage is

$$V_{tot} = RI_0 \sin(\omega t + \varphi) \quad (11)$$

where I_0 is the amplitude of the current, ω is the angular frequency and φ is the phase angle:

$$\varphi = \tan^{-1} \frac{L\omega}{R} \quad (12)$$

The coil resistance R can be easily calculated taking into account the conductor length l , the effective cross-section a , and the resistivity ρ of the conductor material:

$$R = \frac{l\rho}{a} . \quad (13)$$

This leads us to the question of how to calculate the inductance, which is not obvious. The inductance depends on the number of turns and on the coil geometry, but also on the geometry of the iron yoke surrounding the coils, which makes it more difficult to calculate correctly than for a simple cylindrical coil in free space. One possibility is to go via the stored energy U [J, joule] in the magnet

$$L = \frac{2U}{I^2} \quad (14)$$

so that Eq. (10) becomes

$$V_{\text{tot}} = RI + L \frac{dI}{dt} = RI + \frac{2U dI}{I^2 dt} . \quad (15)$$

Unfortunately, this method has the drawback that we now have to compute the stored energy correctly, which is itself not easy. As the stored energy in a magnet depends on the non-uniform field distribution in the gap, the coils, and the iron yoke, it is usually determined by numerical computation.

However, for the very simple case of a window-frame magnet with constant field in the gap as shown in Fig. 17(c), the stored energy can be estimated as follows:

$$U_{\text{gap}} = \frac{B^2}{2\mu_0} V_{\text{gap}} \quad U_{\text{coil}} = \frac{B^2}{6\mu_0} V_{\text{coil}} \quad U_{\text{yoke}} = \frac{B^2}{2\mu_r\mu_0} V_{\text{yoke}}$$

where V_{gap} , V_{coil} , and V_{yoke} are respectively the volumes of the gap, coil and yoke in [m^3]. Hence

$$U_{\text{magnet}} = U_{\text{gap}} + 2U_{\text{coil}} + U_{\text{yoke}} = \frac{B^2}{2\mu_0} \left(V_{\text{gap}} + 2 \frac{V_{\text{coil}}}{6} + \frac{1}{\mu_r} V_{\text{yoke}} \right)$$

4.1.9 Dipole types

Although the design and layout of a dipole magnet can be quite different from case to case depending on the application, we can nevertheless identify three standard families typically used in particle accelerators and transfer lines: the so-called C-magnet, the H-magnet and the O- or window-frame magnet as shown in Fig. 17. Looking into their characteristics there is no optimum solution; they all have their advantages and drawbacks. The choice for one or the other option is led by the constraints and requirements such as the function of the magnet, the available space, and the field quality. The following sections provide a short overview of these three main types pointing out their pros and cons.

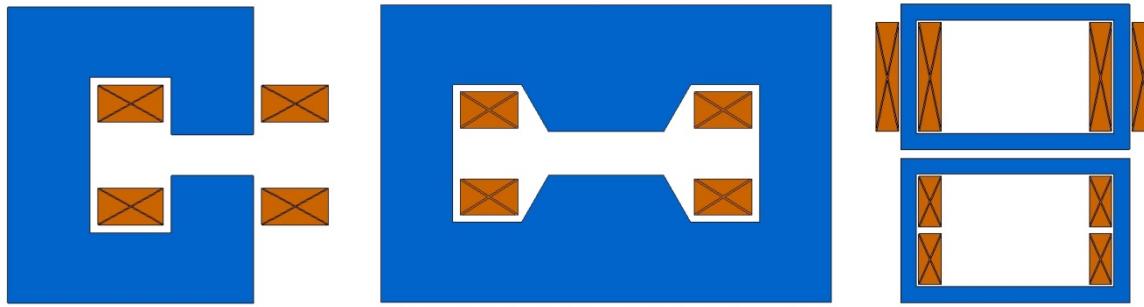


Fig. 17: Standard dipole types: C-magnet (a), H-magnet (b), and O-magnet (c)

4.1.9.1 C-magnet

The C-shaped magnet or C-magnet in Fig. 17 (a) provides a very good accessibility to the beam pipes, which makes it a perfect candidate for light sources where the synchrotron light has to be extracted all along the circumference of the synchrotron. Owing to its asymmetric layout this type is also suitable for injection and extraction regions or zones where adjacent beams are very close to each other like in the transfer lines of experimental areas.

The yoke volume and hence the weight of C-magnets is significantly higher than H-magnets with similar performance. The mechanical stability is less good compared to an H- or O-magnet since it has only one return leg and the attracting magnetic forces may lead to a movement of the poles when the magnet is pulsed. Transversal shims are usually required to achieve a decent field quality.

A drawback of the C-magnet is the asymmetrical field distribution in the gap. Unlike an H-magnet with two-fold symmetry around both axes, a C-magnet has only a one-fold symmetry. Because

$$NI = \oint \vec{H} \cdot d\vec{l}$$

has to be constant, the contribution to the integral in the iron has different path lengths, as shown in Fig. 18. The finite permeability will create lower field densities on the outside of the gap than on the inside which generates so-called ‘forbidden’ harmonics with $n = 2, 4, 6, \dots$. Typically, the dipole produces a gradient across the pole of 0.1% with respect to the central field. In addition, the harmonics change with saturation and display non-linear behaviour depending on the excitation level.

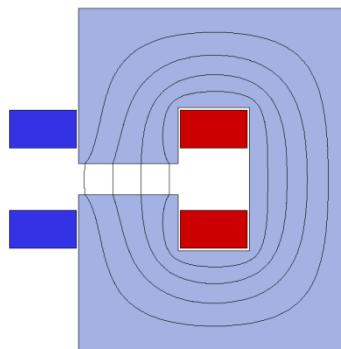


Fig. 18: Flux distribution in a C-magnet

4.1.9.2 H-magnet

The H-type magnet from Fig. 17(b) is used as standard in many accelerators and beam transfer lines. Access to the coils and beam pipes is poor, but they provide a good mechanical stability and a

symmetric field quality. The iron weight is reduced with respect to C-magnets and they are usually made of two parts to allow an easy installation of the coils and the vacuum chamber. Transversal shims are also required here to achieve a decent field quality.

4.1.9.3 Window-frame or O-magnet

If we reduce the pole heights of an H-magnet to zero we basically arrive at the so-called window-frame layout. It has similar characteristics to the H-magnet in terms of symmetry, weight, and mechanical stability, with the difference that the window-frame design provides a very homogenous field quality even without shims. As shown in Fig. 17(c), there are two basic versions of this magnet type which employ different coil designs. The image on the bottom represents a classical window-frame magnet with saddle coils (see Section 4.5.1), while the version on the top uses racetrack coils installed around the vertical legs of the return yokes. The latter is less efficient in terms of excitation: it requires more ampere-turns compared to the version with the saddle coils. In addition, it generates a lot of stray field in the surroundings of the magnet. However, coils can also be installed around the horizontal leg of the magnetic circuit adding a vertical bending function. Such combined horizontal/vertical magnets are often used as steering magnets due to their compact design, but their efficiency is low.

4.2 Quadrupole yoke design

4.2.1 Magnetic induction

Analogous to the dipole, the required quadrupole field gradient B' in [T/m] can be derived by using

$$B' = B\rho k \quad (16)$$

where k is the quadrupole strength in [m^{-2}].

4.2.2 Excitation current

The excitation current in a quadrupole can be calculated using similar considerations to those for a dipole. Choosing the integration path shown in Fig. 19 we get

$$NI = \oint \vec{H} \cdot d\vec{l} = \int_{s1} \vec{H}_1 \cdot d\vec{l} + \int_{s2} \vec{H}_2 \cdot d\vec{l} + \int_{s3} \vec{H}_3 \cdot d\vec{l}$$

For an ideal quadrupole, the gradient $B' = \frac{dB}{dr}$ is constant, $B_x = B'y$ and $B_y = B'x$.

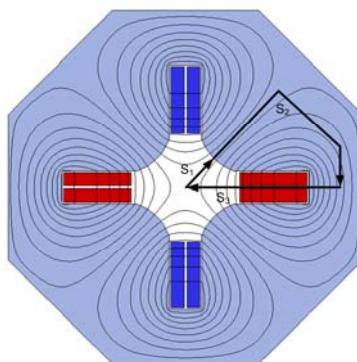


Fig. 19: Closed integration path in a quadrupole

The field modulus along the path s_1 can hence be written as

$$H(r) = \frac{B'}{\mu_0} \sqrt{x^2 + y^2} = \frac{B'}{\mu_0} r$$

Assuming that μ_{iron} is large we can neglect B in the part s_2 because the reluctance $R_{M,s2} = \frac{s_2}{\mu_{iron}}$ in the iron is small compared to the reluctance in the air gap. Since B_x on the x -axis ($y = 0$) is zero, the integral

$$\int_{s_2} \vec{H}_3 \cdot d\vec{l} = 0$$

too, so we can also ignore the contribution of B along path s_3 .

This leads to

$$NI \approx \int_0^R H(r) \cdot dr = \frac{B'}{\mu_0} \int_0^R r \cdot dr$$

and finally to

$$NI_{(per\ pole)} = \frac{B' r^2}{2\eta\mu_0} \quad (17)$$

The highest magnetic field appears at the pole vertex.

4.2.3 Magnetic length

The magnetic length for a quadrupole can be estimated by

$$l_{mag} = l_{iron} + 2r\kappa \quad (18)$$

where κ is a geometry specific constant (typically around 0.45) which can best be determined for a particular yoke geometry by numerical calculation.

It is interesting to note that the number of ampere-turns for a given gradient increases with the square of the quadrupole aperture and the dissipated power even with the power of four.

$$NI \propto r^2 \qquad P \propto r^4$$

This fact makes it more difficult to accommodate the necessary ampere-turns and coil cross-section in the iron yoke and to assure a sufficient cooling. To make space for the coil the hyperbola has to be truncated — digressing from the ideal pole profile. Depending on where the hyperbola is terminated, the resultant (allowed) higher order harmonics may affect the field quality in the aperture sufficiently to warrant correction.

4.2.4 Quadrupole types

In the same way as for the dipoles we can classify the different quadrupole layouts in several categories. The most common examples are presented in Fig. 20.

The standard quadrupole in Fig. 20(a) has 90° poles and provides very limited space for coils.

The image in Fig. 20(b) shows another standard quadrupole with parallel pole sides. It provides maximum space for coils, but the tendency to show saturation around the region of the pole roots

limits operation as a high-field quadrupole. Note that the entire stray flux entering all along the pole side has to pass through the pole root. This design is used when moderate field gradients are required.

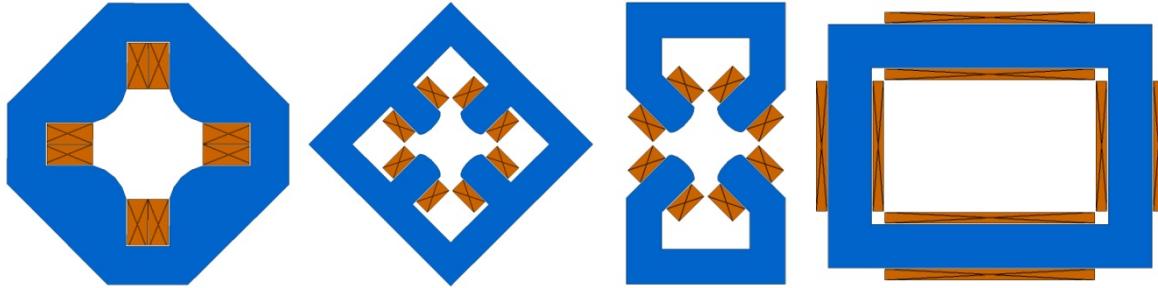


Fig. 20: Quadrupole types: standard type (a) and (b), Collins (c) and Panofsky (d)

A compromise between the two standard types is a quadrupole with tapered pole sides, which is not shown here. It combines the advantage of larger coil windows and less saturation in the pole roots, but is more complicated and costly to manufacture.

The so-called Collins or figure-of-eight quadrupole in Fig. 20(c) is a special type suitable for light sources and narrow beam lines because it provides an opening on the side allowing the extraction of beams or synchrotron light. It is obviously mechanically less stable, more complicated to produce, and therefore more expensive.

A more exotic type is the Panofsky quadrupole shown in Fig. 20(d). This type provides an excellent field quality; however it is used only as corrector magnet, because of its limited field strength. The Panofsky quadrupole looks like a window-frame dipole with horizontal and vertical coils, but is in reality an ironless magnet, since the magnetic field is determined by the current distribution in the copper conductors and not by the iron yoke.

4.3 Sextupole yoke design

4.3.1 Magnetic induction

Analogous to the dipole and the quadrupole, the differential field gradient B'' in [T/m²] of a sextupole can be computed

$$B'' = B\rho m \quad (19)$$

with m being the sextupole strength in [m⁻³].

4.3.2 Excitation current

To identify the required number of ampere-turns for a sextupole we chose the same approach as for

quadrupoles. For a sextupole, the field is parabolic and $B'' = \frac{d^2 B}{dr^2}$ is constant so that

$$H(r) = \frac{B''}{2\mu_0} r^2$$

leading to

$$NI = \oint \vec{H} \cdot d\vec{l} \approx \int_0^R H(r) \cdot dr = \frac{B''}{2\mu_0} \int_0^R r^2 \cdot dr$$

and

$$NI_{\text{per pole}} = \frac{B''r^3}{6\eta\mu_0} . \quad (20)$$

Analogous to the quadrupole, the ampere-turns in the sextupole increase with the third power of the aperture and the power dissipated in the coils rises with the sixth power of the aperture.

$$NI \propto r^3 \quad P \propto r^6$$

Fortunately, sextupolar fields are usually required to be much smaller than quadrupole fields.

4.4 Yoke materials

Magnetic circuits or yokes are made of magnetic steel. They can be built from massive iron or assembled from laminations. Historically, the primary choice for either of these techniques was whether the magnet was cycled or operated in persistent mode. Solid yokes support eddy currents and hence cannot be cycled or pulsed rapidly. To reduce or avoid eddy currents in pulsed operation the yoke has to be laminated.

Yokes machined from cast ingots require less tooling than for the stamping, stacking and assembling of laminated yokes. A major problem with massive yokes is the difficulty of providing magnets with similar magnetic performance. To assure identical characteristics within the accepted tolerances all yokes have to be built using the same melt. This requires very careful documentation.

Today's practice— even for dc operated magnets — is to use cold-rolled, non-grain-oriented (NGO) electro-steel sheets and strips (according to EN 10106). Although laminated yokes are labour intensive and require more and expensive tooling they offer a number of advantages:

- Magnetic and mechanical properties can be adjusted by final annealing
- Reproducible steel quality even over large productions
- Magnetic properties (permeability, coercivity) within small tolerances
- Homogeneity and reproducibility among the magnets of a series can be enhanced by selection, sorting or shuffling of the laminations according to their magnetic properties
- Organic or inorganic coating for insulation and bonding
- Material is usually cheaper

Table 2 summarizes typical material properties of cold-rolled, non-grain-oriented electro-steel. More detailed information on specific materials can be requested from the steel producers.

Table 2: Typical properties of cold-rolled, non-grain-oriented electro-steel

Property	Typical value
Sheet thickness	$0.3 \leq t \leq 1.5 \text{ mm}$
Density	$7.60 \leq \delta \leq 7.85 \text{ kg/dm}^3$
Coercivity	$H_c < 65 \text{ A/m}$
Coercivity spread	$\Delta H_c < \pm 10 \text{ A/m}$
Electrical resistivity at 20°C	$0.16 \text{ (low Si)} \leq \rho \leq 0.61 \mu\Omega\text{m} \text{ (high Si)}$

4.4.1 Permeability

At the beginning of this section we discussed the relation between the magnetic field strength H and the magnetic flux density B , which is defined for free space as

$$B = \mu_0 H , \quad (21)$$

where μ_0 is a universal constant with the value $4\pi \cdot 10^{-7}$ Vs/Am. For the magnetic flux density in a material Eq. (21) becomes

$$B = \mu H$$

where μ is called the absolute permeability which is material specific. Because

$$\mu = \mu_0$$

for free space, we can relate the permeability of matter to the permeability of free space by

$$\mu = \mu_r \mu_0 \quad (22)$$

introducing μ_r as the dimensionless relative permeability, which characterizes the magnetic behaviour of materials. We can distinguish between three main categories of materials:

- diamagnetic materials ($\mu_r < 1$),
- paramagnetic materials ($\mu_r > 1$), and
- ferromagnetic materials ($\mu_r \gg 1$).

For the construction of electromagnets, only the third category is important. Several examples of commonly used steel grades and their relative permeability are presented in Fig. 21.

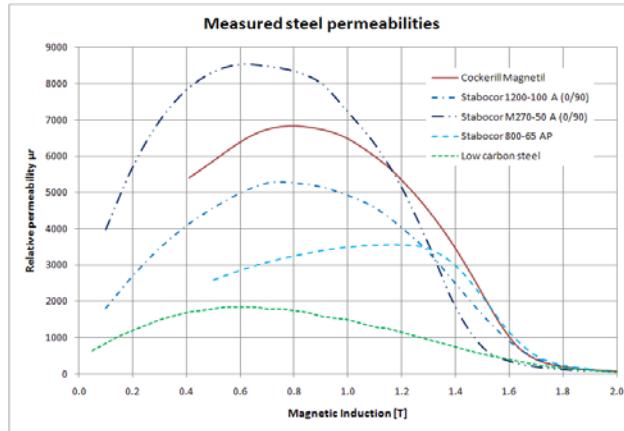


Fig. 21: Permeability of different steel grades

4.4.2 Magnetic polarization

Ferromagnetic materials show a non-linear correlation between the field strength H and the flux density B . The total flux density in the material is the sum of the flux density in free space $\mu_0 H$ and the magnetic polarization J in [T] and is described by the equation

$$B = \mu_0 H + J = \mu_r \mu_0 H . \quad (23)$$

The magnetic polarization J for specific materials is typically presented in tables or graphs by measured data. In addition to the non-linear behaviour, cold- and hot-rolled steel have material characteristics which can be anisotropic to a high degree. This anisotropy, in particular for the permeability can be partly cured by final annealing, but remains to a certain extent and cannot be neglected, so it has to be considered in the magnetic design. Fig. 22 shows the anisotropic polarization and permeability of cold-rolled electro-steel (grade 1200 – 100A) after final annealing.

4.4.3 Hysteresis, remanence and coercivity

On account of complicated material internal processes (movements and growth of magnetic domains) we can observe a hysteresis, which means that the flux density $B(H)$ as a function of the field strength is different when increasing and decreasing excitation. This behaviour is shown in Fig. 23.

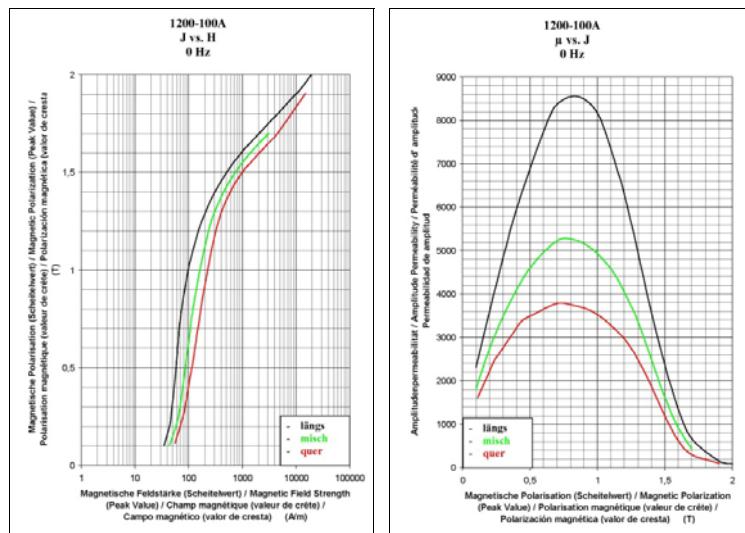


Fig. 22: Anisotropic polarization (a) and permeability (b); data source: Thyssen/Germany

When the current is switched off, some magnetic polarization of the iron remains: this is called remanent field or magnetic remanence B_r . The width of the hysteresis curve is determined by the coercive force or coercivity H_c . The quantity H_c is defined as the value of field strength that reduces the magnetic flux density in the steel to zero. Materials having $H_c < 1000$ A/m are called soft magnetic materials, e.g., electro-steel, those with $H_c > 1000$ A/m are called hard magnetic, e.g., permanent magnets.

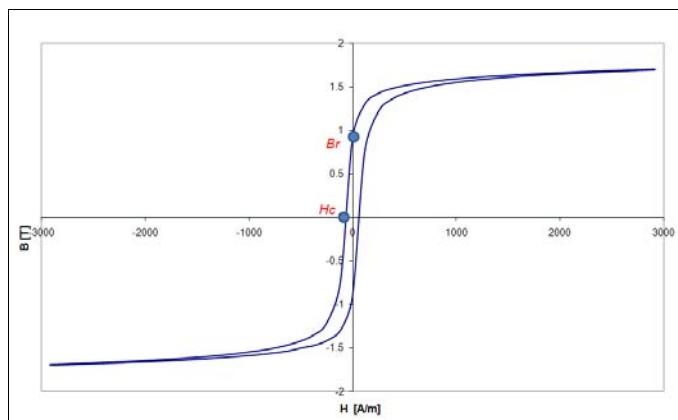


Fig. 23: Hysteresis curve of electro-steel (grade 1200 – 100 A)

In a continuous ferromagnetic core, as is the case of a transformer, the residual field is entirely determined by the remanence B_r . In a magnet where the highest reluctance appears in the magnet gap, the residual field is mainly determined by H_c . If we take a magnet which has been excited to a certain field level and we switch off the current in the coils we get

$$\text{NI} = \oint \vec{H} \cdot d\vec{l} = \int_{\text{gap}} \vec{H}_{\text{gap}} \cdot d\vec{l} + \int_{\text{yoke}} \vec{H}_c \cdot d\vec{l} = 0 \quad (24)$$

and so

$$B_{\text{residual}} = -\mu_0 H_c \frac{\lambda}{g} \quad (25)$$

To set the residual field to zero, a negative current must be sent through the coils. In practice it is often more convenient to set a zero field in the magnet by running it through a certain number of so-called demagnetization cycles. In normal operation, the magnet is always cycled to its maximum value, irrespective of the required field, to ensure that hysteresis effects are reproducible.

4.5 Coil design

In the previous sections it has been shown how to determine the necessary ampere-turns. In this part we will see how to choose a current density, the number of turns and dimensions of the coil.

The design of the coils is not completely independent from the layout of the yoke. Optimizing the coils, e.g., for low power consumption, is usually at the expense of a larger yoke cross-section. It is the duty of the magnet designer to find the right compromise between a good coil design and a good yoke design. A high-quality coil design unifies low electrical power consumption, sufficient cooling performance, adequate insulation thickness, and moderate material and manufacturing costs. To reach this goal, and to achieve a satisfactory overall magnet performance, requires several iterations.

The coil design sequence can be divided into steps:

- Selecting an adequate coil type
- Calculating power requirements
- Cooling circuit configuration
- Selecting the conductor dimensions
- Optimization

4.5.1 Standard coil types

As in the case of the yoke, coil layouts can also differ, depending on the application. Fig. 24 illustrates the types most frequently used for normal-conducting accelerator magnets.

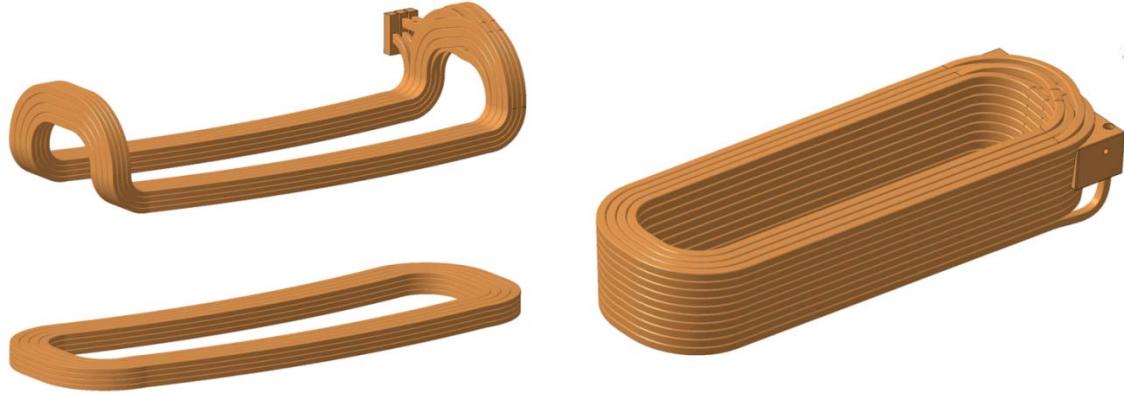


Fig. 24: Standard coil types: bedstead coil (a), racetrack coil (b) and quadrupole coil (c)

Racetrack coils: This coil type is relatively easy to manufacture and therefore the least expensive. It is commonly used in magnets with poles (C-magnet, H-magnet). To facilitate production and installation, the coils are often manufactured in pancakes, in particular for C-shape magnets.

Bedstead or saddle coils: They are more complicated to wind and require a complex impregnation mould, which makes them more expensive. This type is used for O-type and H-type magnets. The bent coil heads allow filling the entire coil window with conductor material and leaving space for the beam pipes and the magnet ends.

Quadrupole coils: This coil can be used in quadrupoles with parallel or slightly tapered poles. A particularity shown here are the integrated terminals for water and electricity.

4.5.2 Power requirements

Once the coil type has been selected we need to determine the power requirements. For this we assume that the magnet cross-section and the yoke length are known. The total dissipated power for the individual magnet types can be calculated accordingly:

$$\text{dipole: } P_{\text{dipole}} = \rho \frac{Bh}{\eta\mu_0} jl_{\text{avg}} \quad (26)$$

$$\text{quadrupole: } P_{\text{quadrupole}} = 2\rho \frac{Br^2}{\eta\mu_0} jl_{\text{avg}} \quad (27)$$

$$\text{sextupole: } P_{\text{sextupole}} = \rho \frac{Br^3}{\eta\mu_0} jl_{\text{avg}} \quad (28)$$

where ρ is the resistivity in $[\Omega\text{m}]$, l_{avg} is the average turn length in $[\text{m}]$ (a useful approximation is $2.5 l_{\text{iron}} < l_{\text{avg}} < 3 l_{\text{iron}}$ for racetrack coils).

The current density j in $[\text{A}/\text{m}^2]$ is defined as

$$j = \frac{NI}{f_c A} = \frac{I}{a_{\text{cond}}} \quad (29)$$

with a_{cond} being the conductor cross section in $[\text{m}^2]$, A being the coil cross section in $[\text{m}^2]$ and f_c being a dimensionless geometric filling factor ($= \frac{\text{net conductor area}}{\text{coil cross section}}$) taking into account insulation

material, cooling duct and the conductor edge rounding. It is interesting to note that for a constant geometry, the power loss P is proportional to the current density j .

4.5.3 Number of turns

The power which we have determined above can be divided into a voltage and a current according to

$$P = VI$$

With the help of the following basic relations

$$R_{magnet} \propto N^2 j$$

$$V_{magnet} \propto Nj$$

$$P_{magnet} \propto j$$

we can choose a number of turns N to match the impedances of the power converter.

A large number of turns implies low current but high voltage which consequently requires thicker insulation for both coils and cables, which gives rise to a poor filling factor. A positive effect is that the transmission power losses are kept low even across long distances between the power converter and the magnets. The choice of coils with many turns is therefore made primarily for magnets with moderate magnetic field strength which are powered individually.

A small number of turns implies high current but low voltage. The drawbacks are large terminals and conductor cross-section. Advantages are a better conductor filling factor in the coils, smaller coil cross-sections and less stringent demands on the coil and cable insulation. Since the transmission power losses are high, this solution is chosen when many magnets have to be electrically connected in series and the distance from one magnet to the next is relatively short, such as in the case of bending magnets in a synchrotron. In this case to have many turns would lead to unreasonably high voltages between the coils and the magnet yokes increasing the risk of short circuits. The transmission power loss can be handled by using water-cooled cables or rigid bus bars with large cross-sections.

4.6 Cooling

The electrical power which is dissipated in the coils has to be removed from the magnets otherwise overheating can seriously damage the coil insulation and cause short circuits between the coil conductor and the surrounding equipment which is usually on ground potential. In the field of normal-conducting magnets, we distinguish between two different cooling techniques: air cooling and water cooling. Sometimes they are also referred to as ‘dry cooling’ and ‘wet cooling’.

4.6.1 Air cooling

Air cooling by natural convection is suitable only for low current densities. This limits the application to magnets with moderate field strength like correctors or steering magnets. As a rule of thumb, the maximum current density for voluminous coils which are almost entirely enclosed in the magnet yoke should not exceed 1 A/mm^2 . For small, thin coils, current density can be higher, but below 2 A/mm^2 .

A precise thermal study of air-cooled magnets by analytic means is difficult if not impossible. Air cooling is a combination of convection, radiation and heat conduction and depends on coil geometry, coil surface (roughness, material, colour), thermal contact to the surrounding materials, etc. Detailed analysis of the thermal behaviour, if needed, would require numerical computations or measurements. Information on air cooling and cooling in general can be found in the relevant text books in the bibliography at the end of this paper.

Air-cooled coils can be made of round, rectangular, or square wires. They are commercially available in various grades and dimensions and can be ordered blank or pre-impregnated with varnish ($0.02 \leq t \leq 0.1 \text{ mm}$) or half-overlapped polyimide (Kapton®) tape ($0.1 \leq t \leq 0.2 \text{ mm}$). Depending on the winding precision, the insulation thickness, and the conductor cross-sections a filling factor

between 0.63 (round) to 0.8 (rectangular) can be obtained. The outer or ground insulation is typically made by epoxy impregnated glass fibre tapes of thickness between 0.5 mm and 2 mm.

The cooling performance of air-cooled coils can be enhanced by mounting an appropriate heat sink with enlarged radiation surface or by forced air flow (cooling fan).

4.6.2 Water cooling

There are two methods of water or wet cooling: direct and indirect. The latter is of minor importance and rarely used, although it has the advantage that normal tap water can be used as coolant, which does not require cooling plants with water treatment for the supply of demineralized water. In the absence of such infrastructure, indirect cooling should be considered as a possible alternative, even though it implies a more complex coil design. In a situation where air cooling is just at the limit, the mounting of an external heat sink cooled with tap water can enhance the cooling performance keeping the thermal load within limits or permitting slightly higher current densities. However, indirect cooling is seldom used, so here we focus on the engineering and construction of direct water-cooled coils.

The current density in direct water-cooled coils can be typically as high as 10 A/mm^2 . This is a conservative value that can be easily realized with standard coil models. It is a good compromise assuring a high level of reliability during operation and a compact coil layout. Although current densities of 80 A/mm^2 can be attained for specific applications, e.g., septum magnets, it is not recommended for standard magnets because the reliability and lifetime of the coils is significantly reduced. High current densities require a sophisticated cooling circuit design with multiple parallel circuits per coil — even single turn cooling — and high coolant velocities increasing the risk of erosion.

Standard water-cooled coils are wound from rectangular or square copper or aluminium conductor with a central cooling duct for demineralized water as shown in Fig. 25. The inter-turn and ground insulation is provided by one or more layers of half-lapped glass fibre tape impregnated in epoxy resin. Inter-turn insulation thickness is normally between 0.3 mm and 1.0 mm, the ground insulation thickness should be between 0.5 mm and 3.0 mm depending on the applied voltage.

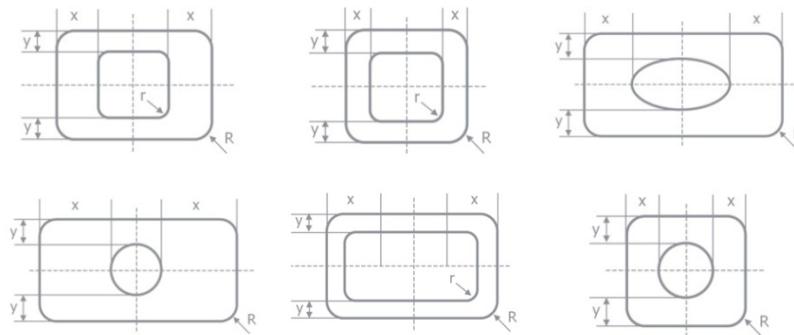


Fig. 25: Hollow conductor profiles for water-cooled coils

4.6.3 Conductor materials

Conductor materials which are most commonly used for the construction of normal-conducting are aluminium and copper. Both materials are available in different grades and purities. While aluminium was a good alternative in the past, since it was much cheaper, nowadays copper coils are more state-of-the-art. The material costs today are similar for a copper coil and an aluminium coil provided they have the same electrical resistance, but aluminium coils have a larger cross-section. On the other hand, aluminium is easier to form since it does not work harden.

Table 3 gives two examples of typical standard materials together with their main characteristics. Magnet designers and engineers who look for more detailed information and other available grades should consult suppliers, material databases, or international standards.

Table 3: Typical properties of conductor materials

Property	Aluminium	Copper (OF grade)
Purity	99.7%	99.95%
Resistivity at 20°C	28.3 nΩ m	17.2 nΩ m
Thermal resistivity coefficient	0.004 K ⁻¹	0.004 K ⁻¹
Density	2.70 kg/dm ³	8.94 kg/dm ³
Thermal conductivity	237 W/m K	391 W/m K
Approx. price	4.7 €/kg	11 €/kg

When winding a coil — regardless of whether it is made of aluminium or copper — special care has to be taken to avoid small bending radii. A tight bending radius increases the risk of insulation damage, decreases the cross-section of the cooling duct and leads to an increase of the outer conductor dimensions, which is known as the ‘keystone effect’ or ‘keystoning’.

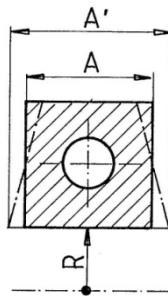


Fig. 26: Keystone effect

Keystoning, as shown in Fig. 26, is caused by the horizontal growth of the conductor on the inside radius of the coil when bending the conductor around the horizontal axis. The effect of keystoning has been determined in the past empirically. For a bending radius of three times the conductor width we can expect a keystoning of 3.6%:

$$R = 3 \cdot A \Rightarrow \frac{\Delta A}{A} = 3.6\%$$

where A means the conductor dimension parallel to the bending axis.

For coils with many turns, the cumulative effect of keystoning can lead to a significant enlargement of the coil dimensions in the bending regions and consequently to problems for the installation of the coils in the yoke. In principle, we can ignore the effect of keystoning by systematically choosing a bending radius four times larger than the conductor width.

4.6.4 Cooling parameters

Water cooling is necessary whenever large coil cross-sections have to be avoided and higher current densities are indispensable. The construction of water-cooled coils requires hollow conductors with a cooling channel. Choosing the correct parameters and dimensions, such as number of cooling circuits, the size of the cooling channel and the flow rate, is not easy and requires several iterations to arrive at a satisfactory solution. This section addresses the calculation and design of an efficient cooling circuit.

Before entering into the subject, I will give a few recommendations and canonical values, which should help the magnet designer with his choices. Please note that these values are rules of thumb and should be critically cross-checked as to their validity and applicability in the particular situation.

As mentioned in the previous section, the current densities for water-cooled coils should be kept between 2 A/mm^2 and 10 A/mm^2 . Below these values, air cooling is probably sufficient; exceeding this threshold implies more complicated and costly designs.

The pressure drop which is typically provided by modern cooling plants ranges between 0.1 MPa and 1.0 MPa , which corresponds to 1 to 10 bar. Advanced cooling stations can supply water with a pressure up to 2.0 MPa . A low pressure drop can be compensated to a certain extent by a higher cooling flow and a more complex (and expensive) coil design with several cooling circuits in parallel.

The velocity of the cooling medium — in general demineralized water — should be sufficiently high to guarantee a turbulent flow but low enough ($u_{\text{avg}} \leq 5 \text{ m/s}$) to avoid erosion and vibration.

A maximum permitted temperature of less than 60°C on the coil surfaces was found to be good practice. It avoids or reduces accelerated ageing of insulation materials in particular in presence of ionizing radiation. Assuming a water inlet temperature of less than 30°C , this corresponds to a maximum water temperature rise of 30°C , also keeping thermal stress within limits. For some applications a high mechanical stability and consequently a good thermal stability is essential. In such cases, the maximum temperature rise has to be reduced accordingly.

In what follows it is assumed that the cooling pipes are always long, mostly straight and smooth inside without perturbations and that the flow is turbulent, meaning a high Reynolds number. We also assume good heat transfer from conductor to coolant, with the temperature of the inner surface of the conductor equal to that of the coolant, and that the conductor is isothermal over its cross-section.

The first step is to determine how the velocity of a coolant depends on the dimensions of the cooling pipe and the pressure drop along it. The pressure drop through a water circuit is known to be

$$\Delta p = f \frac{l}{d} \frac{\delta u_{\text{avg}}^2}{2} \quad (30)$$

with p being the pressure in [$\text{Pa}, \text{N/m}^2$], f a dimensionless friction factor, l and d the cooling circuit length in [m] and diameter in [m], δ the coolant mass density in [kg/m^3], which for water is $1000 \text{ kg/m}^3 = 1 \text{ kg/litre}$, and u_{avg} the average coolant velocity in [m/s].

The friction factor f depends on the Reynolds number Re

$$\text{Re} = \frac{u_{\text{avg}} d}{v} \quad (31)$$

with v describing the temperature-dependent kinematic viscosity of a coolant. For our purpose we can assume it as constant: $9.85 \cdot 10^{-7} \text{ m}^2/\text{s}$ for water at 21°C .

From the text books, the friction factor f for laminar flow (Reynolds number less than 2000) is

$$f = \frac{64}{\text{Re}}$$

For our purpose, the flow is turbulent ($\text{Re} > 4000$), and the friction factor f is transcendental:

$$\frac{1}{\sqrt{f}} = -2 \log_{10} \left(\frac{\varepsilon}{3.7d} + \frac{2.51}{\text{Re} \sqrt{f}} \right) \quad (32)$$

where ε is a measure for the roughness of the cooling channel (typically $1.5 \cdot 10^{-3}$ mm). We can reformulate Eq. (30) to express the coolant velocity u_{avg}

$$u_{avg} = \sqrt{\frac{2 \Delta p d}{\delta f l}} \quad (33)$$

The friction factor f in Eq. (33) depends on the $Re(u_{avg})$, which means that to solve for u_{avg} requires an iterative process. This is also shown in text books on heat transfer. Substituting the Reynolds number in Eq. (32) by

$$Re = \frac{d}{v} \sqrt{\frac{2 \Delta p d}{\delta f l}} \quad (34)$$

yields

$$u_{avg} = -2 \sqrt{\frac{2 \Delta p d}{\delta f l}} \log_{10} \left(\frac{\varepsilon}{3.7d} + \frac{2.51}{\frac{d}{v} \sqrt{\frac{2 \Delta p d}{\delta f l}}} \right) \quad (35)$$

Assuming that water is used as cooling fluid, this complicated expression can be simplified to

$$u_{avg} \approx 0.3926 \cdot d^{0.714} \left(\frac{\Delta p}{l} \right)^{0.571} \quad (36)$$

The heat absorbed by coolant medium across a heated surface is

$$P = \dot{m} c_p \Delta T \quad (37)$$

where P is the power dissipated in [W], c_p is the heat capacity in [W s/kg °C], which is $4.19 \cdot 10^3$ W s/kg °C for water, ΔT is the temperature increase in [°C] and \dot{m} is the mass flow in [kg/s]. Writing

$$\dot{m} = \delta Q \quad (38)$$

with Q being the coolant flow rate in [litre/s], the expression for the necessary flow Q to remove the dissipated heat P by the selection of the maximum permissible temperature ΔT is given by

$$Q = \frac{P}{\delta c_p \Delta T} \quad (39)$$

Using water as cooling medium, Eq. (39) can be simplified and rewritten to

$$Q_{water} = 2.388 \cdot 10^{-4} \frac{P}{\Delta T} \quad (40)$$

Knowing that the coolant flow inside a round tube with a bore diameter d is

$$Q = u_{avg} \frac{\pi d^2}{4} 10^3 \quad , \quad (41)$$

the temperature increase ΔT in a water-cooled circuit due to dissipated power P is

$$\Delta T = 3.04 \cdot 10^{-7} \frac{P}{u_{avg} d^2} \quad (42)$$

Two further relations affect cooling performance and should be considered in the system layout:

Pressure drop is inversely proportional to the third power of the number of circuits K_w per coil:

$$\Delta p \propto \frac{1}{K_w^3}$$

This implies that for a given flow, the pressure drop is reduced by a factor of eight by doubling the number of cooling circuits.

The pressure drop is inversely proportional to the fifth power of the cooling channel diameter d :

$$\Delta p \propto \frac{1}{d^5}$$

This implies that an increase of the cooling channel diameter by a small factor can reduce the required pressure drop significantly.

4.6.5 Cooling circuit design

In the previous section we derived the formulas required to calculate the main parameters of a cooling circuit. Finding a good solution for a cooling layout is an iterative and lengthy exercise and there are certainly different methods which lead to a satisfactory design. In this section I will demonstrate a way to start with the design of a cooling circuit for a dipole magnet, which has always worked fine for me. Other magnet types such as quadrupoles or sextupoles can be treated in a similar way by modifying the parameters accordingly. The steps described contain several approximations and assumptions which may not appear very precise, but they should help setting a starting point for an iterative design process. Depending on how demanding the requirements and constraints are, more or less fine-tuning on initial assumptions will be required in a second or third iteration. To start with we suppose that current density j , power P , current I , and the number of turns N have already been determined.

First, the number of layers m in the coil and the number of turns per layer n has to be selected. Often one needs to round up the number of turns N per coil to get reasonable numbers for m and n . The coil height c and coil width b can be defined more or less arbitrarily, the limiting factor being the available space in the coil window of the yoke. An aspect ratio of $c:b$ between 1:1 and 1:2 should be chosen, and the packing factor f_c somewhere between 0.6 and 0.8. The total coil cross-section area A is

$$A = b c = \frac{N I}{j f_c} \quad (43)$$

The following formula has been found to be a good approximation to estimate the average length of a conductor turn for dipole magnets with poles:

$$l_{avg} = \text{pole perimeter} + 8 \times \text{clearance between pole and coil} + 4 \times \text{coil width}$$

For the first iteration we can assume a coil with a single cooling circuit. If it turns out to be insufficient, the number of circuits per coil has to be increased in the second iteration. On the other hand, one may find that the required water pressure is far less than the available pressure. In this case

one can simplify the magnet design by connecting several coils in series or reducing the size of the cooling duct, such that more conductor cross-section is available to carry the electrical current.

The total length of the cooling circuit, assuming that we can neglect the connection hoses or tubes since they are short compared to the circuit length in the coils, is expressed by

$$l = \frac{K_c N l_{avg}}{K_w} \quad (44)$$

where K_c is the number of coils per magnet and K_w is the number of water circuits per magnet. In the case of our former assumption to have one circuit per coil, K_w is equal to K_c .

The next step is to select the maximum permissible temperature rise ΔT and the available pressure drop Δp and to calculate the cooling hole diameter d

$$d = 5.59 \cdot 10^{-3} \left(\frac{P}{\Delta T K_w} \right)^{0.368} \left(\frac{l}{\Delta p} \right)^{0.21} \quad (45)$$

If the results from Eq. (43) lead to a diameter of the cooling hole that is too large or to small, either the pressure drop Δp or the number of cooling circuits K_w can be changed before repeating the calculation. Once a satisfactory solution is found one can continue to determine the conductor area by

$$a = \frac{I}{j} + \frac{d^2 \pi}{4} + r_{edge}(4 - \pi) \quad (46)$$

where r_{edge} is the edge rounding radius of the conductor profile. Knowing the conductor area a , the conductor dimensions and insulation thickness can be fixed. Now it is time to verify if the resulting coil dimensions, the number of turns N , the coil current I , the coils resistance R and the temperature rise ΔT are still compatible with the initial assumptions and requirements. If this is not the case, another iteration with a new set of parameters has to be launched.

The last step is to compute the coolant velocity and the coolant flow using Eq. (36) and Eq. (40) and verify, using Eq. (31), whether the Reynolds number corresponds to turbulent flow ($Re > 4000$).

4.6.6 Cooling water properties

For the cooling of hollow conductor coils, with the exception of indirect cooled coils, demineralized water has to be made available. Since the water quality is essential for the performance and the reliability of the coil, the following typical water properties should be guaranteed:

- Water resistivity higher than $0.1 \cdot 10^6 \Omega m$
- pH-value between 6 and 6.5
- dissolved oxygen below 0.1 ppm (parts per million)

In addition to filters to be installed in the water circuits close to the magnet to avoid an obstruction of the cooling duct by particles, loose deposits and grease, a constant monitoring of the water quality is indispensable. A poor water quality neglecting the above recommendations might lead sooner or later to failures like electrical short circuits and leaks in the water circuit due to corrosion and erosion effects.

4.7 Cost estimate

Preparing a correct cost estimate for a magnet project requires a lot of experience and a good knowledge of the market. While the material prices are well known, the evaluation of the costs of manufacturing a magnet is not a precise science and depends on many different parameters. Past experience has shown that prices from the lowest to the highest bid may vary by up to a factor three.

In the framework of these lectures we can only give an order of magnitude indication of prices, as shown in Table 4. The inexperienced magnet designer should consult either more experienced colleagues and/or magnet suppliers to draft realistic estimation of costs.

Table 4: Cost indication for standard magnets (valid for 2010)

Item	Cost indication
Production-specific tooling	5000 – 15 000 €/tooling
Steel sheets	1.0 – 1.5 €/kg
Copper conductor	10 – 15 €/kg
<i>Yoke manufacture:</i>	
Dipoles (> 1000 kg)	6 – 10 €/kg
Quadrupoles, sextupoles (> 200 kg)	50 – 80 €/kg
Small magnets	up to 300 €/kg
<i>Coil manufacture:</i>	
Dipoles (> 200 kg)	30 – 50 €/kg
Quadrupoles, sextupoles (> 30 kg)	65 – 80 €/kg
Small magnets	up to 300 €/kg
Contingency	10 – 20%

4.7.1 Cost optimization

One of the main goals in magnet design should be to make an economic design for minimum total cost over the magnet lifetime. Total cost can be divided into capital investment and running costs as shown in Fig. 27. The cost factor ‘power’ is important as it enters into both the capital and running costs.

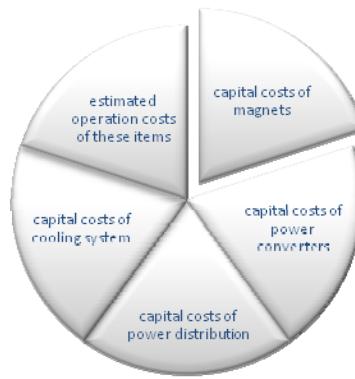


Fig. 27: Magnet cost factors

While it is difficult to estimate the absolute magnet costs precisely, the relative cost of design variants can be compared and optimized relatively easily. An interesting method to minimize overall

cost by optimizing scaling parameters, e.g., magnet length and the current density is described in Ref. [2].

As demonstrated in this paper, a decrease of the current density leads to increases of the coil cross-section and increases of the coil and yoke material and manufacturing cost whilst decreasing the operation costs. The advantages of low current densities are the lower power losses, reducing the power consumption and power converter size and leading to less heat dissipated into the machine tunnel. High current density gives the advantages of smaller coils, smaller magnets, and lower investment cost.

5 Conclusions

I have shown in this paper how to start a magnet design by using simple analytical and empirical formulas. Such a first draft is often sufficient to validate the feasibility of a design proposal and to derive a list of the main magnet parameters. In addition it provides a solid foundation for the following steps where this preliminary design has to be refined by means of numerical programs. With the help of two- and three-dimensional FEM models one has to look in more detail into subjects like end field effects, field homogeneity and possible transient effects due to time variant parameters before ending with a final optimization.

Acknowledgements

I wish to thank the CAS organizing committee for giving me the opportunity to present these lectures, which was a wonderful experience for me. In preparing these lectures and the notes I have profited greatly from the books and articles quoted in the bibliography, in particular from those of N. Marks in the CAS Proceedings, J. Tanabe and G. E. Fischer. Special thanks go to T. Taylor who was kind enough to review this paper by making valuable suggestions. Finally, I would like to thank my colleagues F. Luiz and C. Siedler for helping me to produce the nice figures which make the text more lively and comprehensible.

References

- [1] G. E. Fischer, Iron dominated magnets, 4th Summer School on Particle Accelerators, Fermilab, Batavia, IL, 1984, M. Dienes and M. Month (Eds.), AIP Conference Proceedings 153 (1987), pp. 1120–1227 [SLAC-PUB-3726 (1985)].
- [2] G. Brianti and M. Gabriel, Basic expressions for evaluating iron core magnets – A possible procedure to minimize their cost, CERN/SI/Int. DL/70-10 (1970).

Bibliography

Proceedings of the International Conference on Magnet Technology, MT-1 to MT-20.

CAS: Proceedings of the Fifth General Accelerator Physics Course, Jyväskylä, Finland, 1992, S. Turner (Ed.), CERN 94-01 (1994).

CAS: Proceedings of the Specialist Course on Magnetic Measurement and Alignment, Montreux, Switzerland, 1992, S. Turner (Ed.), CERN 92-05 (1992).

CAS: Proceedings of the Specialist Course on Measurement and Alignment of Accelerator and Detector Magnets, Anacapri, Italy, 1997, S. Turner (Ed.), CERN 98-05 (1998)

J. T. Tanabe, *Iron Dominated Electromagnets* (World Scientific Publishing, Singapore, 2005).

G. Parzen, *Magnetic Fields for Transporting Charged Beams*, BNL 50536 (1976).

G. Schnell, *Magnete: Grundlagen, Aufbau, Anwendungen* (Thiemig Verlag, Munich, 1973), German language.

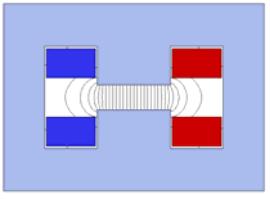
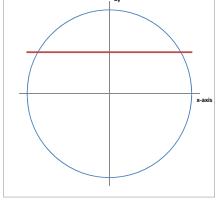
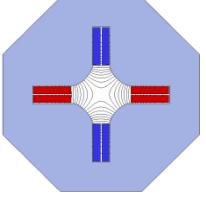
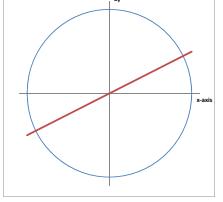
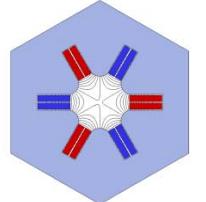
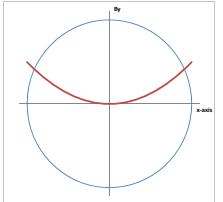
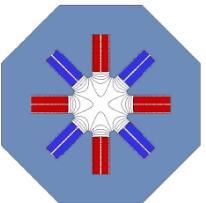
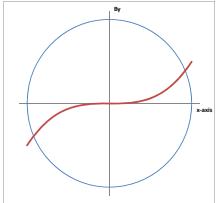
S. Russenschuck, *Field Computation for Accelerator Magnets: Analytical and Numerical Methods for Electromagnetic Design and Optimization* (Wiley-VCH Verlag, Weinheim, 2010).

K. Wille, *Physik der Teilchenbeschleuniger und Synchrotronstrahlungsquellen: eine Einführung* (Teubner Verlag, Stuttgart, 1996), German language.

J. H. Lienhard IV and J. H Lienhard V, *A Heat Transfer Textbook* (Phlogiston Press, Cambridge, Mass., 2003).

H. E. Knoepfel, *Magnetic Fields: A Comprehensive Theoretical Treatise for Practical Use* (Wiley, New York, 2000).

Appendix A: Basic magnet types

Magnet type	Field distribution	Pole equation	Flux density (B_y)
		$y = \pm r$	$B_y = a_1 = B_0 = \text{const.}$
		$2xy = \pm r^2$	$B_y = a_2 x$
		$3x^2y - y^3 = \pm r^3$	$B_y = a_3(x^2 - y^2)$
		$4(x^3y - xy^3) = \pm r^4$	$B_y = a_4(x^3 - 3xy^2)$

Eddy currents in accelerator magnets

G. Moritz

GSI, Darmstadt, Germany

Abstract

This paper covers the main eddy current effects in accelerator magnets — field modification (time delay and field quality) and resistive power losses. In the first part, starting from the Maxwell equations, a basic understanding of the processes is given and explained with examples of simple geometry and time behaviour. Useful formulas are derived for an analytic estimate of the size of the effects. In the second part the effects in real magnets are analysed and described in comparison with numerical and measured results. Finally, based on the previous parts, design recommendations are given regarding how to minimize eddy current effects.

1 Introduction

Eddy currents play an important role in magnets. As the title already indicates, this lecture will be restricted to eddy currents in accelerator magnets.

1.1 Definition

According to Faraday's law a voltage is induced in a conductor loop, if it is subjected to a time-varying magnetic flux. As a result, current flows in the conductor if there exists a closed current path.

'**Eddy currents**' appear in extended conducting media if these media are subjected to time varying magnetic fields. They are now distributed in the conducting media.

A comprehensive treatment is given in Ref. [1], and Tegopoulos et al.

1.2 Effects

We can observe different effects due to these induced eddy currents. They create

- a magnetic field
 - superimposed on the causative time-varying field
 - delayed due to Lenz's law
- electrical power loss
- Lorentz forces, due to the interaction with the magnetic field

1.2.1 Desired effects

Many devices use these effects for their specific applications. Examples are induction heating, magnetic shielding, the levitated train, dampers and brakes [2].

1.2.2 Undesired effects

The above mentioned eddy current effects are mostly undesired in accelerator magnets.

The magnetic field, created by the eddy currents, destroys the high field quality in the ‘good field region’ that is required for this type of magnet. Moreover the field change is delayed, the field lags behind, and one has to wait for the eddy currents to die out before one can reach a stable situation for the beam. For long time constants (of the order of seconds) this can be a real nuisance and reduce the performance of the accelerator.

Power loss due to eddy currents leads to heating of the (conductive) magnet components, which is especially detrimental for the case of superconducting magnets operating at cryogenic temperatures of a few degrees Kelvin.

Especially in the case of fast changes of high magnetic fields the Lorentz forces must not be neglected. They may lead to stresses in the material beyond the elastic limit.

1.3 Goal and outline of the lecture

The primary goal of this lecture is therefore to provide a good understanding of the physics of eddy currents. Analytical formulas will be given to estimate the most common effects under certain assumptions. The results will be compared with numerical calculations.

The effects of eddy currents in different components of a magnet, such as yoke, mechanical structure, coil and beam pipe will be considered with the goal of attaining an optimal magnet design in the presence of a time-varying magnetic field. In most cases this requires a reduction of induced eddy currents as much as possible, in order to minimize their effects.

2 Basics

In this section we will develop a basic understanding of the eddy current processes, first by using the diffusion approach and second by analytical solving of Maxwell’s equations directly, for some selected examples.

2.1 Maxwell equations

In this lecture we consider only the quasi-stationary approach, i.e. neglecting the displacement d . That is a good assumption for accelerator magnets with typical frequencies of the order of some hertz. A further assumption is that we have no excess charge ρ .

Then the Maxwell equations look as follows:

Differential form:

$$\text{Ampere's law: } \nabla \times H = j \quad (1)$$

Integral form:

$$\oint H \cdot ds = \int j \cdot dA \quad (2)$$

$$\text{Faraday's law: } \nabla \times E = -\frac{\partial B}{\partial t} \quad (3)$$

$$\oint E \cdot ds = -\frac{\partial}{\partial t} \int B \cdot dA \quad (4)$$

$$\nabla \cdot E = 0 \quad (5)$$

$$\oint E \cdot dA = 0 \quad (6)$$

$$\nabla \cdot B = 0 \quad (7)$$

$$\oint B \cdot dA = 0 \quad (8)$$

$$\text{Material properties: } \underline{\underline{\mu}} = \mu_0 \mu_r H \quad (9) \quad j = \underline{\underline{\sigma}} E \quad (10)$$

The notation is as used in the literature (Knoepfel).

2.2 Diffusion approach

In this subsection the eddy current problem is handled as a diffusion problem. The physical background is the well-known Lenz's law: **'the electromagnetic field induced in an electric circuit always acts in such a direction that the current it drives around a closed circuit produces a magnetic field which opposes the change in magnetic flux'**.

The reason for this behaviour is the 'minus' sign in Faraday's law. It is demonstrated in Fig. 1, where a permanent magnet is moved in and out of a solenoidal coil. Please note that the flux created in the coil is opposed to the flux change induced by the movement of the permanent magnet. In other words: the eddy current tries to keep the flux status and consequently delays the change of the external field.

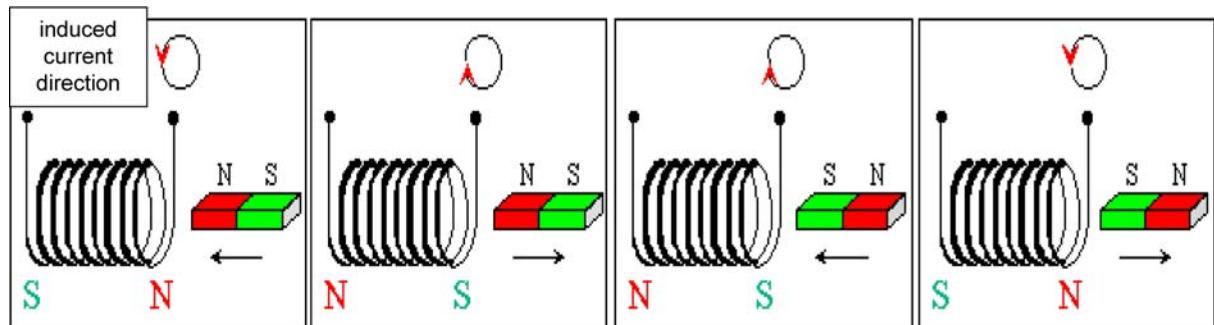


Fig. 1: Demonstration of Lenz's law

Therefore we can consider the process as a diffusion process of the magnetic field with a certain time constant, which depends on the electrical properties of the conductive circuit, and in the case of eddy currents on the electrical properties of the extended media. The advantage of the diffusion approach is (besides the clear physical meaning) that the appropriate methods of solution (also the numerical ones) which are well known from the field of thermal diffusion can be used.

2.2.1 Field diffusion equation (Knoepfel)

Taking the curl of Ampere's law (1) and introducing the material equation (10) we obtain under the assumption that the conductivity σ is uniform in space:

$$\nabla \times \nabla \times \vec{H} = \nabla \times \vec{j} = \sigma (\nabla \times \mathbf{E}) \cdot \quad (11)$$

Application of the well-known vector relation to the left-hand side and introducing Faraday's law (3) on the right-hand side leads to

$$\nabla^2 \vec{H} = \sigma \mu \frac{\partial \vec{H}}{\partial t} \quad (12) \quad \text{with} \quad \mu = \mu_0 \mu_r \cdot$$

This is the magnetic diffusion equation. The expression κ is called the ‘magnetic diffusivity’.

$$\kappa = \frac{1}{\sigma \cdot \mu} . \quad (13)$$

The interpretation is straightforward: the higher the conductivity (i.e., large eddy currents), the lower the diffusivity, i.e., the diffusion process is slow. Large μ means large stored energy, which takes a longer time to be stored.

Having solved the differential equation for the field H (under certain boundary conditions), the eddy current density can be calculated by Ampere’s law and consequently the power loss as well. Diffusion equations can also be derived in a similar way for the magnet vector potential A with the definition

$$B = \nabla \times A . \quad (14)$$

Having found the vector potential by solving the corresponding diffusion equation (under the chosen boundary conditions) one can calculate the current density j by the equation

$$\vec{j} = \sigma \vec{E} = -\sigma \frac{\partial \vec{A}}{\partial t} . \quad (15)$$

2.2.2 Analytical solutions of the diffusion equation (Examples)

Of course analytical solutions can only be given for simple applications, as far as the geometry of the components, the properties of the materials, and the variation in time are concerned. Nevertheless these solutions give an insight into the physics of the eddy current phenomena and give some simple formula for the estimation of eddy current effects in a magnet.

In the following sections we closely follow the treatment given by (Knoepfel).

2.2.2.1 Half-space conductor

The easiest start to understand the diffusion process and the role of the eddy currents in it, is to look at the ‘half-space conductor’ in one dimension only (Fig. 2).

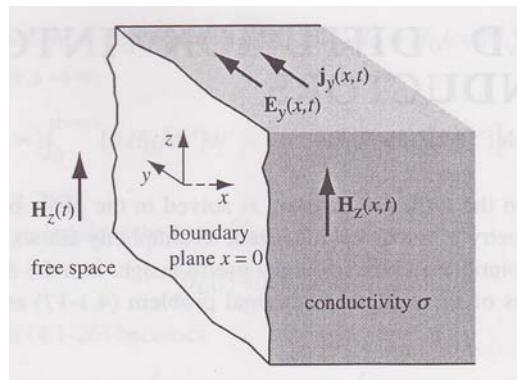


Fig. 2: Half-space conductor

In free space an external field $H_z(t)$ is applied. We seek the solution $H_z(x, t)$ of the differential equation

$$\frac{\partial^2 H_z}{\partial x^2} = \sigma \mu \frac{\partial H_z}{\partial t} \quad (16)$$

in the half-space for the specific boundary conditions

$$\begin{aligned} H_z(0, t) &= 0 & t < 0 \\ H_z(0, t) &= H_z(t) & t \geq 0 \\ H_z(x, 0) &= 0 & 0 < x < \infty \end{aligned}$$

We shall restrict ourselves to the three most common time dependencies: the step function field, the transient linear field, and the transient sinusoidal field.

2.2.2.1.1 Step function field $H_z(t) = H_0 = \text{constant}$

We define a ‘response function’ $S(x, t)$ (schematically shown in Fig. 3) by the following relation:

$$H_z(x, t) = H_0 * S(x, t)$$

with $S(x, 0) = 0$ and $S(x, t \rightarrow \infty) = 1$.

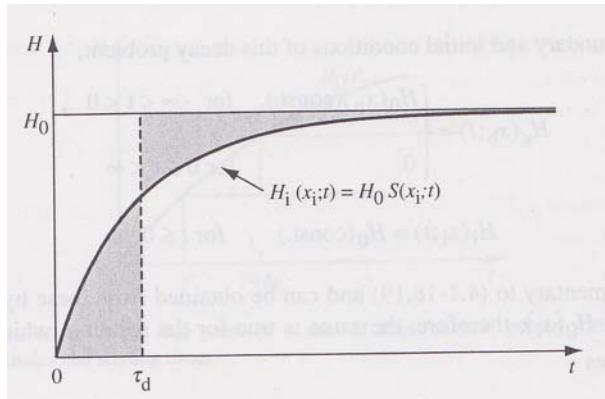


Fig. 3: Response function

An average diffusion time constant $\tau_d(x)$ can then be defined by

$$\tau_d = \int_0^\infty (1 - S(x, t)) dt . \quad (17)$$

It looks plausible to introduce a similarity variable ξ

$$\xi = \frac{x}{2\sqrt{\kappa t}} . \quad (18)$$

With this variable the magnetic diffusion differential equation (16) can be solved. One obtains the following solution

$$H_z(x, t) = H_0 \cdot (1 - \text{erf}(\xi)) . \quad (19)$$

with the special response function $S(x, t) = S(\xi) = 1 - \text{erf}(\xi) = \text{erfc}(\xi)$.

In Fig. 4 is plotted the error function and its complement erfc. Figure 5 shows the dependence of the response function on time, at different coordinates.

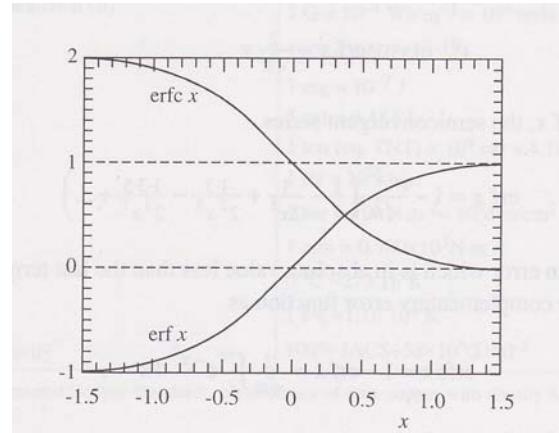


Fig. 4: Error function and its complement

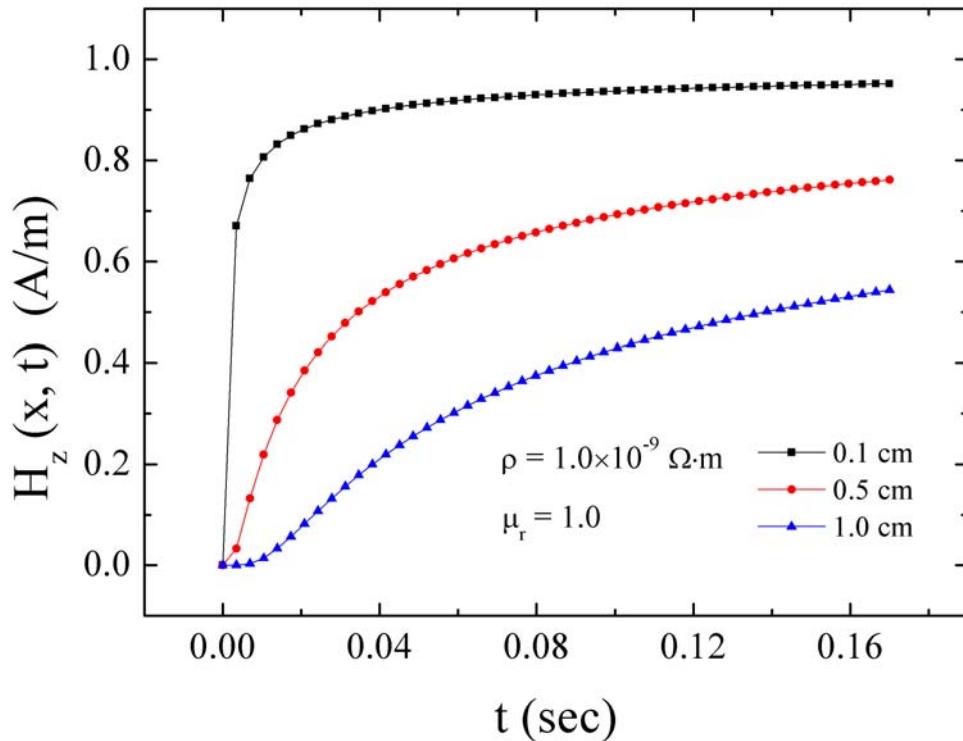


Fig. 5: The diffusion process at different positions x

2.2.2.1.2 Transient linear field $H_z(t) = (H_0/t_0)*t$

An external linear ramp is applied — as typically done in an accelerator: it starts from zero and reaches the maximum field H_0 after the time interval t_0 .

The solution is found as follows:

$$H_z(x,t) = \frac{H_0}{t_0} t \left[\left(1 + 2\xi^2 \right) \operatorname{erfc} \xi - \frac{2}{\sqrt{\pi}} \xi e^{-\xi^2} \right]. \quad (20)$$

It can be split into three parts, of which the first two form the stationary part:

$$H_z(x,t) = \frac{H_0}{t_0} \cdot t - \tau_d(x) \frac{H_0}{t_0} + H_z^t(x,t). \quad (21)$$

with

$$\tau_d(x) = \frac{x^2}{2\kappa}.$$

The first part is the applied external field. Owing to the eddy currents, the internal field lags behind during the ramp (second part) after the transient part (part 3) has vanished. The same is true at the end of the ramp. The situation is visualized in Fig. 6.

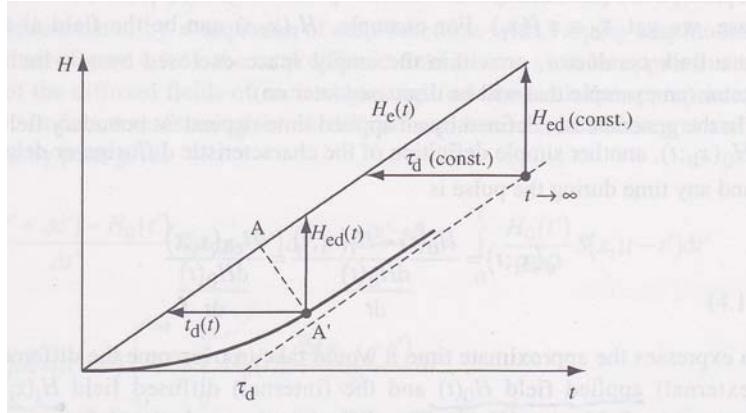


Fig. 6: Sketch of the time behaviour of the half-space field after the application of a linear ramp

Figure 7 visualizes $H_z(x, t)$ as a function of time for different positions x .

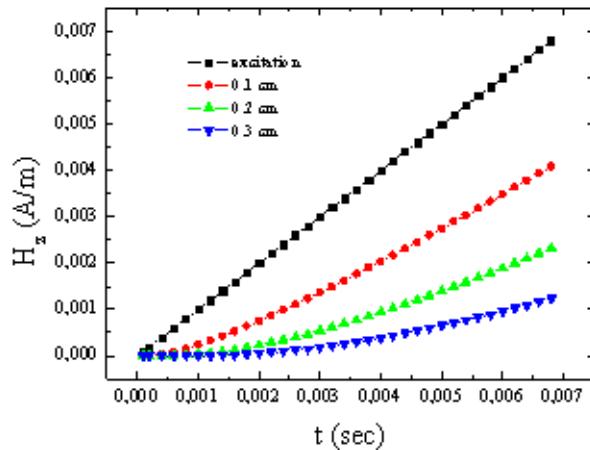


Fig. 7: Field diffusion as a function of time at different positions x

2.2.2.1.3 Transient sinusoidal field $H_z(t) = H_0 * \sin(\omega t)$

Solving the differential equations with this external field leads to the following stationary solution, after the transient part has disappeared:

$$H_z^S(x,t) = H_0 \cdot e^{-\frac{x}{\delta}} \cdot \sin(\omega t - \frac{x}{\delta}), \quad (22)$$

with

$$\delta = \sqrt{\frac{2}{\omega \mu \sigma}}, \quad (23)$$

δ is the well-known harmonic skin depth.

The field can penetrate the half-space only to the skin depth and has a phase shift relative to the externally applied field. Figure 8 shows the skin depth as a function of frequency and the product of conductivity and permeability (material property) [Knoepfel Fig. 4.2–5]. For example, 1 Hz-operation gives (for iron with $\mu_r = 100$) a skin depth of 20 mm (just connecting the corresponding points at the left and right scale). This diagram allows the magnet designer to estimate the lamination thickness of a laminated magnet, requiring that the skin depth δ be large compared to the lamination thickness. A lamination thickness of typically 0.5–1.0 mm seems to be adequate for the example mentioned.

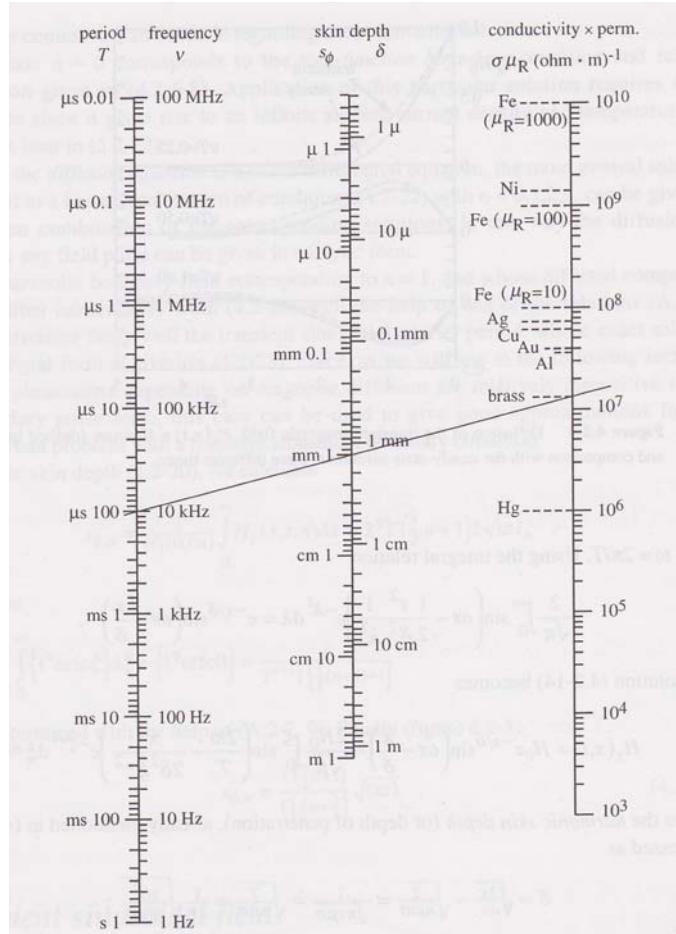


Fig. 8: Skin depth as function of frequency, conductivity and permeability [Knoepfel Fig. 4.2–5]

2.2.2.2 Slab conductor

The slab geometry is shown in Fig. 9. Only in the x direction does it have a limited dimension $2d$.

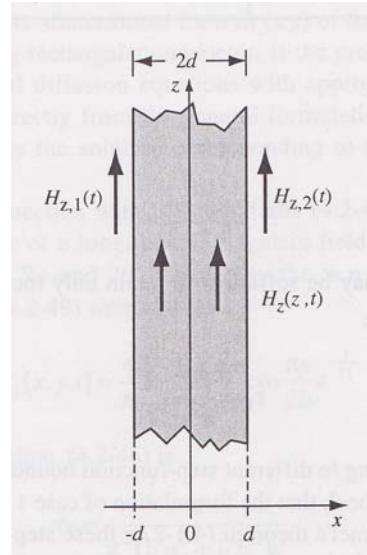


Fig. 9: Slab geometry (1-dimensional)

We consider step function excitation only, i.e., the following boundary conditions:

$$H_z(\pm d, t) = 0 \quad t < 0$$

$$H_z(\pm d, t) = H_0 \quad t \geq 0$$

$$H_z(x, 0) = 0 \quad -d < x < +d$$

The solution of the magnetic diffusion differential equation yields

$$H_z(x, t) = H_0 \left[1 - 4 \sum_{n=1}^{\infty} \frac{\cos \frac{n\pi x}{2d}}{n\pi(-1)^{\frac{n-1}{2}}} e^{-\frac{t}{\tau_n}} \right], \quad (24)$$

n odd, with

$$\tau_n = \frac{4}{n^2 \pi^2} \cdot \frac{d}{\kappa} \quad \text{and} \quad \kappa = \frac{1}{\sigma \cdot \mu}.$$

The second term in Eq. (24) is the transient term. It is the sum over different n including a whole spectrum of time constants τ_n . Please note that the slab dimension appears as the square in the equation of the time constant. Doubling the lamination thickness leads to a factor of 4 in the time constant and, as we shall see later, also in the ohmic losses. One may call it the ‘relevant distance’ of the eddy currents. It is obvious that the time constants are inversely proportional to the diffusivity κ : the smaller the diffusivity (high conductivity, high permeability), the longer the time constants.

The longest possible time constant [for example, for a 1 mm thick low-carbon iron ($\mu_r = 1000$, $\sigma = 10^7$ (Ohm*m) $^{-1}$) lamination] is less than 1 millisecond.

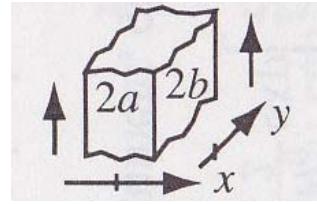


Fig. 10: Slab geometry (2-dimensional)

In a similar calculation the solution for a 2D slab (Fig. 10) can be found [Knoepfel]:

$$H_z(x, y, t) = H_0 \left[1 - 4 \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{\cos \frac{n\pi x}{2a} \cos \frac{m\pi y}{2b}}{\pi^2 \cdot f(n, m)} e^{-\frac{t}{\tau_{n,m}}} \right] \quad (25)$$

n, m odd

$$\tau(n, m) = 4 / \left[\pi^2 \kappa \left(\frac{n^2}{a^2} + \frac{m^2}{b^2} \right) \right]$$

$$f(n, m) = \frac{n * m * (-1)^{\frac{n+m}{2}}}{4}$$

2.2.2.3 Iron-dominated C-dipole

Of course a magnet builder is interested in how a real magnet behaves.

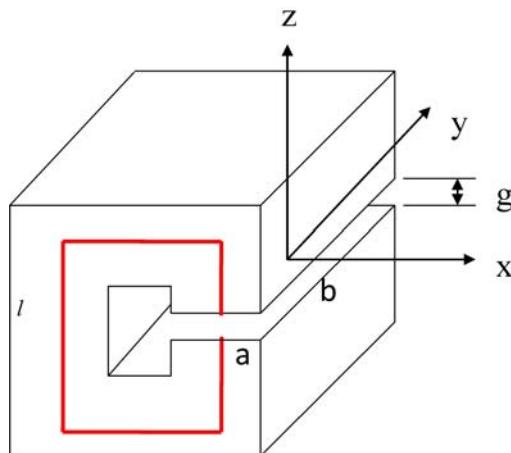


Fig. 11: C-shaped dipole

In a CERN internal note [3] the magnetic diffusion problem of a C-type dipole (Fig. 11) was handled. Here l is the average iron path length, g the gap height. Typically the ratio l/g is about 10 ($l \gg g$) and a and b correspond to the dimensions of the 2D slab in the previous subsection.

Starting from the magnetic diffusion equation, applied to a slab with a gap, they found a special differential equation for the field in the gap of the C-dipole:

$$\frac{\partial^2 H}{\partial x^2} + \frac{\partial^2 H}{\partial y^2} = \sigma \mu_0 \frac{l}{g + \left(\frac{l}{\mu_r}\right)} \frac{\partial H}{\partial t} . \quad (26)$$

It is easy to verify that for $g = 0$ we get the standard diffusion equation. For field levels away from saturation ($l/g \ll \mu_r$) we obtain

$$\frac{\partial^2 H}{\partial x^2} + \frac{\partial^2 H}{\partial y^2} = \sigma \mu_0 \frac{l}{g} \frac{\partial H}{\partial t} = \frac{1}{\kappa_1} \frac{\partial H}{\partial t} \quad \kappa_1 = \frac{1}{\mu_0 \sigma * \frac{l}{g}} . \quad (27)$$

Here we defined a new diffusivity κ_1 that is not determined by the relative permeability, but by the dimensions l and g of the magnet. The magnetic energy is mainly stored in the gap of the magnet!

With this new diffusion constant all the previously mentioned solutions can be used. For the mentioned typical value of $l/g = 10$ we get a smaller value compared to the slab with μ_r between 100 and 1000 and consequently shorter time constants.

2.2.3 Numerical solutions (introduction of numerical codes)

Table 1 compares analytical with numerical methods.

Table 1: Comparison of analytical and numerical methods

	Advantages	Disadvantages
Analytical methods	<ul style="list-style-type: none"> • physical understanding 	<ul style="list-style-type: none"> • simple geometry (mainly 1D/ 2D) • homogeneous, isotropic and linear materials • simple excitation
Numerical methods	<ul style="list-style-type: none"> • complex geometry (3D) • inhomogeneous, anisotropic and nonlinear materials • complex excitation 	<ul style="list-style-type: none"> • long computing times

Analytical methods deliver a good physical understanding and the formulas allow a good estimate of the effects under simple assumptions, but for better results in complex situations one needs numerical calculations.

The most common way to define eddy currents in conducting media is based on solution equations for vector potential A and scalar electric potential V [4]:

$$\nabla \times \frac{1}{\mu} \nabla \times \vec{A} = -\sigma \frac{\partial \vec{A}}{\partial t} - \sigma V , \quad (28)$$

$$\nabla \cdot \sigma \nabla V + \nabla \cdot \sigma \frac{\partial \vec{A}}{\partial t} = 0 , \quad (29)$$

with appropriate boundary and conditions. The current density vector may be found as:

$$\vec{J} = -\sigma \frac{\partial \vec{A}}{\partial t} - \sigma \nabla V . \quad (30)$$

Some codes use the current vector potential T , defined by

$$j = \sigma \cdot \nabla \times T . \quad (31)$$

The following listing gives an overview of widely used commercially available numerical codes for the calculation of eddy currents in magnets.

- Opera[®] (Vector Fields Software, Cobham Techn. Services, Oxford) www.vectorfields.com
 - o Finite Element Method (FEM)
 - o Opera 2d,AC and TR, Opera 3d, ELEKTRA[®] , (TEMPO-thermal and stress-analysis)
- ROXIE[®] (Routine for the Optimization of Magnet X-Sections, Inverse Field Calculation and Coil End Design) (S. Russenschuck, CERN) <https://espace.cern.ch/roxie/default.aspx>,
 - o BEM/FEM
 - o Optimization of $\cos n\theta$ -magnets, coil coupling currents only
- ANSYS[®] (ANSYS Inc.) www.ansys.com
 - o Finite Element Method (FEM)
 - o Direct and indirect coupled analysis (Multiphysics)
 - eddy current → heat → rising temperature → change resistivity → change eddy current (very important, especially for sc magnets)

2.3 Direct solutions of Maxwell equations (small perturbation)

In Section 2.2 we looked at the eddy current effects from the perspective of field diffusion. Owing to eddy currents the magnetic field lags behind (Lenz's rule) and slowly diffuses into the conducting material. The magnetic diffusivity was defined. The field was calculated analytically by solving the diffusion equation. Formulas for the penetration depth and the time constants were given for some simple cases. Numerical codes were introduced which calculate field, current density, and losses in more complex cases. Basically, no special approximations are necessary.

In this section we want to use a second approach by solving the Maxwell equations directly. The eddy currents will be considered as a ‘small perturbation’. Some people call the method ‘quick and dirty’, because it calculates the eddy currents from the Maxwell equations, not taking into account

that these eddy currents also create a magnetic field which affects the current distribution as well. But under the assumption of a ‘small perturbation’, these are second-order effects.

Typically we consider a case as ‘small perturbation’, if the conducting material is ‘magnetically thin’, i.e., the relevant dimension is small compared to the penetration depth. Taking into account the formula for the skin depth

$$\delta = \sqrt{\frac{2}{\omega\mu\sigma}} , \quad (23)$$

we see that this is true for low frequencies (small ramp rates), low conductivity and low permeability.

This approach will be demonstrated with some examples.

2.3.1 Eddy currents in a rectangular, long, thin plate

Figure 12 shows the geometry.

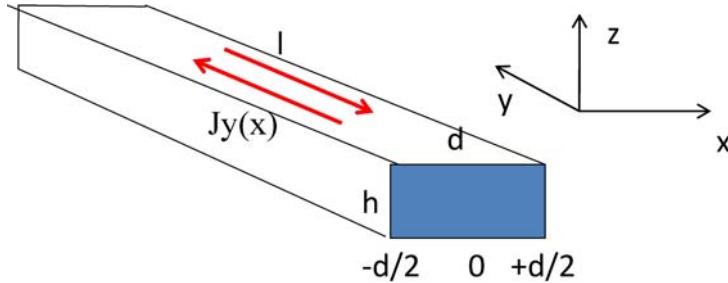


Fig. 12: Thin, long plate

The following conditions must be fulfilled:

- Field vertical to the plate, constant ramp rate \dot{B}_z
- Plate magnetically thin: $d \ll$ penetration depth δ
- Plate geometrically thin: $h \ll d$
- Plate long enough to neglect the ends: $d, h \ll 1$
- Steady-state conditions: $t \gg \tau_d$

From the integral form of Faraday’s law one can calculate the eddy current distribution, neglecting the ends:

$$j_y(x) = -\frac{\dot{B}_z}{\rho} \cdot x . \quad (32)$$

From Ampere’s law (1), one then obtains the field distribution produced by the eddy currents

$$H_z^{eddy}(x) = \frac{\dot{B}_z}{2\rho} \left(x^2 - \frac{d^2}{4} \right) , \quad (33)$$

and finally, the loss by integration:

$$dP = \rho \frac{l}{A} \cdot (j_y(x) \cdot A)^2 = \rho \cdot l \cdot j_y^2(x) \cdot h \cdot dx \quad (34)$$

$$P = 2 \int_0^{d/2} dP = \frac{lh}{12} \frac{d^3}{\rho} B_z^2 \quad \text{or} \quad P / \text{volume} = \frac{1}{12} \frac{d^2}{\rho} B_z^2 \quad (35)$$

Please note:

- The relevant width d enters the formula as squared. The power loss per volume depends only on the square of the relevant dimension d , the resistivity ρ , and of course on the square of the ramp rate.
- The eddy currents produce a negative dipole and a sextupole (for example on the top and bottom of a rectangular beam pipe).
- Because of the geometrical restriction, the loss formula is the same as for a thin slab (thin lamination).
- The resistivity ρ is a material property (Fig. 13). For copper it is more than two orders of magnitude lower than for iron or stainless steel. Therefore beware of eddy currents in copper! The resistivity depends on temperature (Fig. 14). At 4 K it is up to one order of magnitude lower than at room temperature. Therefore, the eddy current losses in a superconducting magnet are much higher than in a room-temperature magnet (besides the fact that they occur at 4 K, where the Carnot efficiency is very low).

	300K	4K
Resistivity ρ (Ohm*m)		
LC steel (3% Silicon)	590×10^{-9}	440×10^{-9}
Stainless steel	720×10^{-9}	490×10^{-9}
Copper	17.4×10^{-9}	0.156×10^{-9}

Fig. 13: Typical resistivity for different materials

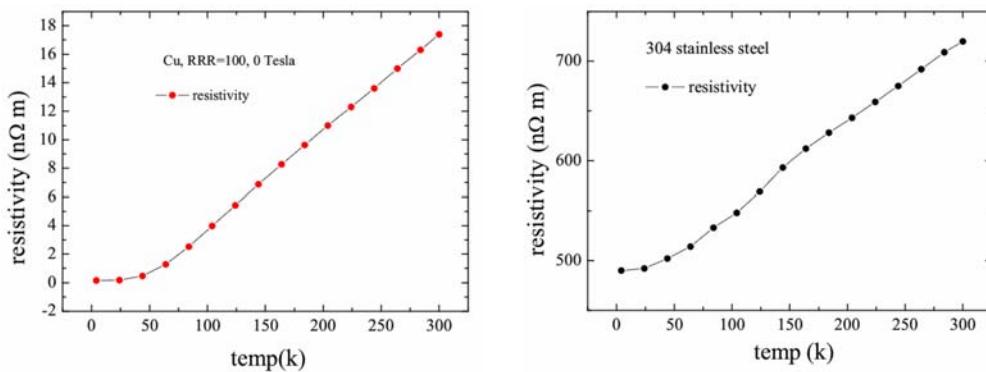


Fig. 14: Resistivity as a function of temperature for copper and stainless steel

Figures 15 and 16 compare a ‘small perturbation’ calculation with a numerical FEM solution. One can clearly see that for the stainless-steel case the results agree, but for copper (with its very high electrical conductivity) the condition ‘magnetically thin’ is no longer fulfilled and the results are completely different.

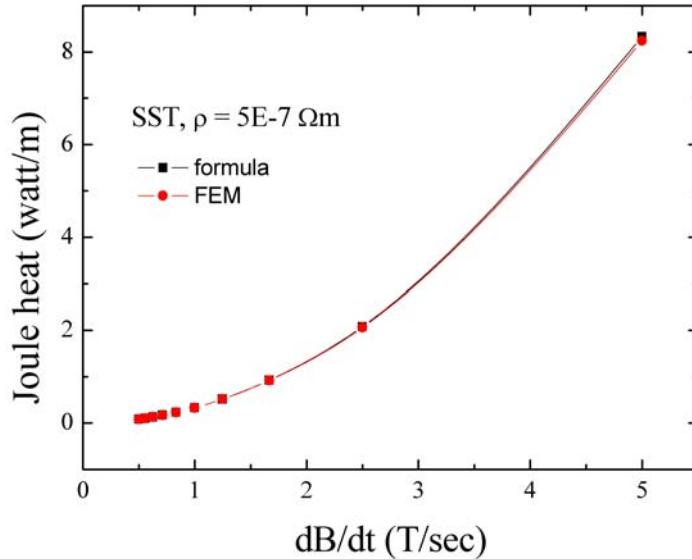


Fig. 15: Joule heating in thin stainless-steel plate, the numerical result is well reproduced by the approximation

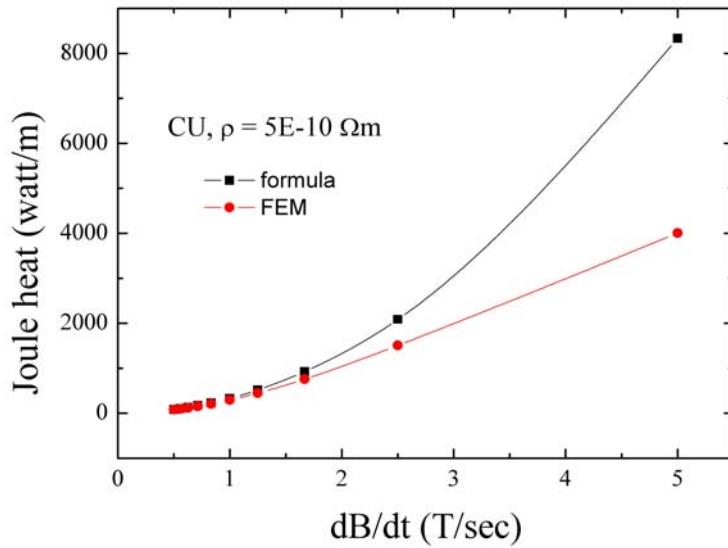


Fig. 16: Joule heating in a copper plate, the ‘small perturbation’ condition is no longer fulfilled due to the high conductivity of the copper

2.3.2 Eddy currents in a long, thin cylinder (beam pipe)

This example handles a thin cylinder as used in many accelerator magnets. The cylinder with radius r , wall thickness d and length l must be magnetically thin (i.e., $r \ll$ skin depth), geometrically thin $d \ll r$ and long ($d, r \ll l$), to ignore the ends and therefore be handled as a 2D problem.

Then, we obtain (in the same way as in 2.3.1) the current density and the power loss

$$j = \frac{r \cos \theta}{\rho} \dot{B} . \quad (36)$$

$$P = \frac{r^3}{\rho} \dot{B}^2 \pi d l \quad \text{or} \quad \frac{P}{V} = \frac{r^2}{2\rho} \dot{B}^2 . \quad (37)$$

Please note:

- The eddy current distribution is proportional to $\cos \theta$, which leads to a pure dipole.
- Again, the loss per volume shows the same dependence as for the thin plate.

2.3.3 Eddy currents in a round thin plate or disk

In a thin plate or disk with radius r and thickness d (with $r \gg d$) we obtain, with the same computation steps

$$\frac{P}{V} = \frac{r^2}{8\rho} \dot{B}^2 . \quad (38)$$

3 Eddy currents in accelerator magnets

In this section we consider real accelerator magnets and investigate the role of eddy currents in these magnets [5–7].

Eddy currents can occur in all conductive elements which see a change in magnetic flux, especially in:

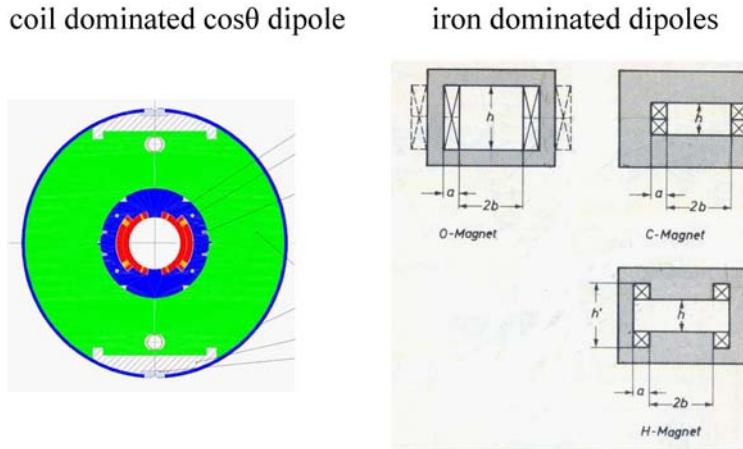
- iron yokes (low carbon iron)
- mechanical structures (mostly low carbon iron or stainless steel)
- coils (copper, superconductor)
- beam pipes (stainless steel)

The main classification of these magnets is:

- coil-dominated magnets ($\cos n\theta$ current distribution or intersecting ellipses)
- iron-dominated magnets

Figure 17 shows some typical examples.

Both types of magnet can be equipped with resistive or superconducting coils. Nevertheless iron-dominated magnets are generally used for fields below 2 T due to iron saturation and are therefore mostly resistive, while coil-dominated magnets use superconducting coils for higher field strength.

**Fig. 17:** Some magnet types

3.1 Eddy currents in the yoke

Accelerator magnets can have either solid or laminated yokes. Since most magnets which are subjected to a time varying field have laminated yokes we shall treat only those.

3.1.1 2D effects within a long magnet

The first questions to answer are the choice of the lamination thickness and of the iron properties.

3.1.1.1 Lamination thickness d

In order to keep the influence of eddy currents small we need to keep

- the relevant geometrical dimensions (here the lamination thickness) small compared to the penetration depth (refer to 2.2.2.1.3). Using Eq. (23) or Fig. 8 one can find the penetration depth δ :

$$\delta = \sqrt{\frac{2}{\omega \mu \sigma}} ; \quad (23)$$

- the longest time constant (refer to Eq. (24) for a one-dimensional slab or Eq. (27) for a C-shaped dipole) small compared to the ramping time or cycle time of the application:

$$\tau_n = \frac{1}{n^2 \pi^2} \cdot \frac{d^2}{\kappa} ;$$

please note that in this formula the thickness of the lamination is d , not $2d$!

with

$$\kappa = \frac{1}{\sigma \cdot \mu} \text{ for a slab ,}$$

$$\kappa = \frac{1}{\sigma \cdot \mu_0 * \sqrt{\frac{l}{g}}} \text{ for a C-shaped dipole .}$$

In addition the power loss/volume in the laminations due to the eddy currents should be calculated [Eq. (35)]:

$$P / \text{volume} = \frac{1}{12} \frac{d^2}{\rho} \dot{B}_z^2 .$$

The losses in the laminations should be small or at least comparable to other losses in the magnet.

Please note that in this formula also the lamination thickness is d , not $2d$!

Conclusion: In order to limit the influence of the eddy currents, we build magnets with carefully insulated laminations of high resistivity iron. Obviously, one has two parameters to play with: the lamination thickness and the resistivity.

3.1.1.2 Choice of the iron

The following three pictures were taken from Ref. [8].

Silicon increases the resistivity of low carbon iron (Fig. 18).

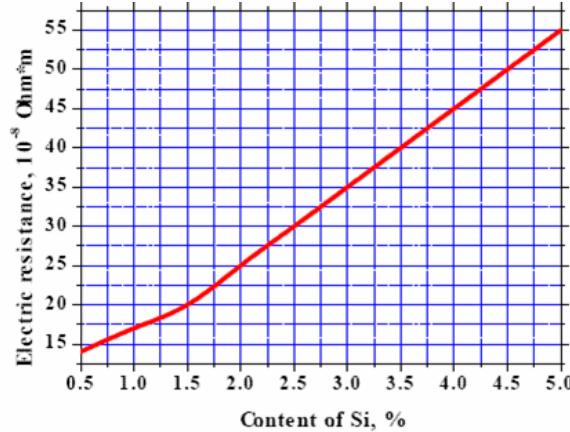


Fig. 18: Resistivity of iron as a function of the content of silicon

Fortunately, silicon also reduces the coercivity, therefore reducing the hysteresis losses in the laminated yoke (Fig. 19).

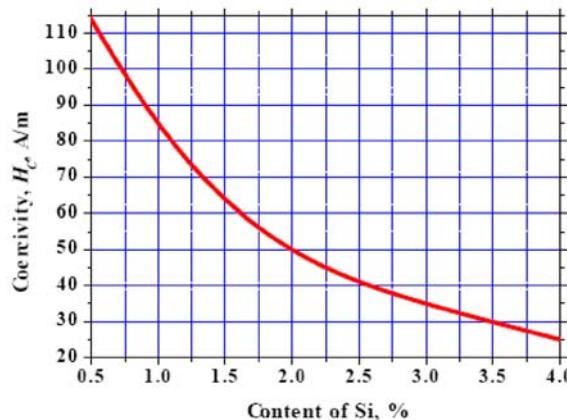


Fig. 19: Coercivity of iron as a function of the content of silicon

Unfortunately, the saturation magnetization decreases with the silicon content, defining a limit of 3.5% silicon content (

Fig. 20).

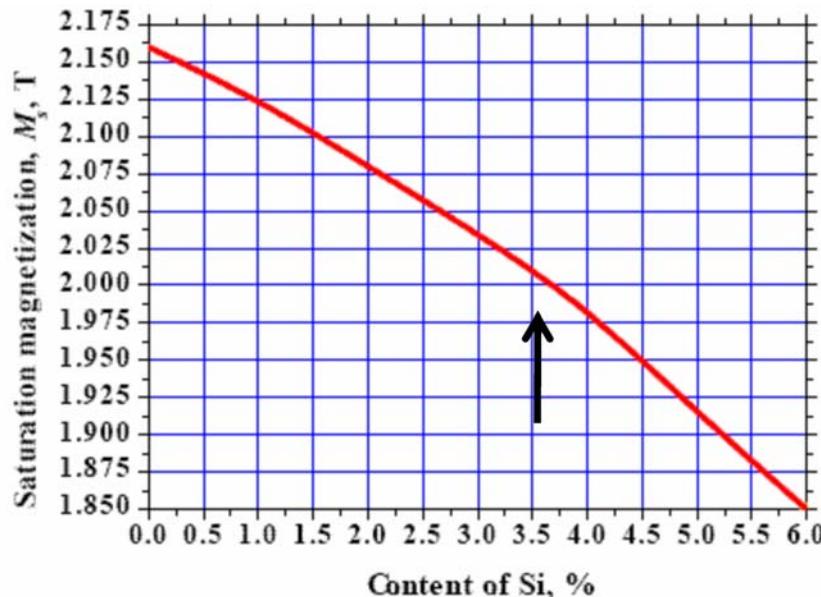


Fig. 20: Saturation magnetization as a function of the content of silicon

There is a practical limit on lamination thickness too: steel sheets are available on the market down to 0.3 mm (transformers) thickness. In addition, thinner laminations reduce the packing factor and increase the labour costs.

Please note: steel suppliers typically give total iron losses at 50 Hz. They have three parts:

- eddy losses, scaling with the square of the frequency;
- hysteresis losses, scaling linearly with the frequency; and the so-called
- anomalous losses, scaling with the power of 1.5 of the frequency.

Eddy current losses are given by

$$P \left[\frac{W}{m^3} \right] = \frac{\pi^2 v^2}{6\rho} B_p^2 d^2 \quad (39)$$

v - frequency (Hz)

d - lamination thickness (m)

ρ - resistivity (Ohm*m)

B_p - induction amplitude (T)

Hysteresis losses can be measured with a permeameter. The rest are anomalous losses. By appropriate scaling with the frequency, you may find the losses for your application [9].

3.1.2 3D effects

3.1.2.1 Anisotropy of a laminated yoke

3.1.2.1.1 Packing factor

We define a packing factor f_p

$$f_p = W_i / (W_i + W_a) \quad \text{with}$$

- W_i thickness of a single lamination (without insulation),
- W_a thickness of insulation.

In reality, the packing factor is slightly lower for mechanical reasons. Typical packing factors are between 0.95 and 0.98.

3.1.2.1.2 Conductivity

Since the laminations are insulated, the conductivity is highly anisotropic:

- perpendicular to the lamination: $\sigma_z = 0$
- within the lamination: $\sigma_{xy} \neq 0$.

3.1.2.1.3 Relative iron permeability μ_r

Generally, the relative permeability is a tensor. Each element $\mu_{ij}(x, y, z; \vec{H})$ depends on position and field strength. We call the permeability homogeneous if it does not depend on position, and linear if it does not depend on \vec{H} . It is isotropic if all diagonal elements are equal to μ_r and the non-diagonal elements are zero.

In the case of a laminated magnet the permeability is basically homogeneous, but it is strongly non-linear, especially at low field, due to coercivity, and at high field, due to saturation.

It is highly anisotropic, since the diagonal elements differ from each other. The permeabilities μ_{xx} or μ_{yy} within the lamination are different from the permeability μ_{zz} perpendicular to the lamination. The permeabilities can be calculated using the continuity equations:

For magnetic flux tangential to the laminations [10] :

$$\mu_{xx} = \mu_{yy} = f_p * (\mu_r(\vec{H}) - 1) + 1 \approx f_p * \mu_r(\vec{H}) . \quad (40)$$

The flux is reduced by the packing factor.

For magnetic flux normal to the laminations [10]:

$$\mu_{zz} = \mu_r(\vec{H}) / (\mu_r(\vec{H}) - f_p * (\mu_r(\vec{H}) - 1)) , \quad (41)$$

and for $\mu_r \gg 1$

$$\mu_{zz} = 1 / (1 - f_p) . \quad (42)$$

In this case, the permeability is drastically reduced to typical values between 20 and 50 at low fields and will further decrease for saturation.

Figures 21 and 22 show the permeabilities as a function of field [11].

ANSYS

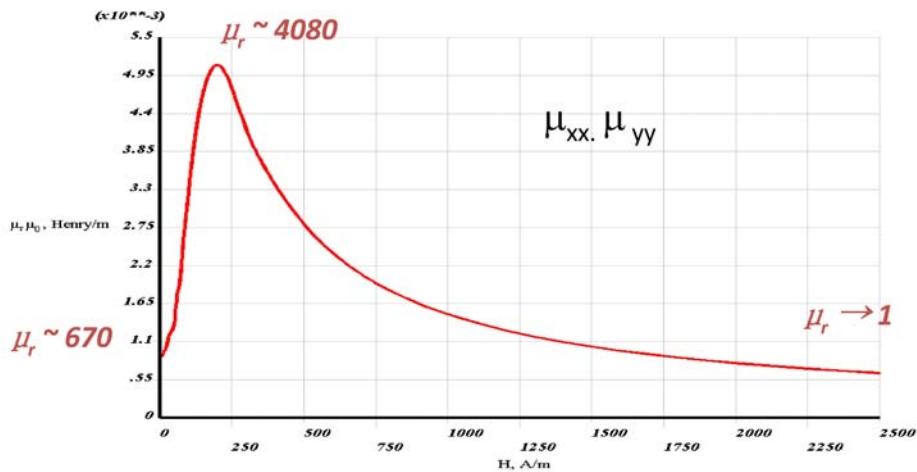


Fig. 21: Permeability μ_{xx} and μ_{yy} (in the plane of the lamination)

ANSYS

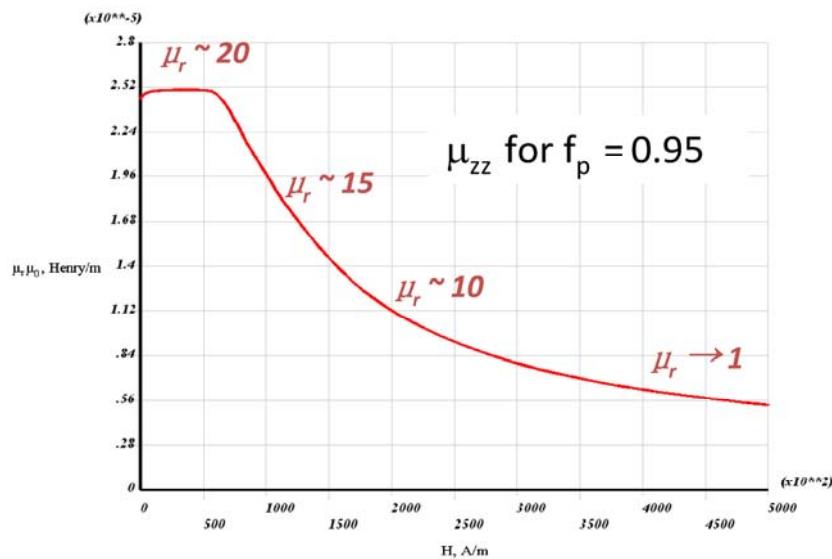


Fig. 22: Longitudinal permeability μ_{zz}

One sees clearly that for low fields the permeability within the lamination is much higher than the longitudinal permeability because the magnetic reluctance is much higher in this direction. However, at high fields due to iron saturation the permeabilities become comparable. As a consequence the distribution of magnetic flux changes along a magnetic ramp from low to high field — just a static phenomenon! Near the magnet ends the longitudinal flux penetrates the iron at high field more than at low field. Since this flux change is responsible for the eddy current loss in the ends, the losses will increase at high fields.

3.1.2.2 Eddy currents due to field components B_z perpendicular to the laminations

An ideal (infinitely long in the z -direction) magnet can be considered as a 2D structure. It has no z -component of the field and its properties do not vary with the longitudinal coordinate z . As was shown in the last paragraph we optimize the magnet as far as eddy currents are concerned by choosing

- the appropriate lamination thickness d (practical limit 0.3 mm),
- low steel conductivity,
- low coercivity (in order to reduce the hysteresis losses).

The dimension relevant for eddy current in this case is the lamination thickness d (Fig. 23).

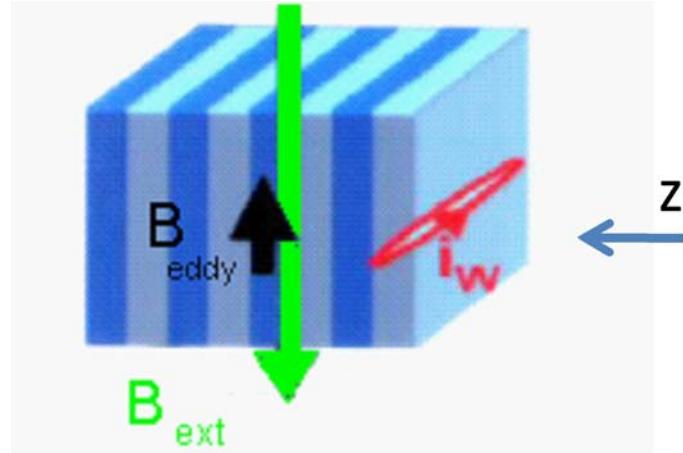


Fig. 23: Schematic view of a 2D-magnet (no B_z component)

A real magnet has of course a limited length and its properties — such as the packing factor — do vary with z . Therefore longitudinal field components B_z exist in

- yoke end regions,
- areas with varying packing factor etc. (Fig. 24).

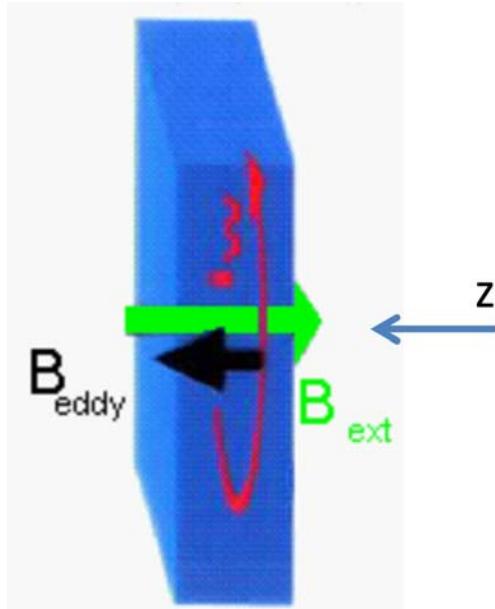


Fig. 24: Schematic view of a 3D-magnet (B_z components exist)

In this case the relevant dimension of the eddy current pattern is not the lamination thickness anymore, but the x - y surface itself (or part of it). Consequently, time constants are larger and losses are not negligible.

3.1.2.2.1 Field variation due to eddy current effects in the end part of magnets

In the following paragraphs time delay and losses caused by eddy currents in the end part of real magnets will be demonstrated. The eddy currents are induced by a linear ramp of the excitation current or a triangular cycle. The corresponding field lag and time constant is a function of z and can be measured.

Figure 25 schematically shows the situation at the end of a dipole.

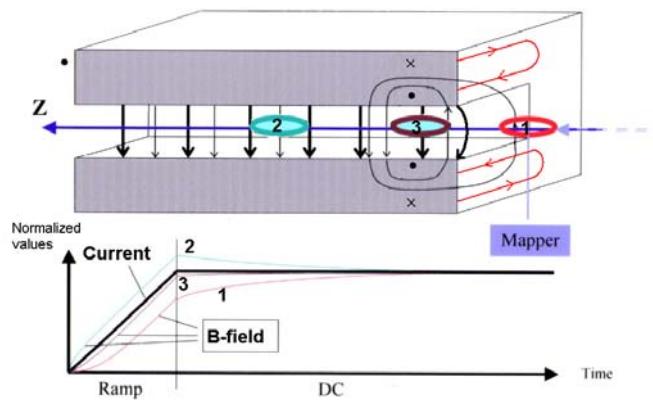


Fig. 25: Schematic view of the eddy current induced vertical field component (courtesy of F. Klos)

The z component of the main field induces eddy currents in the pole sheets, which produce their own field, superimposing the main field. In region 1 the vertical field component B_y decreases (field lag), while in region 2 the vertical field component increases. In the intermediate region 3 we have a transition. Of course the effect becomes smaller towards the longitudinal magnet centre.

3.1.2.2.1.1 SIS 18 dipole

This phenomenon was tested experimentally on the 2.6 m long curved dipole of the synchrotron SIS 18 at GSI (Fig. 26).



Fig. 26: The synchrotron dipole of SIS18 at GSI

The magnet was ramped up to 1.8 T in 1.1 seconds. The vertical field B_y was measured with a Hall probe along the trajectory z . In Fig. 27 is shown the measured field lag/enhancement as a function of time ($t = 0$ corresponds to the end of the current ramp) and longitudinal position z ($z = 0$ corresponds to the longitudinal magnet centre).

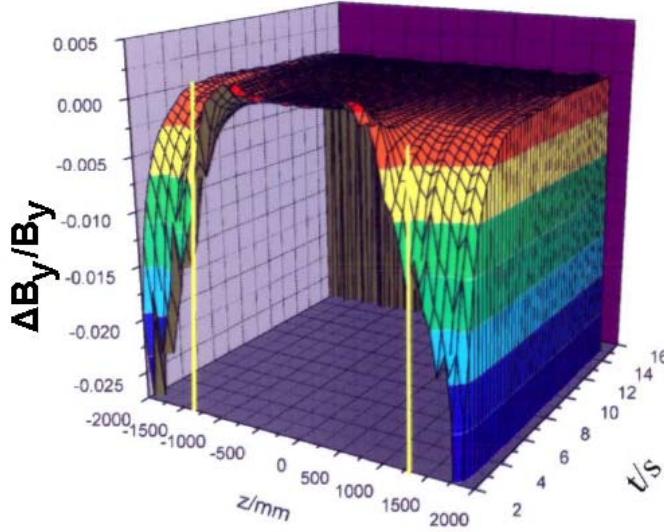


Fig. 27: Field lag/enhancement of the vertical field component due to eddy currents ($t = 0$ corresponds to the end of the ramp, longitudinal magnet centre at $z = 0$, yoke end at $z = \pm 1300$ mm)

One clearly recognizes the field lag outside, but also inside of the magnet. Since the magnet is ramped up to saturation the B_z components exist far inside of the magnet (refer to 3.1.2.1.3).

The much smaller effect of field enhancement is observed only near the center of the magnet. An integral measurement with a search coil shows only an integral field lag (Fig. 30).

Calculations based on current vector potential verified these experimental results [12]. The current density of the pole surface as a function of z at different times is shown in Fig. 28. One realizes that the eddy currents (induced by the B_z components) penetrate into the magnet at higher excitation. That happens because the B_z flux penetrates into the iron due to the anisotropy of the permeability, mentioned earlier.

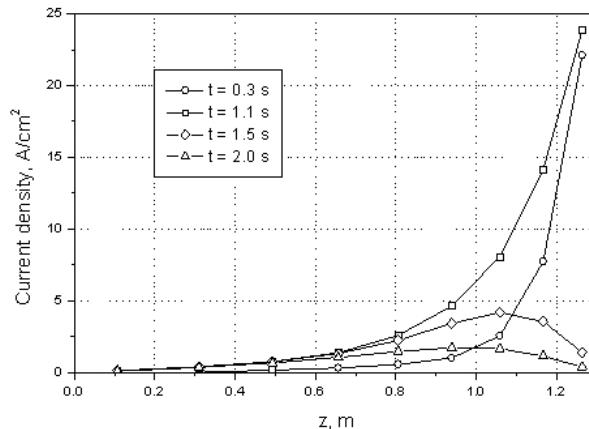


Fig. 28: Current density of the pole surface as a function of z (ramp starts at $t = 0$ and ends after 1.1 seconds).

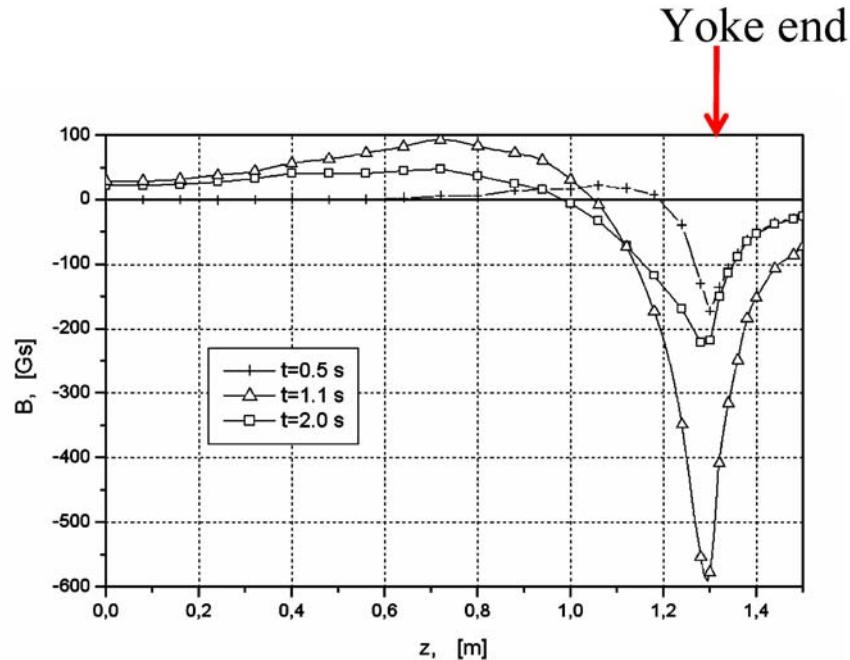


Fig. 29: B_y field component, induced by the eddy currents

Figure 29 confirms the field lag/enhancement of B_y and Fig. 30 compares the measured and calculated B_y (induced by the eddy currents) integrated over the whole magnet length.

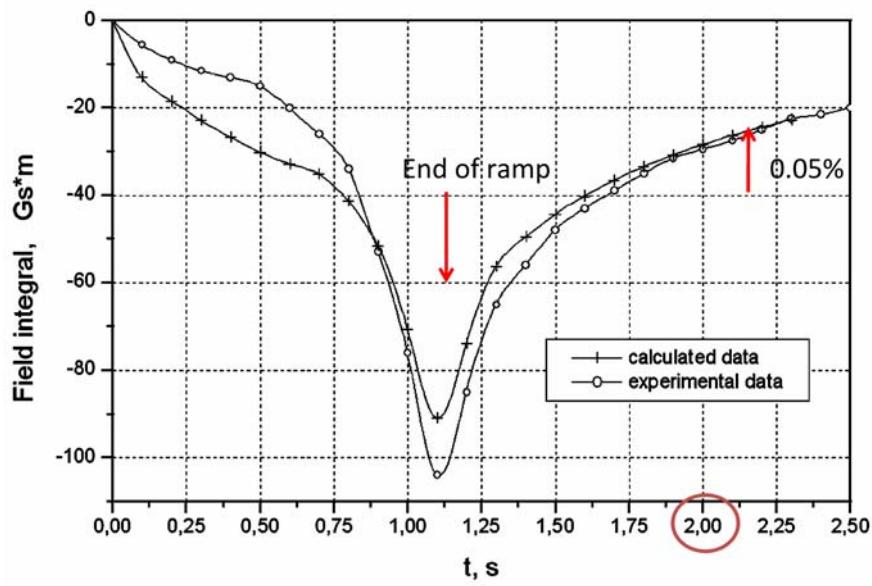


Fig. 30: Integral field induced by eddy currents (ramp 0–1.8T in 1.1 seconds)

The effect is huge: only 2 seconds after the end of the ramp the gap field reaches the required field quality of 5×10^{-4} (total field integral 48000 Gs·m).

3.1.2.2.1.2 CNAO dipole

Field delay was also measured on a 0.44 m long scanner dipole of CNAO (0.3 mm lamination thickness). It was ramped with 500 T/s up to 0.3 T. The gap field B_y was measured with a Hall probe at several longitudinal positions z . The field lag was fitted to an exponential function after the ramp stopped. The results are shown in Fig. 31 [13].

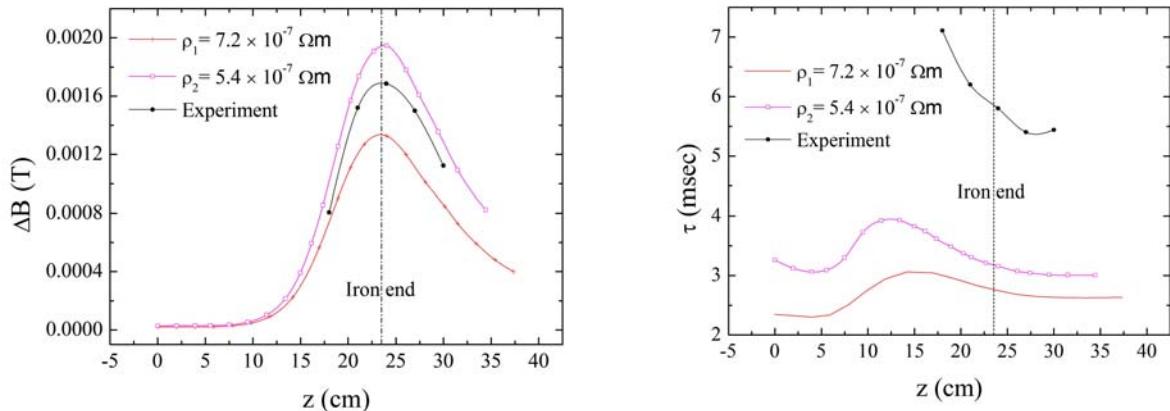


Fig. 31: Field lag and diffusion time constant as a function of the longitudinal coordinate z . ($z = 0$ corresponds to the magnet centre).

One can observe that

- A significant field lag occurs only near the magnet end; the eddy currents do not penetrate much, since the magnet is operated at low field (max. 0.3 T).
- The diffusion time constant near the ends is of the order of some milliseconds, much larger than the value calculated with Eqs. (24) and (27) for the centre of the magnet.
- The field lag is up to 0.7% of the nominal field (the value integrated over the whole magnet is of course smaller).
- The numerical values, calculated with ANSYS, agree reasonably well with the measured data.

3.1.2.2.1.3 SIS 100 dipole model

This type of dipole (2.72 m long, superferric, window-frame type) will be used as synchrotron dipole of the SIS 100 of the FAIR facility. It has a superconducting coil, forming the cold mass together with cold iron. Since it is a synchrotron magnet, correct tracking requires the time lag of the integral field to be known.

The field lag was calculated by ANSYS at four different longitudinal positions [14] as indicated in Fig. 32.

Outside of the magnet, we observe a large field lag and inside at positions 0.689 and 1.01 m a small field enhancement. At the position 1.17 m the initial field enhancement transforms into a field lag, since the eddy current centre moves inside the magnet. All time constants are of the order of 100 ms, smaller than the ramp time of 475 ms.

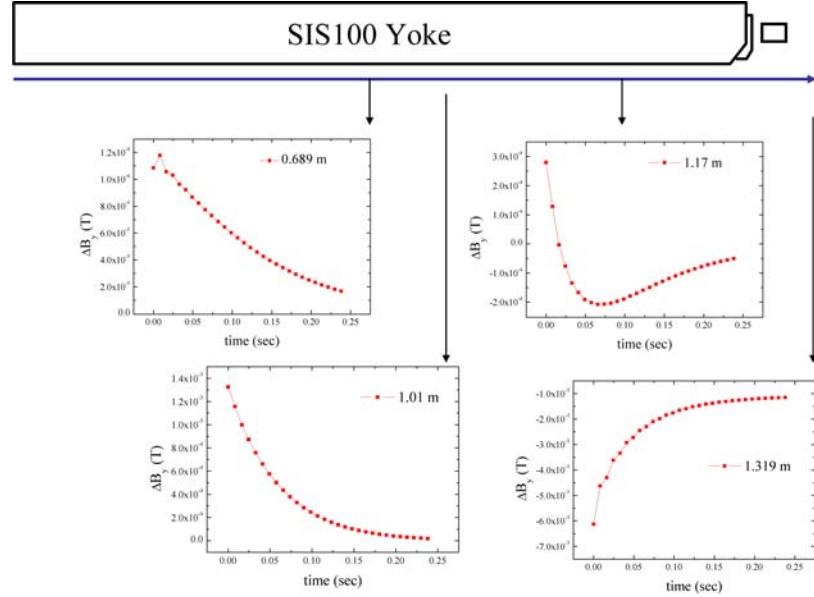


Fig. 32: Field lag/enhancement at several longitudinal positions ($t = 0$ corresponds to the end of a linear ramp up to 1.9 T, ramp rate 4 T/s, $x = 0$ corresponds to the centre, $x = 1.36$ m to the yoke end plate)

3.1.2.2.2 Resistive losses due to eddy current effects in the end parts of magnets

As mentioned before, the eddy current resistive losses are of special interest for superconducting cold iron magnets. Besides the hysteresis losses they form the biggest contribution to the cryogenic load.

3.1.2.2.2.1 SIS 100 dipole

Loss calculation results can be found in several publications [15], [16].

Figure 33 shows the calculated end-part losses during a triangular cycle with a period of 1 s [17], [18]. The current ramp rate is kept constant. Owing to the nonlinearity and anisotropy of the permeability (refer to 3.1.2.1.3) the losses are not constant during the ramp. They depend strongly on the packing factor: a large packing factor leads to a large longitudinal permeability, which reduces the magnetic reluctance and allows the flux to penetrate. Therefore the eddy current density penetrates further into the iron increasing the losses. A well stacked magnet may produce large end-part losses. The diffusion time constants and losses are investigated as a function of operating frequency and stacking factor in Ref. [19].

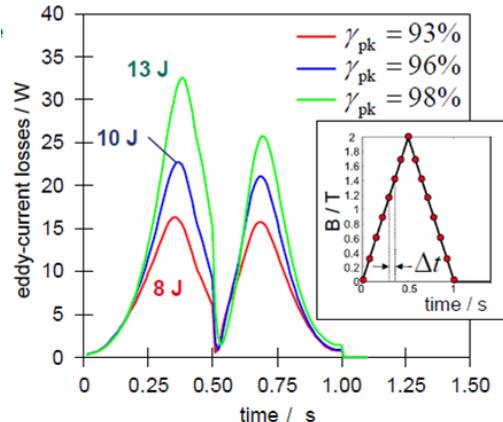


Fig. 33: End-part losses during a triangular cycle (parameter is the packing factor)

Figure 34 makes that even clearer: it shows the loss distribution as a function of the longitudinal coordinate. The variable parameter is the time during the cycle [11], [20]. The higher the field, the more the eddy current density penetrates into the iron.

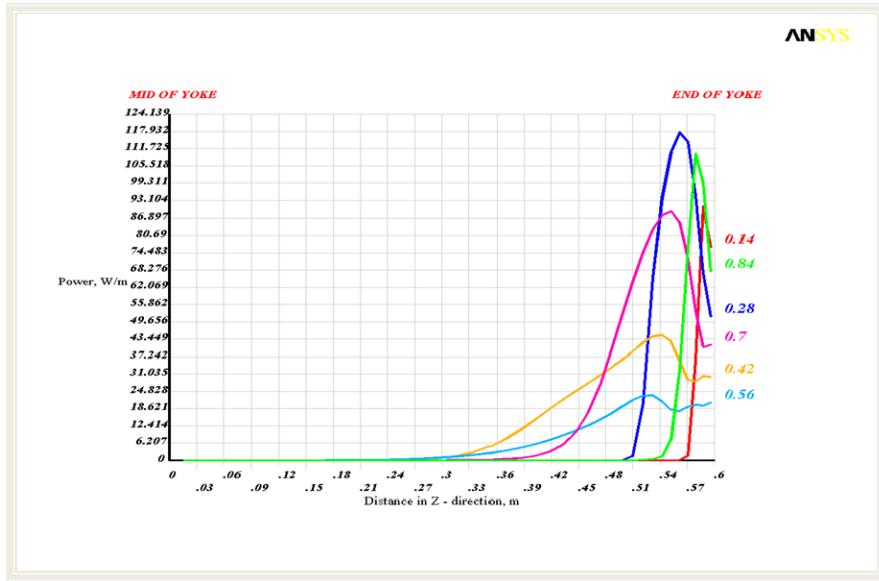


Fig. 34: Power loss as a function of longitudinal coordinate z , at different times

3.1.2.2.2.2 SIS 300 dipole

This dipole is shown schematically in Fig. 35. It is a superconducting $\cos\theta$ dipole with coil, collar, low-carbon iron and stainless-steel laminations, replacing the iron at the magnet ends. The coil is longer than the iron in order to lower the high field point. Flux created by the end of the coil enters the iron with a large B_z component which generates eddy currents in the iron lamination and in the stainless steel and collar laminations as well [21].

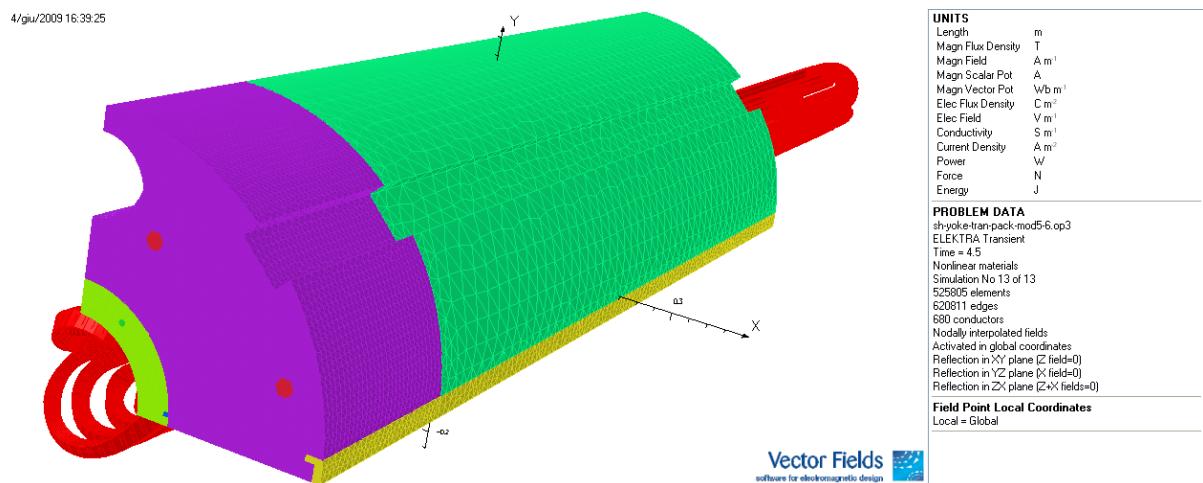


Fig. 35: Schematic view of the SIS 300 prototype dipole (coil, collar, low-carbon iron and stainless-steel ends)

The eddy current pattern was calculated using ELEKTRA[®]. At 4.5 T, on account of saturation the eddy currents penetrate more in the iron and create there higher losses than at 1.5 T [9]. We observe here on a cosθ dipole the same effect which we have previously seen on the superferric SIS 100 dipole.

3.1.2.2.3 Field variation due to variation of the packing factor

Usually the packing factor of a laminated magnet varies slightly along the trajectory on account of mechanical imperfections. Stacked laminated dipoles with large bending angle have to be built out of several sectors. At the boundaries the packing factor is of course reduced, the DC field B_y , therefore drops and field components perpendicular to the lamination surface are created. Figure 36 shows this effect schematically. Figure 37 shows a plot of the DC field and the field lag after a magnet ramp along the trajectory. One can clearly see the coincidence of both effects at the five interfaces of the six magnet sectors [22].

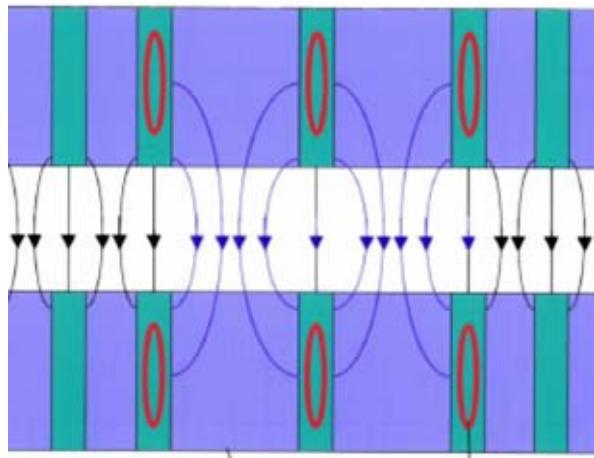


Fig. 36: Eddy currents at positions of low packing factor due to field components perpendicular to the lamination surface (courtesy of F. Klos)

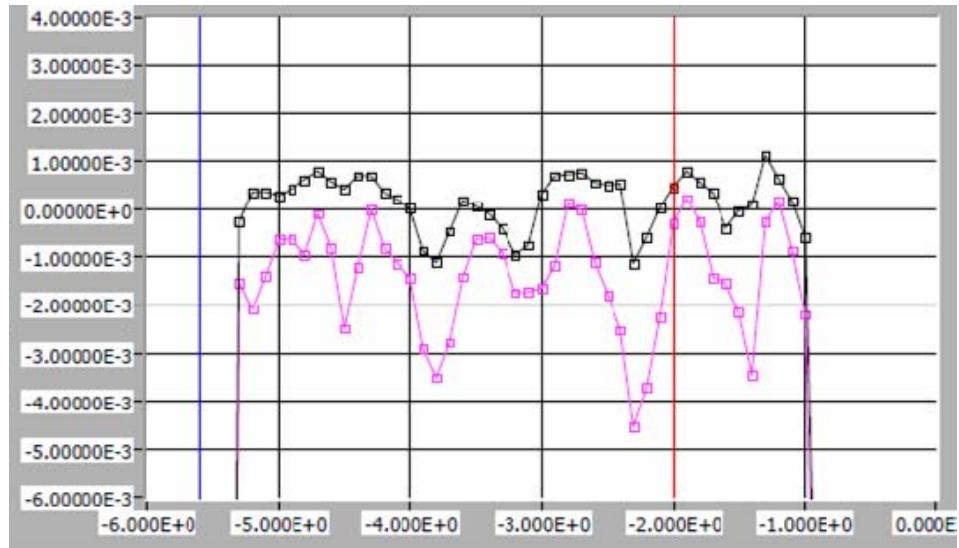


Fig. 37: Coincidence of DC field drop (dark dots) and field lag (light dots) due to eddy currents at the five sector interfaces (points with low packing factor)

3.2 Eddy currents in coils

3.2.1 Eddy currents in resistive coils

As mentioned in the introduction, eddy current effects in accelerator magnets are mostly unwanted. However, there is an example where the homogeneity of a synchrotron magnet benefits from the eddy current contribution to the field [23].

Figure 38 shows the cross-section of a typical ‘hybrid’ dipole. It is a mixture of an H-type dipole and a window-frame dipole. In order to reach the necessary ampere-windings for high field operation one has to place part of the coil between the poles. Such dipoles were built for the SPS at CERN and SIS18 at GSI.

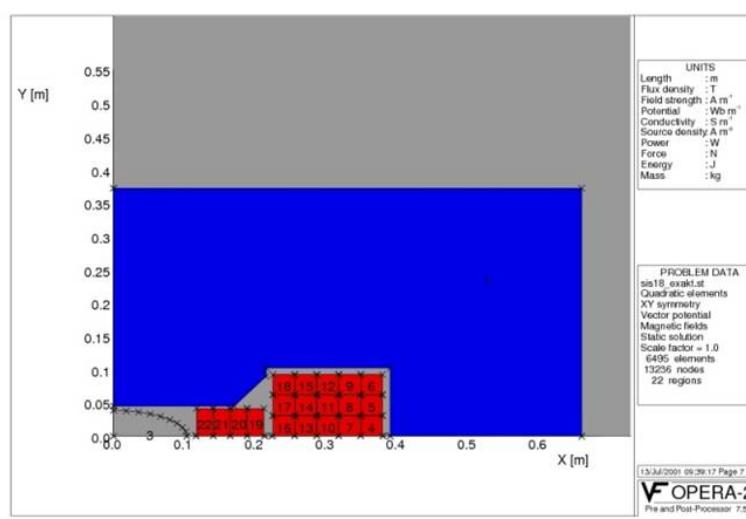


Fig. 38: Cross-section of a ‘hybrid’ dipole

Figure 39 demonstrates the effect. While the transport current (shown in conductor 1-5 with an ‘x’) is ramped up, eddy currents are induced in each of the massive copper conductor (here shown only in conductor 6), which (for symmetry reasons) introduce a positive sextupole-like component.

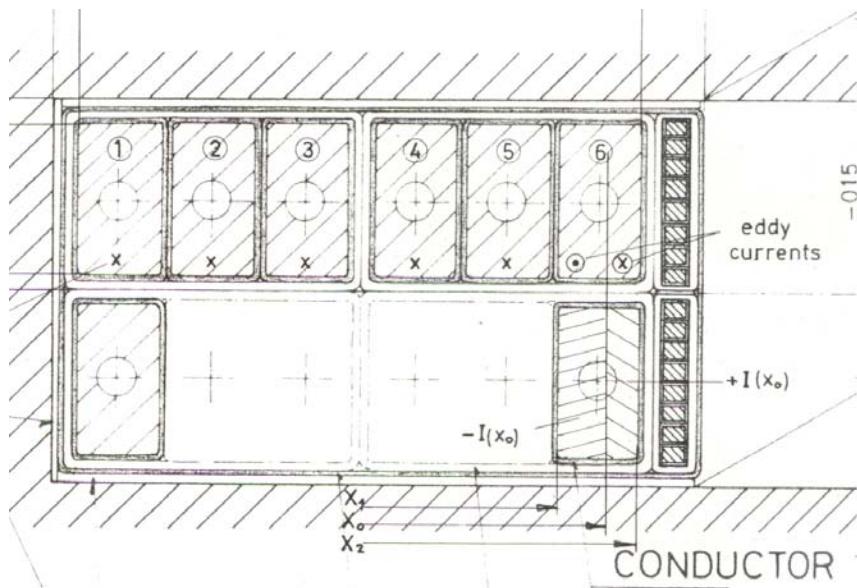


Fig. 39: Cross-section of a copper coil between the pole shoes of a ‘hybrid’ dipole

Figure 40 shows the measured field quality, in the DC mode (three currents) without and in the ramped mode (maximum current 380 A) with eddy currents. The eddy currents compensate the negative sextupolar component. For a ramp rate of 1000 A/s the field quality is almost ideal.

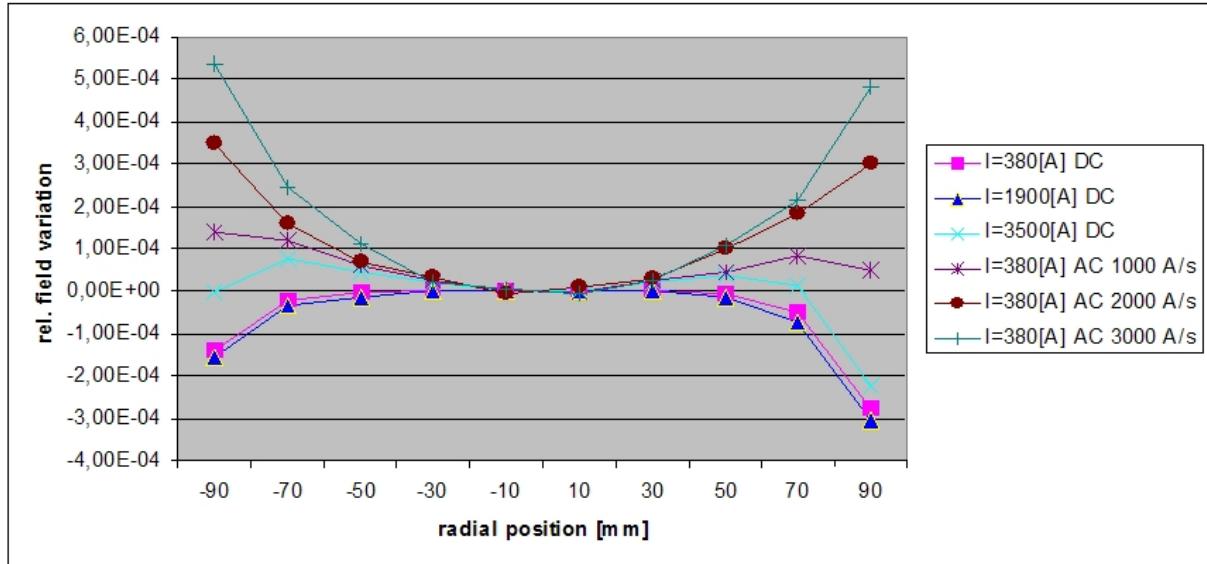


Fig. 40: Field quality in DC mode and at different ramp rates (Courtesy of F. Klos)

3.2.2 Eddy currents in superconducting coils

Coupling between strand filaments and coupling between the strands of a cable are (besides magnetization) the biggest source of losses and field quality distortion in a superconducting coil. Eddy currents in the copper matrix of the strand also play a role. These effects are covered extensively in the literature [Wilson, Verweij].

3.3 Eddy currents in mechanical structure

The mechanical structure of a magnet normally consists of metal parts to provide a stiff structure with a high Young modulus. Therefore, eddy currents are induced if these metallic parts are subjected to a time-varying field. This can happen in

- brackets and welding plates/tension bars
- collars
- end plates and mirror plates
- pins, keys and rods
- shells (helium containment for example) and shields etc.

Of course, closed flux loops as would be formed by welding seams on the pole surface are strictly forbidden.

Figure 41 shows the cold mass of an LHC dipole including its mechanical structure.

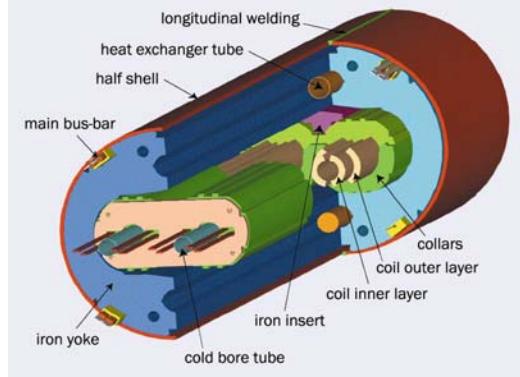


Fig. 41: Cold mass of the LHC dipole with structural parts

Very often eddy currents occur only at high fields, when, because of saturation, flux leaks out of the iron yoke. For example, for the R&D magnet GSI001 the nominal field was increased from 3.5 T to 4 T.

Figure 42 (a) demonstrates that at this high saturation level the field leaks and induces eddy currents in the copper shield. Figure 42 (b) gives the power density for a ramp rate of 4 T/s. The maximum value is 270 kW/m³ [24].

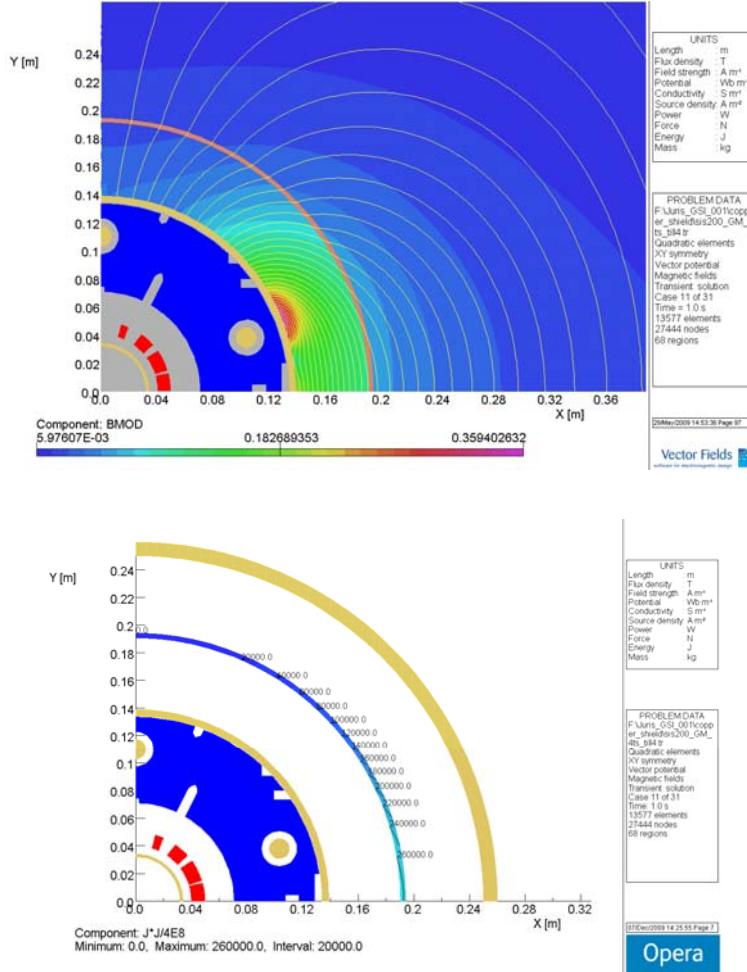


Fig. 42: (a) B_{mod} at 4T nominal bore field and (b) corresponding power loss density in the copper shield at 4 T/s

Collars and iron laminations are held together by rods, pins and keys. They usually form a flux loop. Rods should be insulated from the iron by insulating sleeves and washers. If that is not possible for mechanical reasons, they have to be placed in positions where the corresponding flux is minimal (Fig. 43) [21].

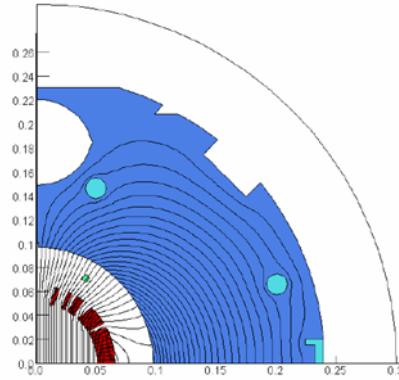


Fig. 43: Optimal position of the rods (flux enclosed between symmetric rods is minimized)

Figure 44 shows the eddy current density in rods and keys of the SIS 300 dipole prototype, calculated with ELEKTRA[©]. The iron key shows a contribution only at the end, the central part ‘sees’ no flux change [21].

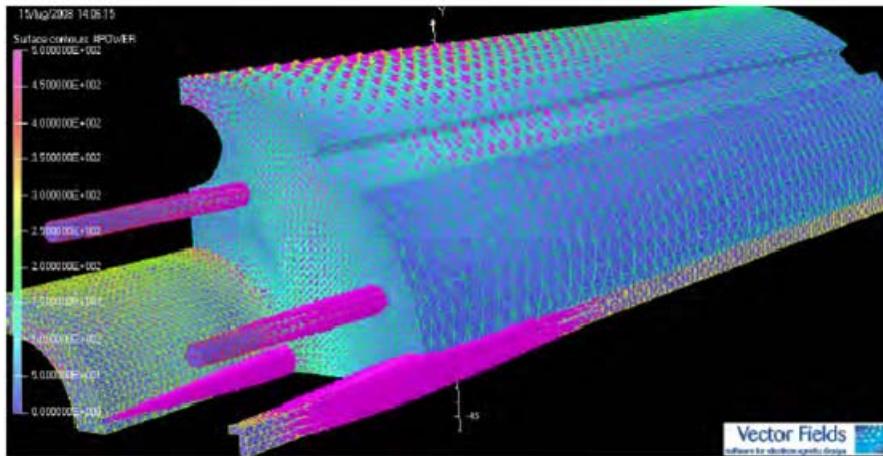


Fig. 44: Eddy current density in rods and keys of the mechanical structure of the SIS 300 dipole prototype

3.4 Eddy currents in beam pipe

The analytical formulas for a round beam pipe were given in Eq. (37).

Here we discuss elliptical beam pipes for the special case of a pipe at 4 K, as part of the cold mass of the synchrotron dipole of the planned synchrotron SIS 100. In Fig. 45 is shown the eddy current loss density in watts per metre of an infinitely long stainless-steel pipe at 4 K as the result of an ANSYS-calculation [25], [26]. The main loss occurs on both sides of the pipe. The wall thickness d is 0.3 mm, the loss (averaged over the whole pipe) is 4.9 W/m for a ramp rate of 4 T/s. Here $a = 64$ mm and $b = 29$ mm are major and minor semi-axes of the elliptical cross-section.

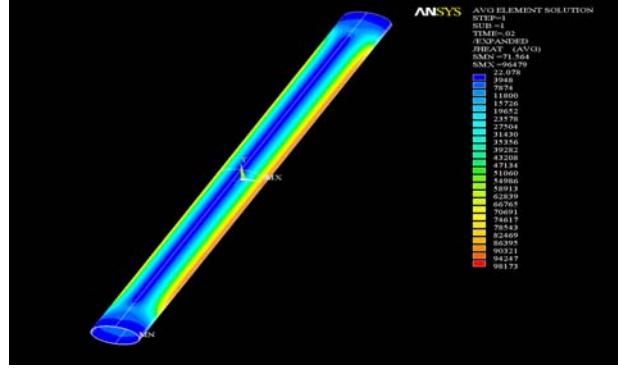


Fig. 45: Eddy current loss density distribution in an elliptical beam pipe at 4 K

An analytical formula can be derived for this simple 2D geometry (l length of the pipe) [27]:

$$P = \frac{a^2}{\rho} \dot{B}^2 f(\varepsilon) a d l \quad (43)$$

with

$$\varepsilon = \sqrt{1 - \left(\frac{b}{a}\right)^2}$$

$$f(\varepsilon) = \int_0^{2\pi} \cos^2 \eta * \sqrt{1 - \varepsilon^2 \cos^2 \eta} d\eta .$$

Please note again the quadratic dependence on the ‘relevant’ dimension a , whereas the product $f(\varepsilon)adl$ is proportional to the volume.

The main reason for the use of a cold beam pipe is cryogenic pumping, especially of hydrogen. That requires that large parts of the pipe stay below 10 K. Of course, the heat produced by the eddy currents has to be carried away. One option is to use cooling pipes connected to the beam pipe with liquid helium flowing inside. It is clear that the design must avoid any flux loop coupled to the main field component (Fig. 46) [28].

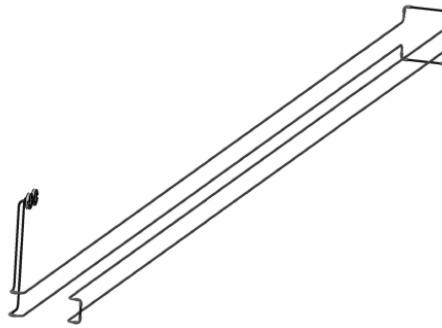


Fig. 46: Design of the cooling pipes of an elliptical beam pipe

In addition, reinforcement ribs have to be added to the tube (in order to guarantee the mechanical stability), leading to a real 3D problem. ANSYS calculations were made for the central and the end part of the beam pipe [29]. Figure 47 shows the results.

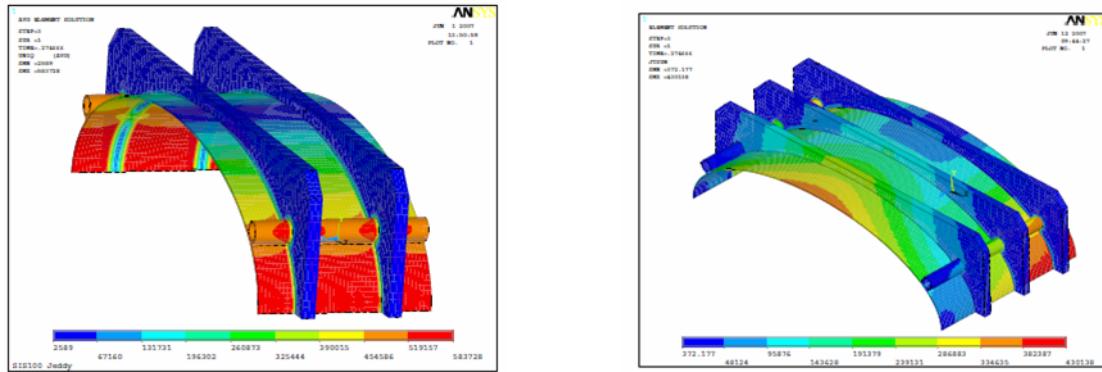


Fig. 47: Eddy current density distribution in the central (left) and end part of the beam pipe (right)

Unfortunately, the average loss in watts per metre is now increased from 4.9 W/m to 8.7 W/m. Electrical insulation of the cooling pipes helps to reduce the losses.

4 Magnet design principles

Based on the previous section we are now able to summarize the design principles for pulsed or ramped magnets, in order to minimize eddy current effects:

- The magnet must be laminated. The lamination sheets must be well insulated. The thickness of the sheet is determined by the ramp rate or the frequency of the operation.
- Iron with an increased resistance (mostly by increasing the silicon content) should be used. However, more important is a low coercivity.
- The magnet design must be appropriate:
 - Saturation should be avoided in the magnet-end field region. Owing to the anisotropy and non-linearity of the permeability, the longitudinal field component penetrates further into the iron at high fields (static phenomenon!), creating more eddy currents in this end region in case of a time-varying field (see 3.1.2.1.3).
 - Saturation should also be avoided in the central part of the magnet, because flux leakage may induce eddy currents in structural components [30].
 - A Rogowski or a simple chamfer pole profile at the magnet pole ends reduces the power loss, compared to a ‘sharp edge’, since the longitudinal magnetic field component perpendicular to the lamination sheets comes down.
 - Vertical or horizontal slits in the end laminations limit the ‘relevant dimension’ of the eddy currents and therefore reduce the effects [31]. Be aware of problems with mechanical stability and field quality.
 - Non-conductive pole material at the magnet ends prevents eddy currents, but reduces field quality.
 - ‘Long’ magnets are advantageous, also from the eddy current perspective, since most of the eddy current contributions come from the magnet ends.
 - Be aware that the coil shape at the magnet end has an influence on the horizontal field component.

- For superconducting coils special low-loss wire and cable must be used.
- The mechanical structure of the magnet must be carefully designed to avoid flux loops and flux concentration in bulky components, especially if they are made out of low-carbon iron. Stainless steel may be a better choice.
- Replace conductive materials by non-conductive materials wherever possible.

Field control is a very efficient way to reduce the waiting time for the eddy currents to die out [32].

5 Summary

This paper covers the main eddy current effects in accelerator magnets: field modification (time delay and field quality) and resistive power losses. In the first part, starting from the Maxwell equations, a basic understanding of the processes was given and explained with examples of simple geometry and time behaviour. The ‘magnetic diffusion’ approach was used and the ‘small perturbation’ method was explained. Useful formulas were derived for an analytic estimate of the size of the effects.

In the second part, the effects in real magnets were analysed and described with a comparison between numerical and measured results. It was demonstrated that eddy currents play a role in the yoke, coil, mechanical structure, and beam pipe of a magnet.

Finally, based on the previous main parts, design recommendations were given, regarding how to minimize eddy current effects.

Acknowledgements

The author wishes to thank Dr. S.Y. Shim for his help during the preparation of this paper and Prof. A. Kalimov for many helpful discussions.

References

- [1] E.E. Kriezis, T.D. Tsiboukis, S.M. Panas, and J.A. Tegopoulos, Eddy currents: theory and applications, Proc. IEEE **80** (1992) 1599–1589.
- [2] M. I. Gonzalez, Experiments with eddy currents: the eddy current brake, Eur. J. Phys. **25** (2004) 463–468.
- [3] G. Brianti, I. Gumowski, and K. Schindl, Behaviour of the magnetic induction B in solid steel cores submitted to an excitation variable with time, CERN/SI/Int. DL/71-3 (1971).
- [4] A. Kalimov, private communication.
- [5] K. Halbach, Some eddy current effects in solid core magnets, Nucl. Instrum. Methods, **107** (1973) 529–540.
- [6] D. Reistad, Simple Expressions to evaluate eddy current effects in laminated magnets, CERN/ISR-BT/68-42 (1968).
- [7] A. Asner, G. Brianti, and I. Gumowski, Eddy currents in pulsed magnets, CERN/SI/Int. DL/69-2 (1969).
- [8] P. Sherbakov et al., Design Report SIS 300 6 T dipole (2004).

- [9] F. Alessandria et al., Technical Design Report of a superconducting model dipole for SIS 300, INFN/T-09/4, May 2009.
- [10] S. Russenschuck, CAS Magnets: Foundations of analytical and numerical field computation I, these proceedings.
- [11] E. Fischer, P. Shcherbakov, and R. Kurnyshov, *Cryogenics*, **47** (2007) 583–594.
- [12] A. Kalimov, F. Klos, and B. Langenbeck, and G. Moritz, Dynamic processes in laminated magnets: simulation and comparison with experimental results, *IEEE Trans. Appl. Superconductivity*, **12** (2002) 98–101.
- [13] S.Y. Shim, F. Klos, C. Muehle, and G. Moritz, CNAO Scanning Magnet and its Operation Property, MT-INT-SYS-FK-CM-GM-2008-001 (2008).
- [14] S.Y. Shim, private communication.
- [15] A. Kalimov, E. Fischer, G. Hess, G. Moritz, and C. Muehle, Investigation of the power losses in a laminated dipole magnet with superconducting coils, *IEEE Trans. Appl. Superconductivity*, **14** (2004) 267–270.
- [16] S. Koch, H. de Gersem, E. Fischer, G. Moritz, and T. Weiland, Simulation of eddy current losses in a fast ramped superconducting dipole magnet, *IEEE Trans. Appl. Superconductivity*, **16** (2006) 334–337.
- [17] S. Koch, H. de Gersem, and T. Weiland (TU Darmstadt), private communication.
- [18] P. Shcherbakov, E. Fischer, and R. Kurnyshov, 3D magnetic field and eddy loss calculations for iron dominated accelerator magnets using ANSYS compared with results of non-commercial codes, WEPLS094, Proceedings of EPAC 2006, Edinburgh, Scotland, 2006.
- [19] E. Fischer, P. Shcherbakov, and R. Kurnyshov, Analysis of the eddy current relaxation time effects in FAIR SIS100 main magnets, *IEEE Trans. Appl. Superconductivity*, **17** (2007) 1078–1082.
- [20] E. Fischer, R. Kurnyshov, G. Moritz, and P. Shcherbakov, 3D transient process calculations for fast cycling accelerator magnets, *IEEE Trans. Appl. Superconductivity*, **16** (2006) 497–410.
- [21] M. Sorbi, F. Alessandria, G. Bellomo, S. Farinon, U. Gambardella, P. Fabbricatore, R. Musenich, and G. Volpini, Electromagnetic design of the coil ends for the FAIR SIS300 Model Dipole, in Proceedings ASC08, Chicago, 2008.
- [22] F. Klos, private communication.
- [23] A. Asner, The effects of eddy currents in the Booster bending magnets excitation windings on the magnetic field uniformity, CERN/SI/Note MAE/69-6 (1969).
- [24] H. Leibrock, Calculations of eddy current losses, MT-INT-HL-2005-001 (2005).
- [25] E. Fischer et al., Numerical analysis of the operation parameters of fast cycling superconducting magnets, Proceedings of ASC08, Chicago, 2008.
- [26] E. Fischer, R. Kurnyshov, and P. Sherbakov, Analysis of coupled electromagnetic thermal effects in superconducting accelerator magnets, IOP Conf. Series, 2008, 97.
- [27] H. de Gersem, S. Koch et al., Transient finite element simulation of the eddy current losses in the beam tube of the SIS 100 magnet during ramping, *IEEE Trans. Appl. Superconductivity*, **18** (2008) 1613–1616.
- [28] S.Y. Shim, M. Kauschke, A. Krämer, and G. Moritz, Thermal analysis of UHV beam tube of SIS100 for FAIR, Proceedings of ICEC22- ICMC2008, 705, (2008).

- [29] S. Y. Shim, Analysis of Eddy current loss and cooling of SIS100 beam tube, GSI Tech. Report, MT-INT-SYS-2007-002 (2007).
- [30] T. Zickler, CAS Magnets: Basic magnet design IV, these proceedings.
- [31] M. Negrazus, D. George, V. Vrankovic, and M. Werner, Eddy current reduction in fast ramped bending magnets, IEEE Trans. Appl. Superconductivity, **16** (2006) 228–230.
- [32] A. Peters, E. Feldmeier, et al., Magnetic field control in synchrotrons, In Proceedings of PAC 2009, Vancouver, Canada.

Bibliography

- Y. Iwasa, *Case Studies in Superconducting Magnets* (Plenum Press, New York and London, 1994).
- Heinz E. Knoepfel, *Magnetic Fields* (John Wiley and Sons, New York, 2000).
- Jack T. Tanabe, *Iron Dominated Electromagnets: Design, Fabrication, Assembly and Measurements* (World Scientific, Singapore, 2005).
- J.A. Tegopoulos, and E.E. Kriezis, *Eddy Currents in Linear Conducting Media* (Elsevier, Amsterdam, 1985).
- A.P. Verweij, *Electrodynamics of Superconducting Cables in Accelerators*, PhD Thesis, University of Twente, The Netherlands, 1995.
- M.N. Wilson, *Superconducting Magnets* (Clarendon Press, Oxford, 1983).

Injection and extraction magnets: kicker magnets

M.J. Barnes, L. Ducimetière, T. Fowler, V. Senaj, L. Sermeus
CERN, Geneva, Switzerland

Abstract

Each stage of an accelerator system has a limited dynamic range and therefore a chain of stages is required to reach high energy. A combination of septa and kicker magnets is frequently used to inject and extract beam from each stage. The kicker magnets typically produce rectangular field pulses with fast rise- and/or fall-times, however, the field strength is relatively low. To compensate for their relatively low field strength, the kicker magnets are generally combined with electromagnetic septa. The septa provide relatively strong field strength but are either DC or slow pulsed. This paper discusses injection and extraction systems with particular emphasis on the hardware required for the kicker magnet.

1 Introduction

An accelerator stage has limited dynamic range: a chain of accelerator stages is required to reach high energy. Thus beam transfer into (injection) and out of (extraction) an accelerator is required. The design of the injection and extraction systems aims to achieve the following:

- minimize beam loss,
- place the newly injected or extracted particles onto the correct trajectory, with the correct phase space parameters.

A combination of septa and kickers is frequently used for injection and extraction. Septa can be electrostatic or magnetic: they provide slower field rise- and fall-times, but stronger field, than kicker magnets. Some septa are designed to be operated with DC. Kicker magnets provide fast field rise- and fall-times, but relatively weak fields.

In general, a septum (plural: septa) is a partition that separates two cavities or spaces. In a particle accelerator a septum is a device which separates two field regions. Important features of septa are an ideally homogeneous (electric or magnetic) field in one region, for deflecting beam, and a low fringe field (ideally zero magnetic and electric field) next to the septum so as not to affect the circulating beam. Hence a septum provides a space separation of circulating and injected/extracted beam. In contrast a kicker magnet provides time selection [separation] of beam to be injected/extracted: a kicker system is used for fast, single-turn, injection and extraction.

The processes of injection and extraction are covered in the proceedings of this CERN Accelerator School, in the paper *Injection and extraction magnets: septa*; the aforementioned paper also discusses the hardware associated with septa. The present paper discusses the hardware associated with kicker magnets.

The field produced by a kicker magnet must rise/fall within the time period between the beam bunches (see Sections 2 and 3). In addition, the magnetic field must not significantly deviate from the flat top of the pulse or from zero between pulses (i.e., very small ripple/excursions). Typical field rise/fall-times range from tens to hundreds of nanoseconds and pulse widths range from tens of nanoseconds to tens of microseconds. If a kicker exhibits a time-varying structure, in the field pulse

shape, this can translate into small offsets with respect to the closed orbit (betatron oscillations). Thus a fast, low-ripple, kicker system is generally required.

2 Single-turn (fast) injection

Figure 1 shows an example of fast single-turn injection in one plane. The injected beam passes through the homogeneous field region (gap) of the septum: circulating beam is in the field-free region (i.e., space separation of injected and circulating beam). The septum deflects the injected beam onto the closed orbit at the centre of the kicker magnet; the kicker magnet compensates the remaining angle. The septum and kicker are either side of a quadrupole (defocusing in the injection plane) which provides some of the required deflection and minimizes the required strength of the kicker magnet.

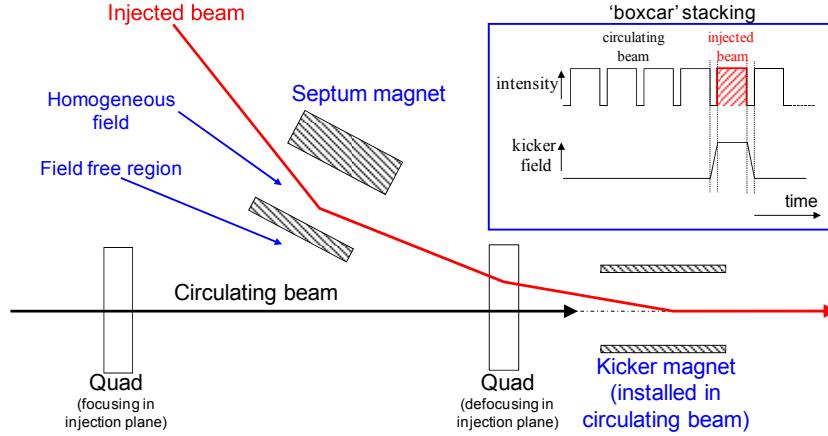


Fig. 1: Fast single-turn injection in one plane

A kicker magnet is installed in the accelerator and hence the circulating beam is in the aperture of the kicker. Thus the kicker field must rise from zero to full field in the time interval between the circulating beam and the start of the injected beam (Fig. 1, top right) and fall from full field to zero in the time interval between the end of the injected beam and the subsequent circulating beam (Fig. 1, top right). The kicker system is described in more detail in Section 4.

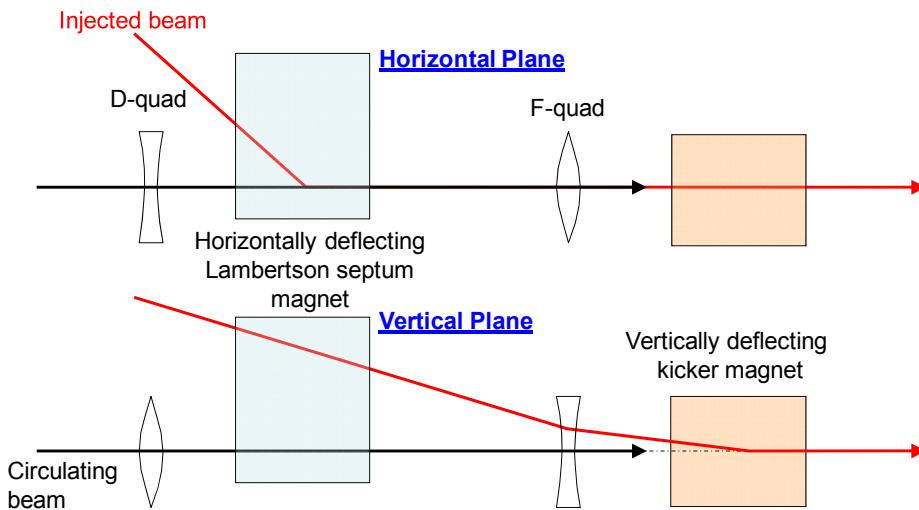


Fig. 2: Fast single-turn injection in two planes

Figure 2 shows an example of fast single-turn injection in two planes: a Lambertson septum is used for a two-plane injection scheme. The injected beam passes through the homogeneous field region of the septum: circulating beam is in the field-free region of the septum. In the example shown in Fig. 2 the septum deflects the beam horizontally and the downstream kicker magnet deflects the beam vertically onto the closed orbit of the circulating beam. The septum and kicker are either side of an F-quadrupole (horizontally focusing but vertically defocusing) to minimize the required strength of the kicker magnet. The Lambertson septum magnet is discussed in more detail in the proceedings of this CAS, in Section 5.2.4 of *Injection and extraction magnets: septa*.

3 Single-turn (fast) extraction

Extraction is the process of ejecting a particle beam from an accelerator and into a transfer line or a beam dump, at the appropriate time, while minimizing beam loss and placing the extracted particles onto the correct trajectory, with the correct phase space parameters. Extraction usually occurs at higher energy than injection, hence stronger elements (larger $\int B \cdot dl$) are required. At high energies many kicker and septum modules may be needed. To reduce the required strength of the kicker magnet, a closed orbit bump can be applied to bring the circulating beam near to the septum.

Figure 3 shows an example of fast single-turn extraction in one plane. The kicker magnet deflects the entire beam into the septum in a single turn [time selection (separation) of beam to be extracted]. The extracted beam passes through the homogeneous field region of the septum: the circulating beam, prior to extraction, is in the field-free region of the septum (space separation of circulating and extracted beam). The septum deflects the entire kicked beam into the transfer line.

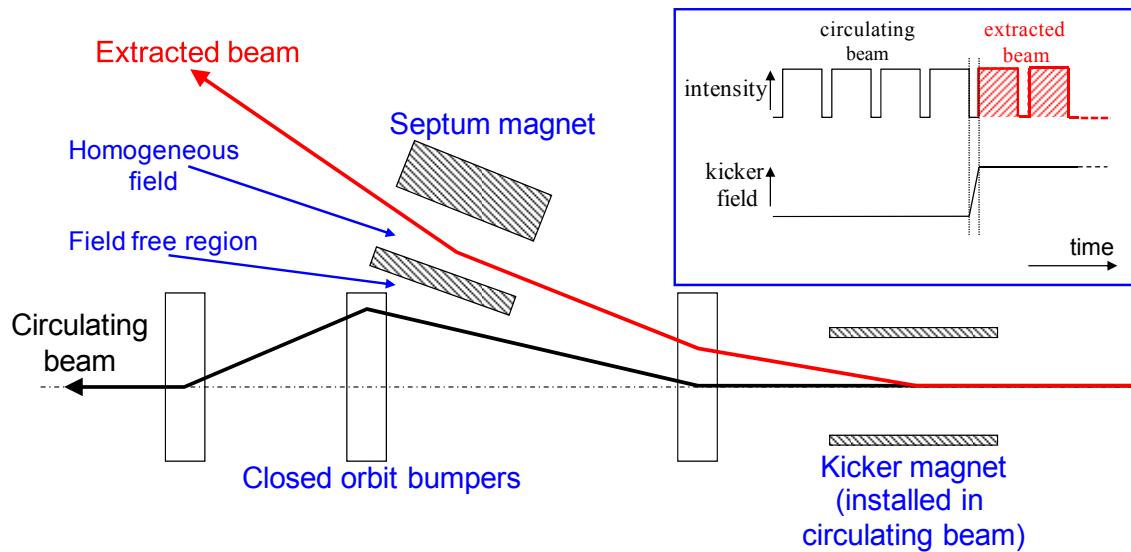


Fig. 3: Fast single-turn extraction in one plane

The kicker magnet is installed in the accelerator and hence the circulating beam is in the aperture of the kicker. Thus the kicker field must rise from zero to full field in a beam-free time interval deliberately created in the circulating beam (Fig. 3, top right). The entire beam is generally extracted and hence fast fall-time is typically not required: however, sometimes, bunch-by-bunch transfers are made and then the field of the kicker magnets must have fast rise- and fall-times [1].

4 Kicker system

4.1 Overview

Figure 4 shows a simplified schematic of a kicker system. The main sub-systems ('components') of a kicker system are

- PFL = Pulse Forming Line (coaxial cable) or PFN = Pulse Forming Network (lumped elements),
- kicker magnet,
- fast, high power, switch(es),
- RCPS = Resonant Charging Power Supply,
- transmission line(s) [coaxial cable(s)],
- terminators (resistive).

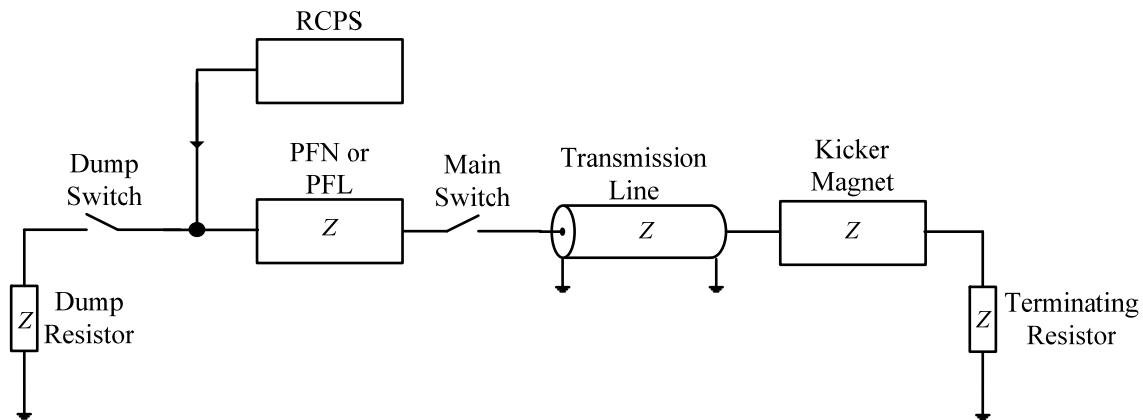


Fig. 4: Simplified schematic of a kicker system

4.2 Pulse forming circuit: general case

Figure 5 shows a simplified schematic of a pulse-forming circuit: the switch is initially open and the coaxial cable (PFL) is pre-charged, through the large valued resistor or inductor, to a voltage V .

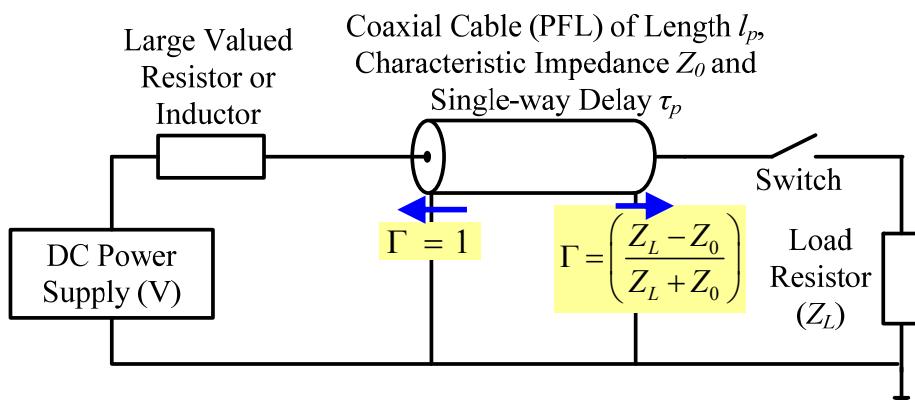


Fig. 5: Simplified schematic of a pulse-forming circuit

At time $t = 0$, when the ideal switch closes, the load voltage is given by Eq. (1):

$$V_L = \left(\frac{Z_L}{Z_0 + Z_L} \right) V = \alpha V , \quad (1)$$

where

$$\alpha = \left(\frac{Z_L}{Z_0 + Z_L} \right);$$

V_L is the load voltage (V);

V is the initial voltage to which the PFL is charged (V);

Z_L is the load impedance (Ω);

Z_0 is the characteristic impedance of the PFL (Ω).

Figure 6 shows the lattice diagram for the general case (impedances not necessarily matched) for the voltage on the PFL. At time $t = 0$, when the switch closes, a voltage pulse of ' $(\alpha - 1)V$ ' propagates from the load end of the PFL towards the charging end. At the charging end of the PFL the reflection coefficient (Γ) is +1 and hence a voltage of ' $(\alpha - 1)V$ ' is reflected back towards the load end of the PFL. At the load end of the PFL the reflection coefficient is given by Eq. (2):

$$\Gamma = \left(\frac{Z_L - Z_0}{Z_L + Z_0} \right) = \beta , \quad (2)$$

and hence a voltage of ' $\beta(\alpha - 1)V$ ' is reflected back towards the charging end of the PFL, etc.

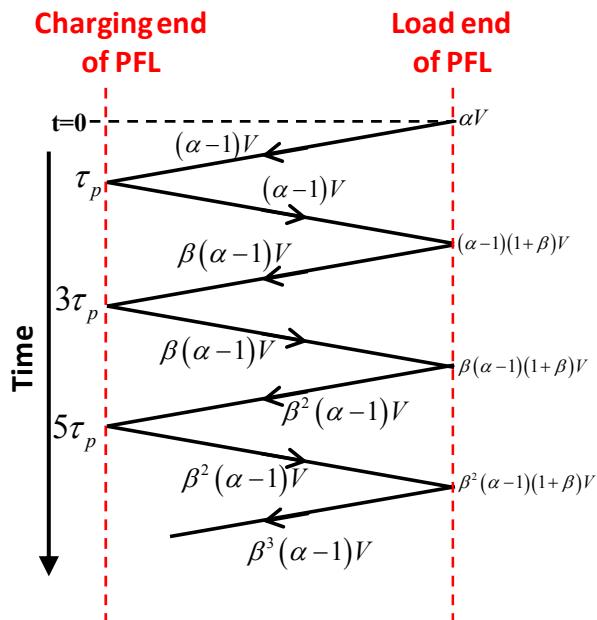


Fig. 6: Lattice diagram for the pulse forming circuit of Fig. 5: general case

Impedances need to be matched to avoid reflections, i.e., from Eq. (2), when $Z_L = Z_0 \rightarrow \beta = 0$. When the impedances are matched:

- PFN/PFL is charged to a voltage V by the RCPS;
- The Main Switch (MS) closes (Fig. 4) and, for a matched system, a pulse of magnitude $V/2$ is launched, through the transmission line, towards the kicker magnet. A voltage pulse of magnitude $-V/2$ propagates from the load end of the PFN/PFL towards the charging end;
- Once the current pulse reaches the (matched) terminating resistor full-field has been established in the kicker magnet (Section 4.3.3);
- The length of the pulse in the magnet can be controlled in length, between 0 and $2\tau_p$, by adjusting the timing of the Dump Switch (DS) (Fig. 4) relative to the MS.

Note: if the magnet termination is a short-circuit, the magnet current is doubled but the required ‘fill-time’ of the magnet is doubled too (Section 4.3.6). In this case the DS may be an inverse diode: the inverse diode ‘automatically’ conducts when the PFN voltage reverses, at the charging end of the PFL/PFN, but there is no control over the pulse-length in the magnet.

4.3 Kicker magnet

4.3.1 History

Figure 7 shows a 1970’s ‘plunging’ kicker magnet which was hydraulically operated [2]: the aperture was too small for the kicker to be in the beam-line during circulating beam. Developments leading to higher current pulses permitted larger apertures: kicker magnets developed later at CERN were not hydraulically operated.



Fig. 7: 1970’s plunging kicker magnet for ejection from the CERN Antiproton Accumulator (AA)

4.3.2 Overview

Figure 8 shows a cross-section of a typical C-core kicker magnet. Fast kicker magnets are generally ferrite-loaded transmission line type magnets with a rectangular-shaped aperture of dimensions H_{ap} by V_{ap} (Fig. 8).

The flux density in the aperture (B_y) of the kicker is given by Eq. (3):

$$B_y \cong \mu_0 \left(\frac{N \cdot I}{V_{ap}} \right), \quad (3)$$

where

- μ_0 is permeability of free space ($4\pi \times 10^{-7}$ H/m);
- N is the number of turns (usually 1 for a kicker magnet);
- I is magnet current (A);
- V_{ap} is the distance between the inner edges of the ‘legs’ of the ferrite (m).

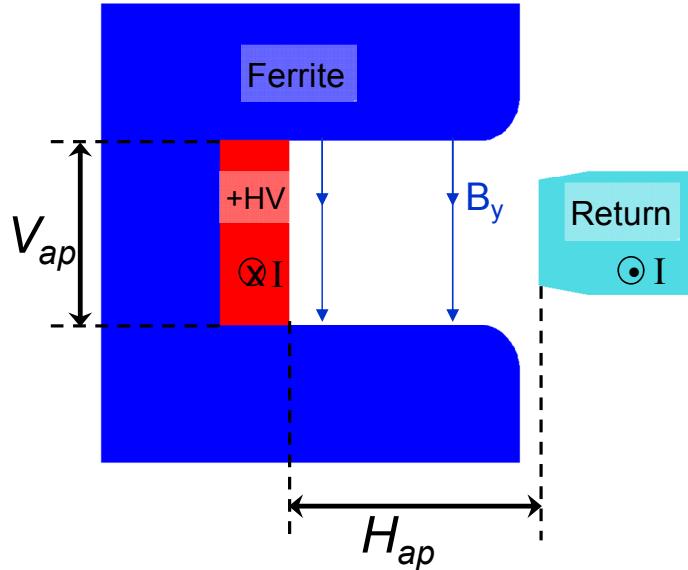


Fig. 8: Cross-section of a typical C-core kicker magnet

Skin effect and proximity effect result in current flow on the inside surface of both the High Voltage (HV) and return conductors. Hence inductance is given by Eq. (4):

$$L_{m/m} \cong \mu_0 \left(\frac{N^2 \cdot H_{ap}}{V_{ap}} \right), \quad (4)$$

where

- H_{ap} is the distance between the inner edges of the HV and return conductors (m);
- $L_{m/m}$ is inductance per metre length of the kicker magnet (H/m).

Since kicker magnets generally need to be fast they usually have only single-turn coils: multi-turn coils are used only for slower, lumped-inductance kicker magnets.

The deflection of a charged particle beam in a magnetic field and an electric field is given by Eq. (2) and Eq. (3), respectively, of *Injection and extraction magnets: septa* in the proceedings of this CAS. For a transmission line kicker magnet, where ferrite C-cores are sandwiched between HV capacitance plates (see Section 4.3.3.2), end effects result in an effective length of each end-cell approximately $(V_{ap}/4)$ greater than its physical length [3]. Thus the effective length of the kicker magnet is increased by an amount approximately equal to $(V_{ap}/2)$, in comparison to the physical length of the aperture.

4.3.3 Design options for kicker magnets

Design options for kicker magnets include [2]

- Type: ‘lumped-inductance’ or ‘transmission-line’ (with specific characteristic impedance (Z))?
- Machine vacuum: install in or external to machine vacuum?
- Aperture: window-frame, closed C-core or open C-core?
- Termination: matched impedance or short-circuited?

4.3.3.1 Lumped-inductance kicker magnet

Although a lumped-type magnet has a simple structure, in most cases it cannot be applied to a fast kicker system because of its impedance mismatch and its slow response [4]. The lumped-inductance kicker is generally useable only when a rise-time above a few hundreds of nanoseconds is required. The lumped-inductance kicker either has a resistor in series with the kicker magnet input or else the resistor is omitted. In both cases the kicker magnet only sees (bipolar) voltage during pulse rise and fall. With a short-circuit termination, magnet current is doubled for a given PFN/PFL voltage and system characteristic impedance.

For a magnet inductance L_m in series with a parasitic inductance L_s , combined with a pulse generator of impedance Z , the rise of the magnet current is exponential with a time constant t_{cl} given by Eq. (5):

$$t_{cl} = \left(\frac{L_m + L_s}{Z} \right). \quad (5)$$

For a short-circuit magnet the 5% to 95% rise-time of the current is as long as three time-constants. To help to overcome the long rise-time a capacitor can be connected in parallel with, at the entrance of, the lumped-inductance magnet, but this can provoke some overshoot.

4.3.3.2 Transmission-line kicker magnet

To overcome the long rise-time, the first transmission-line kicker magnet was developed at CERN in the early 1960s [4]. A transmission-line magnet consists of few to many ‘cells’ to approximate a broadband coaxial cable (Fig. 9). Ferrite C-cores are sandwiched between HV capacitance plates: plates connected to ground are interleaved between the HV plates. The HV and ground plates form a capacitor to ground (Fig. 9). One C-core, together with its ground and HV capacitance plates, is termed a cell. Each cell conceptually begins and ends in the middle of the HV capacitance plates. The ‘fill-time’ of the magnet (τ_m) is the delay required for the pulse to travel through the ‘ n ’ magnet cells.

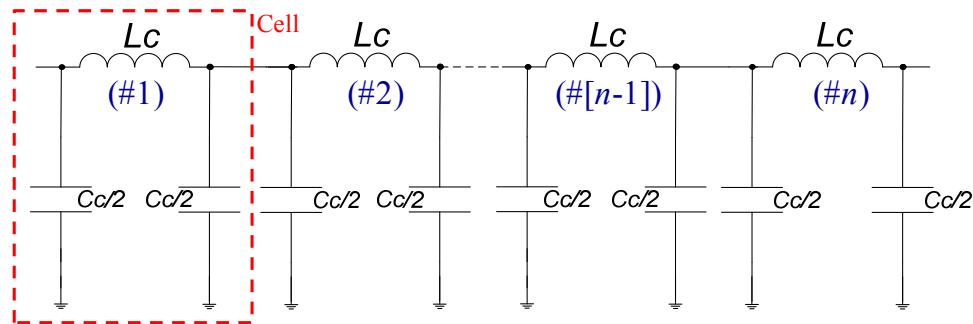


Fig. 9: Simplified equivalent electric circuit of a transmission-line kicker magnet

The characteristic impedance (Z) of the kicker magnet, which is matched to the impedance of the generator to minimize reflections [5, 6] (see Section 4.2), is given by Eq. (6):

$$Z = \sqrt{\frac{L_c}{C_c}} , \quad (6)$$

where

- L_c is the inductance of a cell of the kicker magnet (H),
- C_c is the capacitance of a cell of the kicker magnet (F).

The fill-time of the kicker magnet is given by Eq. (7):

$$\tau_m = n\sqrt{L_c \cdot C_c} = n \left(\frac{L_c}{Z} \right) = \left(\frac{L_m}{Z} \right) , \quad (7)$$

where

- neglecting end effects, $L_m = n(L_c)$,
- L_m is the inductance of the kicker magnet (H).

For a kicker magnet terminated with a matched resistor, field rise-time starts with the beginning of the voltage pulse at the entrance of the kicker magnet and ends with the end of the same pulse at the exit of the magnet. Flux is given by the time integral of the difference between the voltage at the magnet entrance and exit [Eq. (8)]:

$$\Phi = \int (V_{in} - V_{out}) dt , \quad (8)$$

where

- Φ is flux (V·s),
- V_{in} is the voltage at the entrance to the kicker magnet (V),
- V_{out} is the voltage at the exit of the kicker magnet (V).

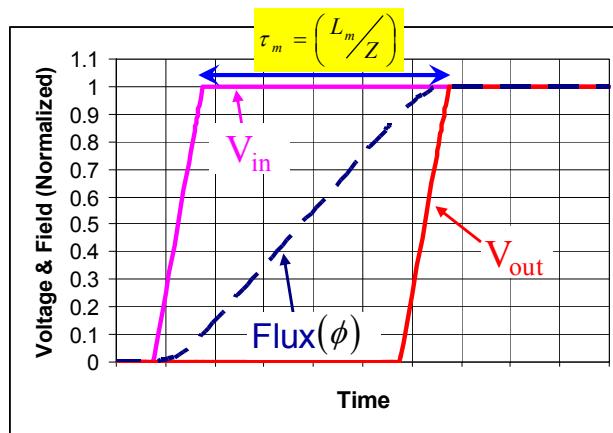


Fig. 10: Flux in an ‘ideal’ transmission-line kicker magnet

The flux builds up until the end of the voltage rise at the exit of the kicker magnet (Fig. 10), thus, for an ideal transmission-line kicker magnet, flux rise-time is given by the sum of the voltage rise-time and the magnet fill-time. Therefore it is important that the pulse not degrade while travelling through the kicker magnet. Hence the magnet cut-off frequency is a key parameter, especially with field rise-times below ~ 100 ns. Cut-off frequency (f_c) depends on series inductance ($L_c s$) associated with the cell capacitor (C_c):

$$f_c = \frac{1}{\pi \cdot \sqrt{(L_c + 4L_c s) \cdot C_c}} = \frac{Z}{\pi \cdot \sqrt{(L_c + 4L_c s) \cdot L_c}} . \quad (9)$$

Thus, to achieve a high cut-off frequency, $L_c s$ should be kept as low as possible and the cell size small. However, cells cannot be too small because adequate distance is required between HV and ground capacitance plates to avoid voltage breakdown. In addition many very small cells would significantly increase the cost and complexity of the kicker magnet.

Figure 11 shows the results of a low-voltage measurement on each of the HV capacitance plates of a transmission-line kicker magnet. The fast rise-time of the input voltage pulse, used for the measurements, contains frequency components above the cut-off frequency of the cells. Thus there is an increase in the rise-time of the voltage pulse between the entrance HV capacitance plate and the second HV capacitance plate; in addition there is significant ripple on the pulses.

The choice of the characteristic impedance, for a transmission-line magnet, depends upon the requested field rise-time and the available length for the kicker magnet. In general the highest impedance, up to the impedance of commercially available coaxial cable (50Ω), is used while still respecting the available length for the kicker magnet and chosen level of PFN voltage [2]. The higher the characteristic impedance, the higher the magnet cut-off frequency [Eq. (9)]: a high cut-off frequency reduces the field ripple.

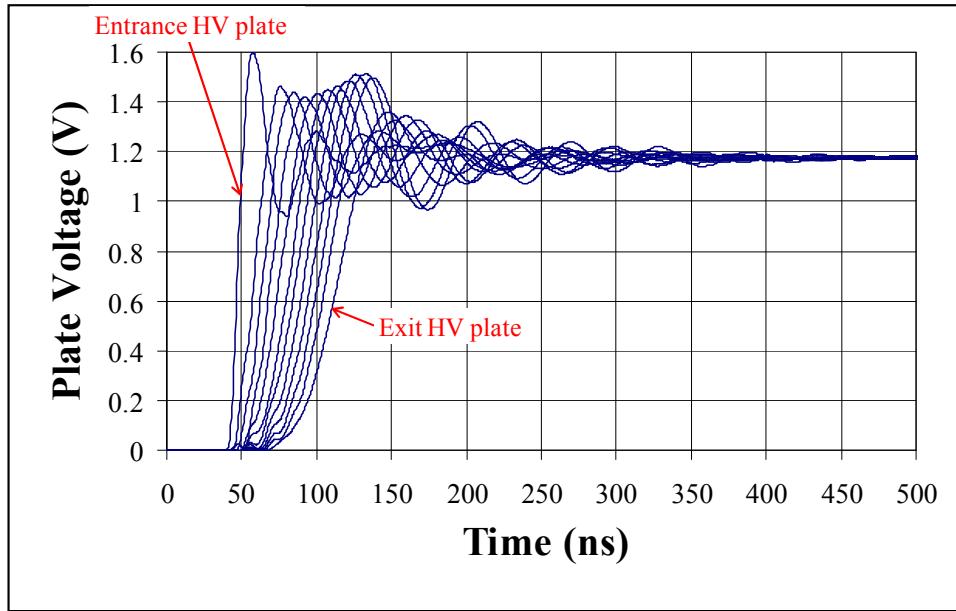


Fig. 11: Low voltage measurement on each of twelve HV capacitance plates of an eleven-cell transmission-line kicker magnet

A vacuum dielectric transmission-line kicker magnet can require a very large area of metal plates to form the capacitance, especially for a low-impedance system. Since the minimum separation of the capacitor plates is determined by practical considerations, such as voltage breakdown and cost, the capacitance is determined by the area of the plates: however, Eq. 6 shows that the capacitance (area) must increase by a factor of 4 for a decrease in impedance by a factor of 2 (neglecting edge effects). It has been shown that it is feasible to design an ‘hybrid’ kicker magnet, which is an effective transmission kicker magnet with a fast rise-time using only 3 to 5 large cells [7], but which requires large capacitance values. However, it is impractical for lower impedance magnets or for magnets with larger cell inductances to use vacuum dielectric capacitors. Hence, with the space limitation at many accelerator facilities, the application of higher permittivity dielectric capacitor media is necessary [6, 8].

4.3.4 Machine vacuum

The minimum-sized aperture for a kicker magnet can be achieved if the magnet is in vacuum: if the magnet is outside vacuum a chamber must be inserted in the kicker aperture thus increasing the dimensions of the aperture. The minimum value of both H_{ap} and V_{ap} (Fig. 8) are determined by beam parameters. Equation (4) shows that magnet inductance is proportional to H_{ap} and inversely proportional to V_{ap} . However, if V_{ap} is increased, although the magnet inductance is reduced, this is at the expense of increased current to obtain a given flux density (Eq. 3): increased current also requires, for given impedance, increased PFN voltage and thus increased insulation requirements.

Advantages of putting a transmission line kicker magnet in the machine vacuum are

- Aperture dimensions are minimized, therefore the number of magnets and/or voltage and current are minimized for a given $\int B \cdot dl$ and rise-time.
- Machine vacuum is a reliable dielectric (70 kV/cm is OK [2]) and generally ‘recovers’ after a flashover, whereas a solid dielectric, outside vacuum, may not recover.

Disadvantages of putting a transmission line kicker magnet in the machine vacuum are

- The kicker magnet is costly and time consuming to construct (all parts must be appropriately cleaned and handled, a vacuum tank is required, suitable pumps are needed, the kicker magnet must be baked-out and the design must allow for the thermal expansion during bake-out, etc.).
- In the event of failure of something inside the vacuum tank the vacuum must be broken to repair the fault. Afterwards pumping, bake-out, and HV conditioning are required — a time-consuming process.

The beam coupling impedance may be an issue irrespective of whether the magnet is in vacuum. Hence a beam screen (see Section 4.5), in the kicker aperture, may be required which will require an increase in aperture dimensions in any case.

4.3.5 Kicker magnet aperture

Normally a magnetic circuit is used which contains magnetic material [2, 4]: without magnetic material the effective value of V_{ap} (Fig. 8) is greatly increased, therefore requiring more current to achieve the required field (Eq. 3). In addition, magnetic material improves field uniformity. Nickel-zinc (NiZn) ferrite, with $\mu_r \approx 1000$, is usually used [2, 9]. If the ferrite is in machine vacuum (Section 4.3.4) the out-gassing properties must be acceptable: in addition, the density of the ferrite must be such that its water absorbent characteristics are not too high and no undesirable lubricants must be used during the grinding process [9]. NiZn ferrite has the following properties:

- field rise can track current rise to within ~ 1 ns [2],
- low remnant field,
- low out-gassing rate, after bake-out.

Figure 12 shows a full-aperture C-core magnet and a window-frame magnet. The C-core magnet shown in Fig. 8 has its aperture closed by the return conductor. The C-core magnet on the Left Hand Side (LHS) of Fig. 12 has the return conductor behind the yoke: this is for beam gymnastic reasons and has the effect of increasing the effective width ($H_{ap(eff)}$) of the magnet aperture, of given dimensions, from H_{ap} to a value given by Eq. (10) [10, 11]:

$$H_{ap(eff)} \approx (H_{ap} + V_{ap}/2). \quad (10)$$

Similarly the inductance of a cell of the full-aperture kicker magnet is increased by a factor of $(H_{ap(eff)}/H_{ap})$ with respect to the return conductor closing the aperture of the kicker magnet.

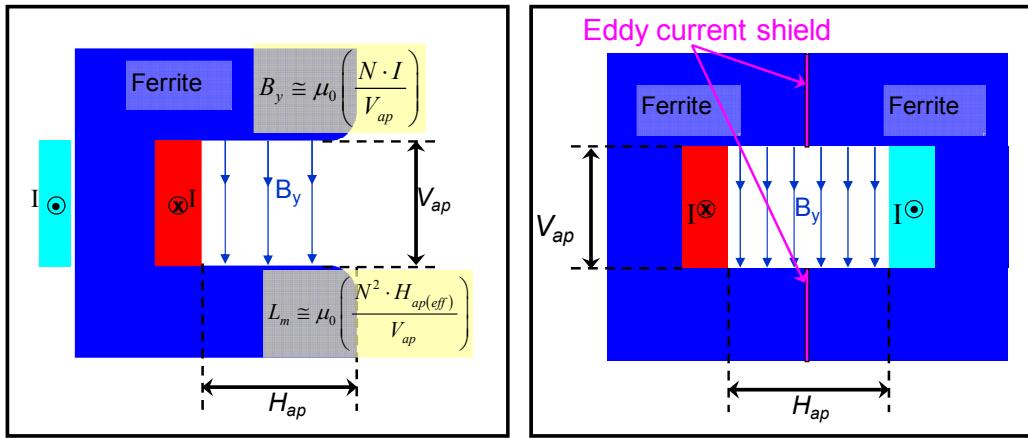


Fig. 12: Full-aperture C-core magnet (LHS) and window frame magnet (RHS)

To reduce magnet fill-time by a factor of two, FNAL and KEK use a window-frame topology (Right Hand Side (RHS) of Fig. 12): it can be considered as two symmetrical C-magnets energized independently. However, the window-frame magnet requires two generators to achieve the reduced magnet fill-time [4]. It is necessary to include eddy current ‘shields’ between the two ferrite C-cores to reduce beam coupling impedance (see Section 4.5).

C-core kicker magnets are generally used at CERN.

4.3.6 Kicker magnet termination

When space along the beam line is at a premium, a short-circuit termination has the advantage over a matched resistive termination of doubling kick (for a given system impedance) [2]: in addition, a short-circuit termination reduces the time during which the kicker magnet is exposed to high voltage. However, disadvantages include:

- fill-time of the kicker magnet is doubled: to establish full field in the magnet, the current/voltage pulse must propagate from the entrance to the exit of the magnet and then reflect back to the input;
- the kicker magnet experiences voltage of both polarities: the incident voltage pulse is inverted at the short-circuit termination (reflection coefficient is -1);

- if the DS is used to control pulse length it must be bidirectional (unidirectional dump-switch, acting as an inverse diode, is suitable for a fixed length pulse);
- beam coupling impedance can be affected [12] (owing to resonances, below magnet cut-off frequency, with kicker circuitry).

4.4 Kicker magnet design tools

Circuit simulation and finite element codes greatly assist the goal of obtaining high performance kicker systems. Simulation of circuits which include almost all known parasitic elements and nonlinearities is now possible [4].

2D and 3D finite element codes now include AC and transient analysis with eddy currents. These tools are used for kicker magnet design to predict magnetic field, cell inductance, and electric field [13–15]. In order to obtain realistic predictions for inductance and magnetic field distribution, in the aperture of the kicker magnet, it is necessary that the skin effect and proximity effect, in the HV and ‘ground’ (return) conductors (Fig. 13), be properly accounted for: this requires an AC or transient analysis to be carried out. Codes such as Opera2D and Opera3D [16] are also used to study the shape of the ferrite and conductors and thus optimize the field homogeneity [14, 15]; the current distribution is calculated by the code. The LHS of Fig. 13 shows a transmission-line kicker magnet used to deflect beam horizontally. The magnet was modelled using Opera2D to predict magnetic and electric field distribution. The predictions were post-processed and the deflection calculated; the RHS of Fig. 13 shows a plot of deflection uniformity for the optimized geometry of ferrite and return conductor [14, 15].

The penetration of the pulsed field through a beam screen (Section 4.5), in a magnet aperture, can also be accurately calculated and the frequency dependence of magnet inductance, due to eddy currents in the screen, predicted [17]. An equivalent circuit can then be fitted to the resulting predictions to account for the frequency dependence in an analog circuit simulation [18].

An AC analysis can also be used to predict the frequency dependence due to eddy currents in PFN coils [18, 19] (see Section 4.6). An equivalent circuit can be fitted to the resulting predictions to account for the frequency dependence in an analog circuit simulation.

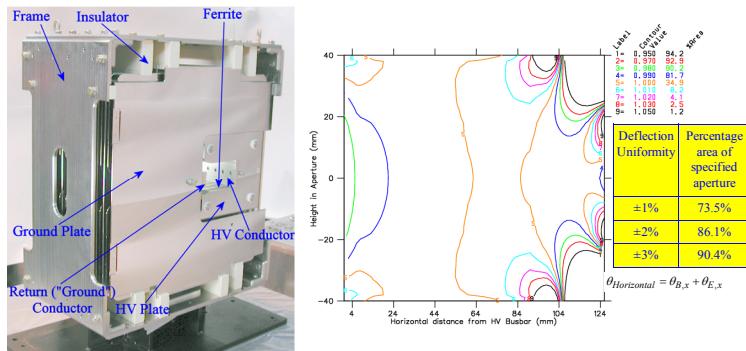


Fig. 13: Transmission-line kicker magnet and corresponding deflection uniformity plot

In addition capacitance to ground of an HV plate, in a kicker magnet, is influenced by insulators and nearby ground planes:

- ground plate,
- magnet frame,
- return conductor.

Boundary element software such as *Coulomb*, a 3D code from Integrated Engineering Software [20], can be used to accurately predict capacitance of a cell of a kicker magnet. The predictions resulting from *Coulomb* have been carefully checked against measurements and found to be in good agreement [3].

4.5 Beam-coupling impedance

High-intensity accelerators are very sensitive to longitudinal and transverse beam-coupling impedance. Kicker magnets, with their ferrite yoke, can result in considerable beam-coupling impedance [21, 22]. The beam coupling also induces heating of the ferrite yoke [23, 24] sometimes beyond the Curie temperature. In extreme cases it can affect the generator as well [25].

Some beam-coupling high-impedance resonances are attributable to the image path of the beam not being continuous, for example, a gap between the end of the kicker magnet and the vacuum tank. In this case these high-impedance resonances can be eliminated by inserting transition pieces inside the tank to electrically connect the tank flange to the kicker magnet at each end [23, 26, 27]. When several kicker magnets are installed in a common tank, the transition between magnets is carried out in a similar way [4].

Beam-coupling impedance of the ferrite yoke is addressed by providing a beam screen in the magnet aperture [12, 17, 21, 22, 24, 28–30]. The screen conducts the beam image current but at the same time must not significantly attenuate the pulsed field; thus eddy-current loops must be avoided.

The LHS of Fig. 14 shows beam-coupling impedance reduction techniques currently being implemented for the MKE kickers [21], used for extraction from the CERN SPS ring. Silver combs are printed directly on the ferrite, by serigraphy, and connected to the high-voltage plates. The serigraphy does not reduce the available aperture, and the high permittivity of the ferrite enhances the capacitive coupling between the serigraphy: the capacitive coupling provides a continuous path for the beam image current while preventing eddy-current loops. This solution is less attractive for magnets with short cells since the capacitive coupling is reduced.

The RHS of Fig. 14 shows beam-coupling impedance reduction techniques implemented for the MKI kickers, used for injection into the CERN LHC rings. A 3 m long ceramic tube is manufactured with 24 slots on its inside diameter [17]. Carefully radiused conductors are inserted into the slots: the ceramic tube acts as a support and in addition provides insulation to the ferrite and high-voltage plates. Capacitive coupling is made at one end between the beam screen conductors and an outside metallization which is connected to ground (Fig. 14): the capacitive coupling provides a continuous path for the beam image current while preventing eddy-current loops. 3D electromagnetic simulations have been carried out to optimize the geometry of the beam screen [29].

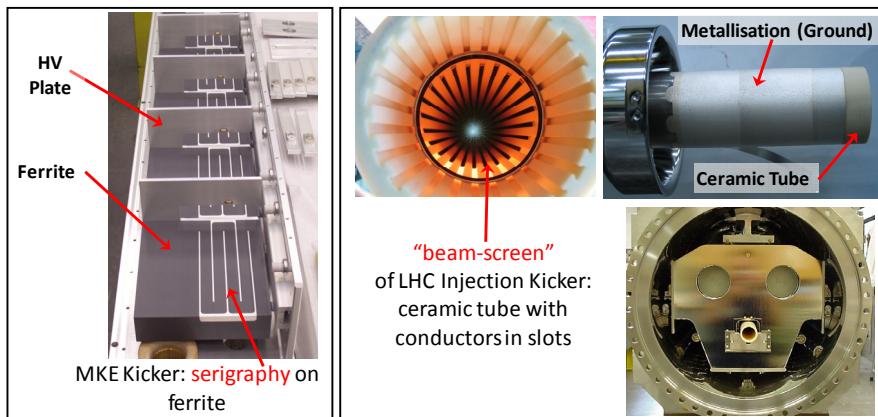


Fig. 14: Beam-coupling impedance reduction techniques for the MKE (LHS) and MKI (RHS) kickers

Figure 15 shows longitudinal beam-coupling impedance for several systems: unless mentioned otherwise the impedance is derived from measurements.

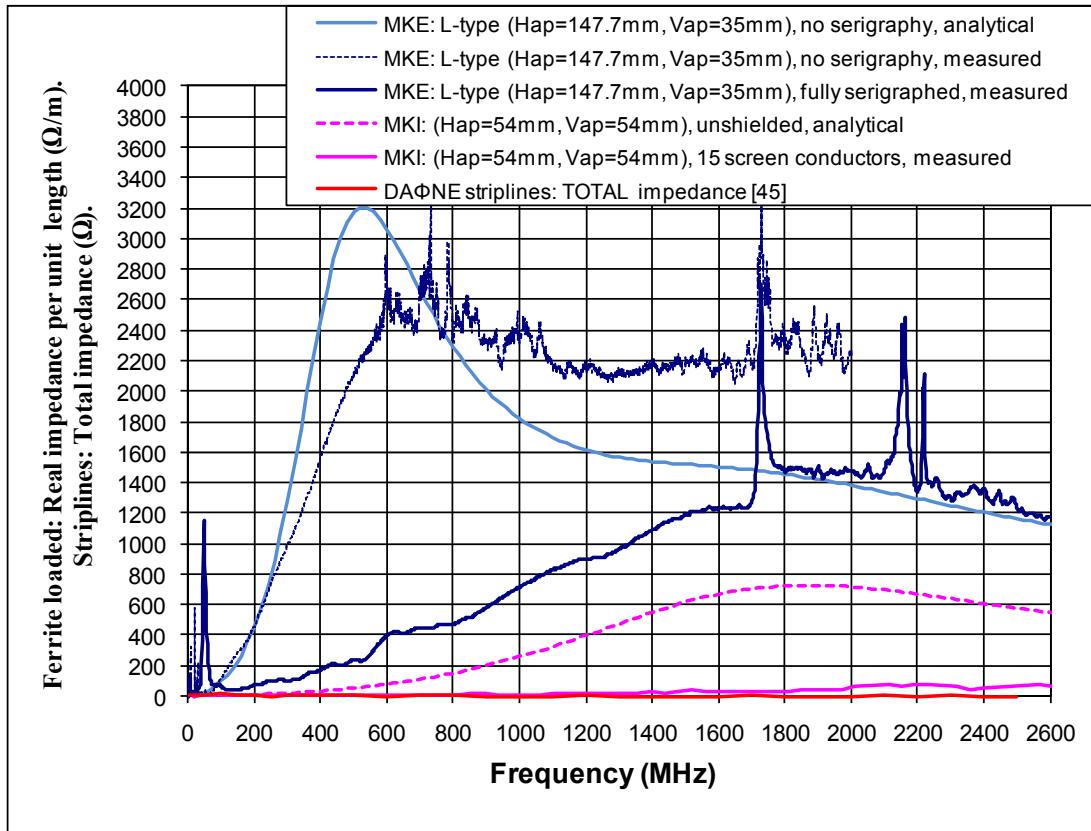


Fig. 15: Longitudinal beam-coupling impedance for several systems

Longitudinal beam-coupling impedance is significantly reduced by

- serigraphy of ferrites (painted stripes): there is negligible loss of aperture;
- beam screen conductors within the kicker magnet aperture: however this can result in approximately 15 mm loss of aperture and is thus usually not applicable as a retrofit to an existing kicker magnet;
- striplines instead of a ferrite loaded magnet: this is not feasible for obtaining a large kick in a limited length.

4.6 Pulse-forming line/network

The simplest configuration for a pulse-forming circuit is a PFL (coaxial cable) charged to twice the required pulse voltage. The PFL gives fast and ripple-free pulses, but low attenuation is essential, especially with longer pulses, to keep droop and ‘cable tail’ within specification [2]. Attenuation is adversely affected by the use of semiconductor layers to improve voltage rating [2]. Hence, for PFL voltages above 50 kV, SF₆ pressurized polyethylene (PE) tape cables are used. PFL becomes costly, bulky, and the droop becomes significant (e.g. ~1%) for pulses exceeding about 3 µs duration.

Where low droop and long pulses are required a PFN is used: a PFN is an artificial coaxial cable made of lumped elements. The SPS extraction PFNs at CERN [31, 32], which are approximately 30 years old, have 17 cells, connected in series in a single line (Fig. 16, RHS), which are individually

'adjustable': these PFNs have 'corners' (Fig. 16, LHS), therefore mutual inductance between cell inductances is not well defined and can result in ripple on the pulse. Adjusting the pulse flat top is difficult and time-consuming.

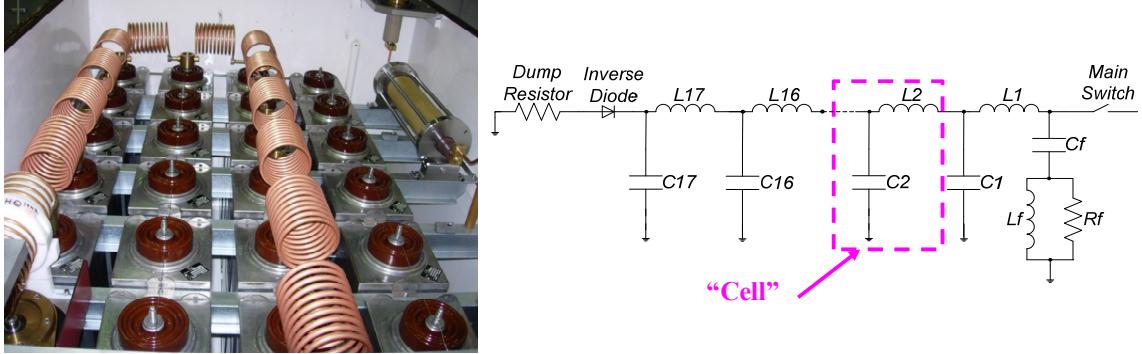


Fig. 16: The SPS extraction PFNs at CERN have 17 cells which are individually 'adjustable'

The preservation of the transverse emittance of the proton beam at injection into the LHC is crucial for luminosity performance. The transfer and injection process is important in this respect, and injection offsets are a well-known source of error [32]. To limit the beam emittance blow-up due to injection oscillations, the reflections and the flat top ripple of the field pulse must be lower than $\pm 0.5\%$, a very demanding requirement [8].

Each of the two LHC injection kicker systems has four 5Ω PFNs [33]. Each PFN consists of two lumped element delay lines, each of 10Ω , connected in parallel [33]. Each 10Ω line consists of 26 central cells plus two end-cells. A cell consists of a series inductor, a damping resistor connected in parallel, and a capacitor connected to ground (Fig. 17, LHS). The inductors are part of a single continuous coil, 4.356 m long, with 198 turns and a pitch of 22 mm [33]. The central cell inductors are made of seven turns each. The nominal MS and DS end-cell inductors have nine turns and five turns respectively, but are built with one extra turn to allow some adjustment to compensate for end-effects. The 26 central cells of the coils are not adjustable and therefore are defined with high precision: the coil conductor is a copper tube wound on a rigid fibreglass coil former. Both delay lines are mounted in a rectangular tank (Fig. 17, RHS), with mild steel walls, that is filled with insulating silicone fluid. Each line is surrounded by a 3 mm thick, Ω -shaped, aluminium shield, which has an inner radius of 140 mm. Two thyatron switches, a MS and a DS, are connected to the PFN.

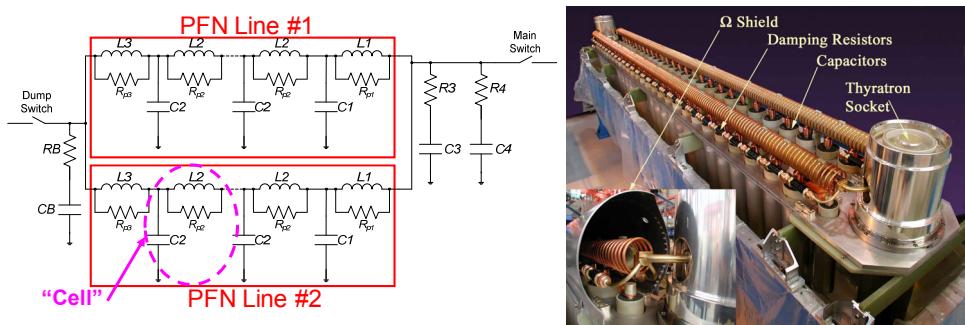


Fig. 17: Each LHC injection PFN consists of two parallel 10Ω lines with precision wound coils

Opera2D simulations have been carried out to assess the frequency dependence of inductance and resistance of a coil of the LHC injection PFN [19]. Figure 18 shows the predicted current distribution in the PFN coil at frequencies of 0.1 Hz, 200 Hz, 1 kHz and 40 kHz: the nominal DC current density is 27.3 kA/cm^2 , for a total current of 6 kA.

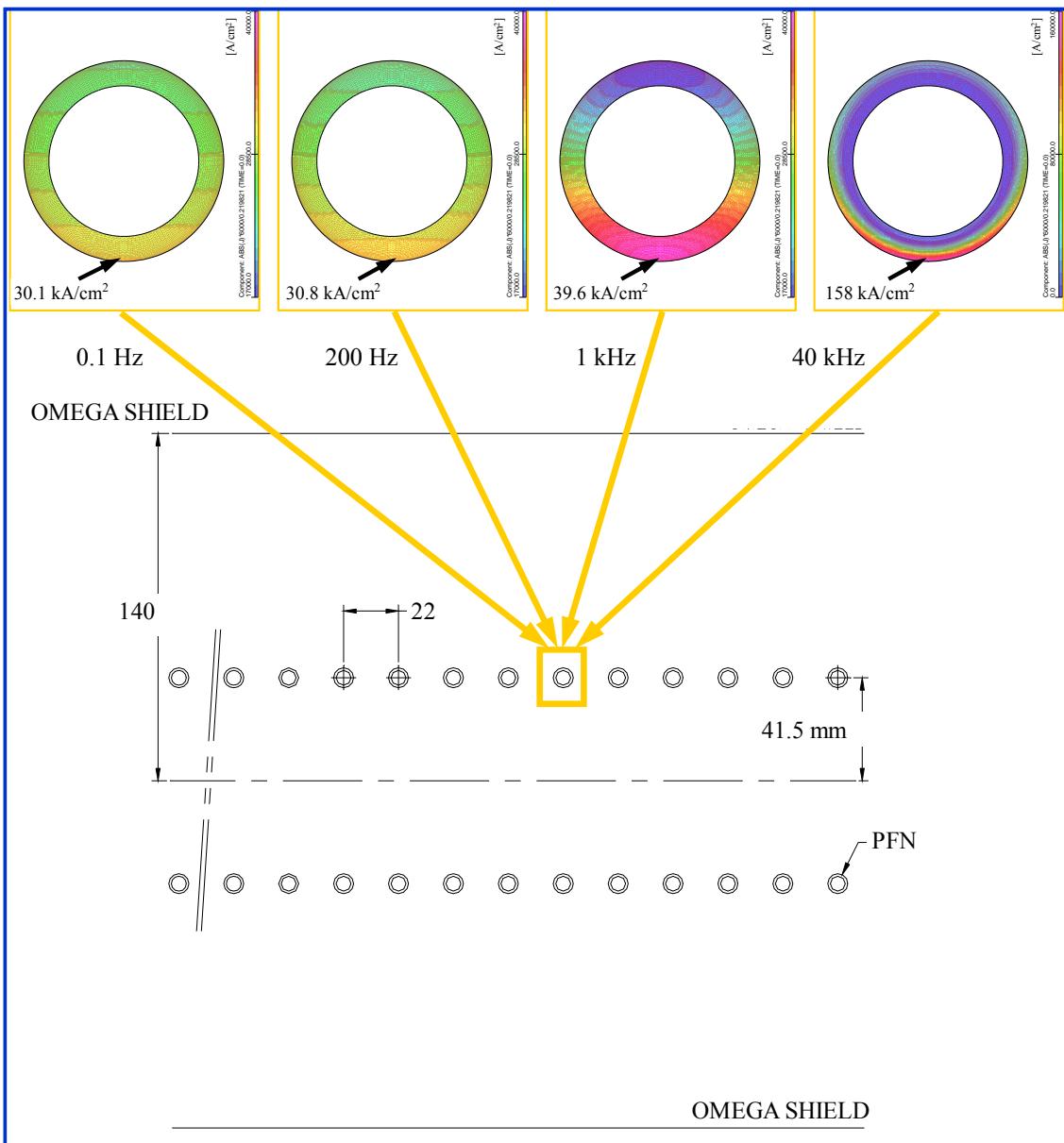


Fig. 18: Opera 2D prediction for current distribution in LHC injection PFN for a current of 6 kA

Figure 19 shows the predicted inductance of a 7-turn cell versus frequency for a mean radius of the coil of 41.5 mm. The ‘Grover limits’ refer to values calculated from equations [19]. The reduction in inductance as frequency is increased from DC to a few hundred Hertz is mainly due to screen shielding. The reaction field from the eddy currents induced in the Omega shield reduces the flux density along the axis of the coil from 0.343 T near DC to 0.315 T at a few hundred Hertz, for a current of 6 kA. As the frequency is increased beyond a few hundred Hertz the inductance decreases, mainly due to skin effect and proximity effect within the coil. Conduction losses along the coil result in droop of the pulse of approximately 0.5% in the kicker magnet. PSpice simulations, subsequently confirmed by measurements, show that conduction losses in the PFN coil can be compensated for by grading the PFN capacitor values linearly from the MS end to the DS end [33].

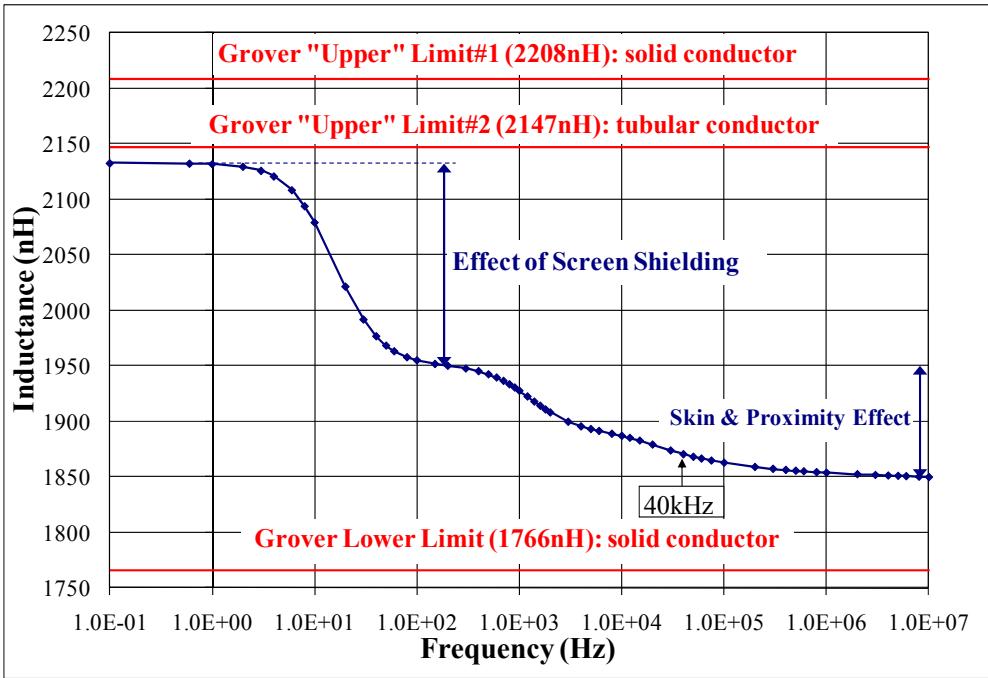


Fig. 19: Opera 2D prediction for inductance of a 7-turn PFN cell, versus frequency, for a mean radius of the coil of 41.5 mm

4.7 Power switches

4.7.1 Thyratrons

Despite the advances in high-power solid-state switches, deuterium thyratrons are still widely used as the power switch in kicker systems. The deuterium thyratron is a high peak power electrical switch which uses deuterium gas as the switching medium. The switching action is achieved by a transfer from the insulating properties of neutral gas to the conducting properties of ionized gas [34]. Voltage breakdown in the deuterium-filled gap is initiated by free charges (electrons and ions) crossing the gap under the influence of an electric field. If sufficient energy is available, gas molecules are ionized producing more free charges. The positive ions are accelerated towards the lower potential electrode and cause the release of secondary electrons. Under the right circumstances, the processes become self-sustaining and voltage breakdown occurs [34].

Thyratron commutation is achieved by introducing plasma into the grid/anode region via slots in the grid structure. The plasma is created in the cathode/grid region by a fast-rising trigger pulse applied to the grid(s), which then diffuses to the grid slots where it comes under the influence of the anode field [34].

Three-gap thyratrons can hold-off 80 kV and switch 6 kA of current with a 30 ns rise-time (10% to 90%) [$\sim 150\text{k A}/\mu\text{s}$]. However, special care must be taken with the high-power thyratron:

- Coaxial housings are used to achieve low inductance but there must be adequate insulation to the housing.
- Current flow in the connections around the thyratron should be symmetrical. An asymmetrical magnetic field can impose forces on the internal plasma preventing uniform current density which may affect performance and lifetime [34].

- The appropriate thyratron must be selected (e.g., suitable rating for peak current and average current) and properly applied for the anticipated short-circuit and fault conditions.
- The thyratron must be adequately cooled to ensure that the maximum specified envelope temperature is not be exceeded.
- Erratic turn-on (turn-on without a trigger being applied) can result in operational problems: erratic turn-on is reduced significantly by ‘fast’ (\sim ms) charging of the PFN/PFL immediately before the kicker magnet is required to be pulsed.
- The reservoir voltage of the thyratron should be optimized. The gas density in the grid/anode gap must be maximized, to minimize current rise-time and switching losses and maximize thyratron lifetime, consistent with achieving a low rate of erratic turn-on.

In order to reduce as much as possible the number of erratic turn-ons, Resonant Charging Power Supplies (RCPS) are used to charge PFNs for kicker magnets [2, 35]. The number of erratic turns-on of a thyratron switch is dependent upon several variables which include the time period for which there is a high voltage across the thyratron: thus, in order to minimize the number of erratics, it is advantageous to minimize this time period. The fast RCPS developed for the LHC injection kicker systems charges two PFNs to 54 kV in approximately 800 μ s: approximately 600 μ s later the PFNs can be discharged into the kicker magnets [36]. An RCPS can also allow the reservoir voltage to be increased, provided that the rate of erratic turn-on is still acceptable, and therefore help to extend thyratron lifetime.

To achieve significantly improved performance and lifetime a thyratron can be double pulsed to turn it on: the first pulse pre-ionizes the cathode grid space which prepares the cathode region for conducting the main current pulse. The second pulse, which should be delayed by at least 500 ns, provides a fast rising voltage to ensure precise triggering (1–2 ns) of the thyratron [34].

4.7.2 Power semiconductor switches

In some applications thyratron switches cannot be used, e.g., for the dump (abort) kickers in the LHC where no self-firing is allowed [37]. For the LHC system, acceleration of the beam to the required energy and subsequent physics may require more than 10 hours: the generators of the LHC dump kickers must be charged to the correct voltage, which must track the beam energy, and be ready to operate at any time in the LHC cycle. In this application high-power semiconductor switches are used. The semiconductor switches also allow a wide dynamic range of operation [37]. Maintenance is significantly reduced with a semiconductor switch, in comparison with a thyratron. However, high-power semiconductors have a maximum rate-of-rise of current which is significantly less than a thyratron: the current rise-time for a high-power semiconductor switch is typically \sim 1 μ s.

Semiconductors may be influenced by ionizing radiation and neutron flow. Depending on the type of radiation, the component type and its working conditions, the radiation effects can be cumulative with relatively slow deterioration of semiconductor performance and/or sudden malfunction or failure: the latter is called ‘Single Event Effects’ (SEE). The cumulative effects, which result from Total Ionizing Dose (TID) and displacement damage, are responsible for the modification of component parameters such as leakage current, bipolar transistor gain, opto-coupler efficiency, MOS threshold voltage, voltage reference value, etc. On the other hand, a SEE provokes sudden malfunctions which, in the case of HV power semiconductors, usually lead to component failure: this is known as Single Event Burnout (SEB). For a given radiation dose the SEB rate is strongly dependent on the applied voltage: there is a very steep increase of the failure rate when the applied voltage is higher than a certain percentage of the components’ rated voltage. The general recommendation for high voltage components, to maintain a reasonably low SEB rate, is to apply a maximum DC voltage of approximately 50% of the semiconductor rating. For the LHC abort kickers the power semiconductors can be affected by radiation and neutron flow from the LHC [38].

Each semiconductor switch module for the LHC abort kickers consists of ten series Gate Turn Off (GTO) thyristors (Fig. 20) that have been optimized for turn-on. The maximum forward voltage rating of each GTO is 4.5 kV and the switch module is operated over a voltage range from 2.2 kV (450 GeV) to 30 kV (7 TeV) [37]. The current switched is in the range from 1.3 kA to 18.5 kA; the maximum rate of rise of current is 18 kA/ μ s, which corresponds to approximately 1/8th of the capability of a thyratron. However, recent tests on similar GTOs, with a high-current gate drive, have demonstrated a capability of more than 32 kA/ μ s.

Ultra-fast kicker systems generally use either fast high-voltage MOSFETs [39–45] or a Fast Ionization Dynistor (FID) [45–47]; the FID is also sometimes called a Fast Ionization Device.



Fig. 20: Stack of high-power GTOs (LHS) and individual GTOs (RHS)

4.8 Ultra-fast kickers

Where sub-nanosecond jitter is required, only semiconductor switches can be used. A case in point is the tail-clipper kicker for the CLIC Test Facility 3 (CTF3) [48] at CERN: the tail-clipper must have a fast field rise-time, of 5 ns or less, to minimize uncontrolled beam loss. For this application there are eight pulse generators: each is composed of a $50\ \Omega$ PFL, a fast semiconductor MOSFET switch, $50\ \Omega$ stripline plates (no magnetic material) and a matched terminating resistor [43, 44]. The deflection of the electron beam makes use of both the electric and magnetic field in the stripline plates: in order that the effects of the electric and magnetic field do not cancel, the striplines are fed from the beam exit end of the plates (Fig. 21).

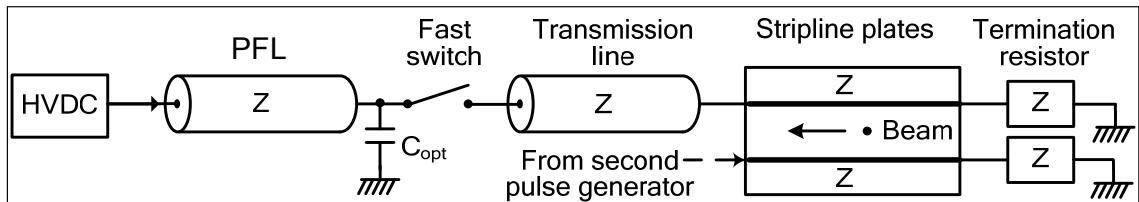


Fig. 21: Schematic of CTF3 tail-clipper kicker

The overall length of the stripline assembly is mechanically sub-divided into four sections of equal length of 380 mm; a pair of stripline plates is energized a time delay of 1.27 ns (0.38 m/c) after the previous section to minimize the overall apparent rise-time of the kick [43]. Each set of stripline plates is driven by two fast switches, one connected to a positively charged PFL and the other connected to a negatively charged PFL (Fig. 21). Figure 22 shows a measured current pulse from a MOSFET switch: the 56 A of current has a 10% to 90% rise-time of 2.5 ns. A novel design of gate driver results in an overall 3-sigma jitter of less than ± 300 ps for the modulator system [43].

A FID was considered for use for the CTF3 tail-clipper but a suitable device was not delivered in time: thus the FID is currently being evaluated.

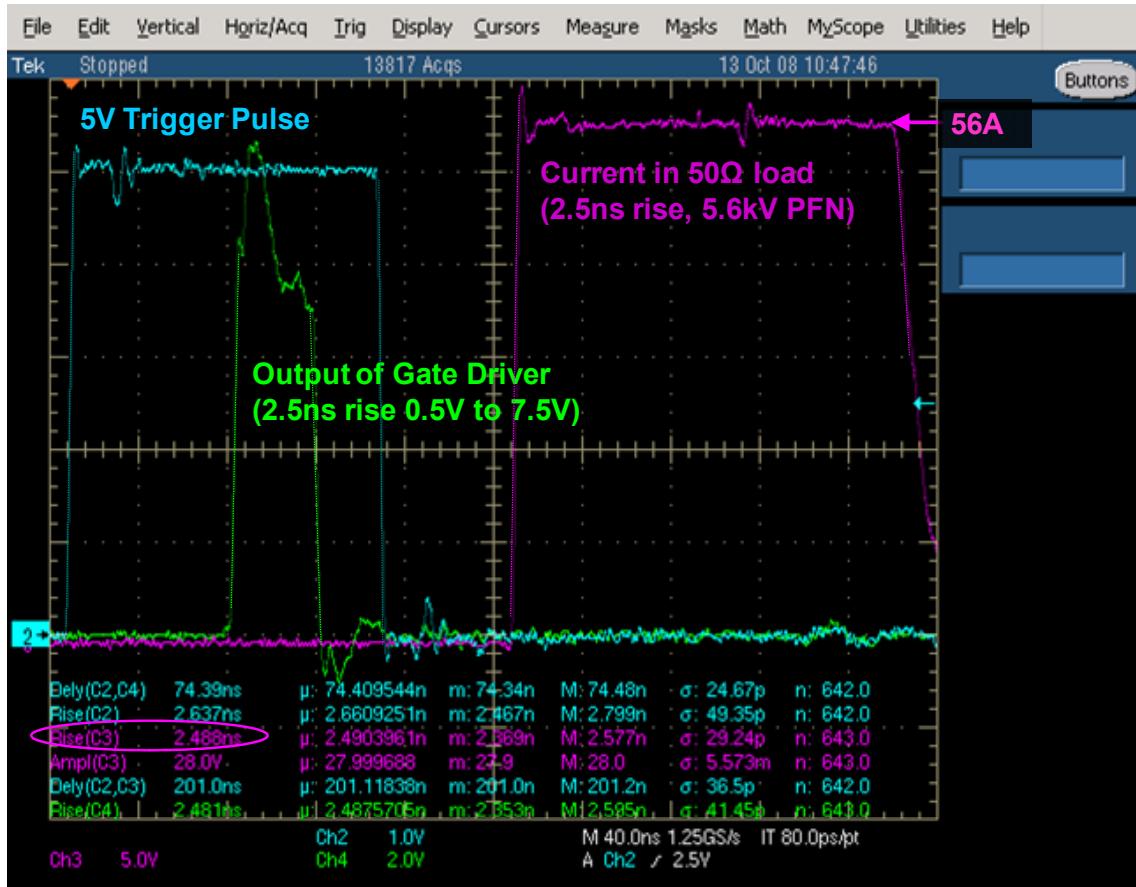


Fig. 22: Measured trigger pulse (cyan trace), gate driver output (green trace) and load current (lilac trace) for a CTF3 tail-clipper kicker system

4.9 Resistive terminators

In order that the impedance of the resistive terminator be matched to the system impedance, over a wide range of frequencies, high-power resistor disks are generally used, which are housed in a coaxial structure to minimize inductance of the terminator. The coaxial housing is normally tapered, with the maximum internal diameter of housing at the input end [49]: this design permits the resistive terminator to withstand a high pulse voltage while minimizing the parasitic inductance of the terminator.

Current distribution within the resistor discs, during the pulse, is dependent upon the resistivity of the disks and the frequency content of the pulse. As a result of proximity effect, image current

flows on the inside diameter of the coaxial housing. For cooling and insulation purposes the terminator is generally filled with oil [18]. Long-term stability of the resistance value is linked to the ageing of the resistor discs: the ageing process is affected by the oil [33]. In order to accelerate the stabilization towards a final value of resistance, the discs are pre-impregnated under vacuum at high temperature. The short-term stability is linked to the temperature coefficient of the resistor material. Each current pulse will raise the temperature of the resistor stack: thus, where high stability is required, a suitable heat-exchanger is required to maintain the oil temperature.

The resistor discs also exhibit a voltage coefficient of resistance, which is typically around $-1.5\%/\text{kV/cm}$: thus the resistive terminator changes in value during the pulse. For high-precision applications, where a well matched system is required, the voltage dependence of the terminator must be taken into account at the design stage.

References

- [1] G.H. Rees, Extraction, CAS - CERN Accelerator School: 5th General accelerator physics course, Jyväskylä, Finland, 1992, CERN 94-01 v 2.
- [2] D. Fiander, K.D. Metzmacher, P.D. Pearce, Kickers and septa at the PS complex, CERN, KAON PDS Magnet Design Workshop, Vancouver, Canada, October 3–5, 1988, pp. 71–79.
- [3] M.J. Barnes, G.D. Wait, Low power measurements on an AGS injection kicker magnet, Proc. 22nd Particle Accelerator Conference (PAC'07), Albuquerque, New Mexico, USA, June 25–29, 2007, pp. 2188–2190.
- [4] L. Ducimetière, Advances of transmission line kicker magnets, Proc. 21st Particle Accelerator Conference (PAC'05), Knoxville, USA, May 16–20, 2005, pp. 235–239.
- [5] J. N. Weaver et al., Design, analysis and measurement of very fast kicker magnets at SLAC, Proc. 13th Particle Accelerator Conference (PAC'89), Chicago, USA, March 20–23, 1989, pp. 411–413.
- [6] W. Zhang, J. Sandberg, J. Tuozzolo, R. Cassel, L. Ducimetière, C. Jensen, M.J. Barnes, G.D. Wait, J. Wang, An overview of high voltage dielectric material for travelling wave kicker magnet application, Proc. 25th International Power Modulator Conference and High Voltage Workshop, Hollywood, California, USA, June 30–July 3, 2002, pp. 674–678.
- [7] M.J. Barnes, G.D. Wait, I.M. Wilson, Comparison of field quality in lumped inductance versus transmission line kicker magnet, Proc. 4th European Particle Accelerator Conference (EPAC'94), London, UK, June 27–July 1, 1994, pp. 2547–2549.
- [8] L. Ducimetière, N. Garrel, M.J. Barnes, G.D. Wait, The LHC injection kicker magnet, Proc. 20th Particle Accelerator Conference (PAC'03), Portland, Oregon, USA, May 12–16, 2003, pp. 1162–1164.
- [9] G.H. Rees, Injection, CAS - CERN Accelerator School: 5th General accelerator physics course, Jyväskylä, Finland, 1992, CERN 94-01 v 2.
- [10] D.C. Fiander, K.D. Metzmacher, L. Sermeus, AA ejection kicker for AA Complex, CERN PS/BT/Note 87-22.
- [11] K.D. Metzmacher, L. Sermeus, AA injection kicker for AA Complex, CERN PS/BT/Note 87-5.
- [12] M.J. Barnes, F. Caspers, T. Kroyer, E. Metral, F. Roncarolo, B. Salvant, Measurement of longitudinal and transverse impedance of kicker magnets using the coaxial wire method, to be publ. in proceedings of 23rd Particle Accelerator Conference (PAC'09), Vancouver, Canada, May 4–8, 2009.
- [13] M.J. Barnes, G.D. Wait, Comparison of measured and predicted inductance per cell for a travelling wave kicker magnet, Proc. 5th European Particle Accelerator Conference (EPAC'96), Sitges, Spain, June 10–14, 1996, pp. 2588–2590.

- [14] R.B. Armenta, M.J. Barnes, E.W. Blackmore, O. Hadary, G.D. Wait, L. Ahrens, C.J. Gardner, W. Zhang, Design concept for AGS injection kicker upgrade to 2 GeV, Proc. 21st Particle Accelerator Conference (PAC'05), Tennessee, USA, May 16–20, 2005, pp. 1380–1382.
- [15] R.B. Armenta, M.J. Barnes, E.W. Blackmore, O. Hadary, D.T. Norn, G.D. Wait, Electromagnetic modeling of the AGS A10 injection kicker magnet, *IEEE Transactions on Applied Superconductivity*, Vol. 16, No. 2, June 2006, pp. 293–296.
- [16] Cobham Technical Services, 24 Bankside Kidlington, Oxford OX5 1JE, UK, www.vectorfields.com.
- [17] M.J. Barnes, F. Caspers, L. Ducimetière, N. Garrel, T. Kroyer, The beam screen for the LHC injection kicker magnets, Proc. 10th European Particle Accelerator Conference (EPAC'06), Edinburgh, Scotland, June 26–30, 2006, pp. 1508–1510.
- [18] L. Ducimetière, U. Jansson, G.H. Schröder, E.B. Vossenberg, M.J. Barnes, G.D. Wait, Design of the injection kicker magnet system for CERN's 14TeV proton collider LHC, Proc. 10th International Pulsed Power Conference, Albuquerque, New Mexico, USA, July 10–13, 1995, pp. 1406–1411.
- [19] M.J. Barnes, G.D. Wait, L. Ducimetière, Inductance calculations and measurements for the CERN LHC injection pulse forming network, Proc. 7th European Particle Accelerator Conference (EPAC'00), Vienna, Austria, June 26–30, 2000, pp. 2355–2357.
- [20] Integrated Engineering Software, Winnipeg, Manitoba, Canada, www.integratedsoft.com.
- [21] T. Kroyer, F. Caspers, E. Gaxiola, Longitudinal and transverse wire measurements for the evaluation of impedance reduction measures on the MKE extraction kickers, CERN AB-Note-2007-028.
- [22] F. Caspers, L. Ducimetière, P. Faure, E. Gaxiola, T. Kroyer, B. Versolatto, E. Vossenberg, The fast extraction kicker system in SPS LSS6, Proc. 10th European Particle Accelerator Conference (EPAC'06), Edinburgh, Scotland, June 26–30, 2006, pp. 3125–3127.
- [23] J. Uythoven, G. Arduini, T. Bohl, F. Caspers, E.H. Gaxiola, T. Kroyer, M. Timmins, L. Vos, Beam induced heating of the SPS fast pulsed magnets, Proc. 9th European Particle Accelerator Conference (EPAC'04), Lucerne, Switzerland, July 5–9, 2004, pp. 623–625.
- [24] J. Rümmler, DESY, DESY III dump system with one fast kicker, Proc. 16th Particle Accelerator Conference (PAC'95), Dallas, Texas, USA, May 1–5, 1995, pp. 1933–1935.
- [25] W. Zhang, L.A. Ahrens, J. Glenn, J. Sandberg, N. Tsoupas, Beam coupling phenomena in fast kicker systems, Proc. 19th Particle Accelerator Conference (PAC'01), Chicago, USA, June 18–22, 2001, pp. 3708–3710.
- [26] F. Caspers, A. Mostacci, H. Tsutsui, Impedance evaluation of the SPS MKE kicker with transition pieces between tank and kicker module, CERNNSL-2000-071 (AP).
- [27] H. Hahn, A. Dunbar, C.I. Pai, R.T. Sanders, N. Tsoupas, J.E. Tuorzzolo, The RHIC beam abort kicker system, Proc. 18th Particle Accelerator Conference (PAC'99), New York, USA, March 29–April 2, 1999, pp. 1100–1102.
- [28] M.J. Barnes, F. Caspers, K. Cornelis, L. Ducimetière, E. Mahner, G. Papotti, G. Rumolo, V. Senaj, E. Shaposhnikova, Measurement and analysis of SPS kicker magnet heating and outgassing with different bunch spacing, to be published in proceedings of 23rd Particle Accelerator Conference (PAC'09), Vancouver, Canada, May 4–8, 2009.
- [29] M.J. Barnes, F. Caspers, L. Ducimetière, N. Garrel, T. Kroyer, An improved beam screen for the LHC injection kickers, Proc. 22nd Particle Accelerator Conference (PAC'07), Albuquerque, New Mexico, USA, June 25–29, 2007, pp. 1574–1576.
- [30] J. Rümmler, Kicker Types for in- and ejection in the damping rings of TESLA, DESY/MIN December 2000.

- [31] G.Arduini, P.Burkel, E.Gaxiola, W.Höfle, F.Roncarolo, E.Vogel, E.Vossenberg, Performance of the CERN SPS fast extraction for the CNGS facility, Proc. 21st Particle Accelerator Conference (PAC'05), Knoxville, USA, May 16–20, 2005, pp. 1757–1759.
- [32] B. Goddard, M. Barnes, L. Ducimetière, W. Höfle, G. Kotzian, Emittance growth at LHC injection from SPS and LHC kicker ripple, Proc. 11th European Particle Accelerator Conference (EPAC'08), Genoa, Italy, June 23–27, 2008, pp. 3629–3631.
- [33] M.J. Barnes, G.D. Wait, L. Ducimetière, Low voltage measurements on nine PFNs for the LHC injection kicker systems, Proc. 8th European Particle Accelerator Conference (EPAC'02), Paris, France, June 3–8, 2002, pp. 2520–2522.
- [34] E2V Technologies, 'Hydrogen Thyratrons Preamble',
http://www.e2v.com/assets/media/files/documents/thyratrons/thyratron_preamble.pdf.
- [35] M.J. Barnes, G.D. Wait, E. Carlier, L. Ducimetière, G.H. Schröder, E.B. Vossenberg, A fast 60 kV resonant charging power supply for the LHC inflectors, Proc. 17th Particle Accelerator Conference (PAC'97), Vancouver, Canada, May 12–16, 1997, pp. 1325–1327.
- [36] M.J. Barnes, G.D. Wait, L. Ducimetière, G.H. Schröder, E.B. Vossenberg, Measurements on a fast 66 kV resonant charging power supply for the LHC inflectors, Proc. 6th European Particle Accelerator Conference (EPAC'98), Stockholm, Sweden, June 22–26, 1998, pp. 2278–2280.
- [37] J. Bonthond, J.H. Dieperink, L. Ducimetière, U. Jansson, E. Vossenberg, Dual branch high voltage pulse generator for the beam extraction of the large hadron collider, Proc. 2002 Power Modulator Symposium, Hollywood, USA, June 30–July 3, 2002, pp. 114–117.
- [38] V. Senaj, N. Voumard, M.J. Barnes, L. Ducimetière, Optically isolated circuit for failure detection of a switch in an HV series connected stack, to be publ. in proceedings of 17th IEEE International Pulsed Power Conference, Washington DC, USA, 29 June – 2 July 2009.
- [39] M.J. Barnes, G.D. Wait, J. Dilling, J.V. Vaz, L. Blomeley, O. Hadary, M.J. Smith, A high frequency MOSFET driver for the TITAN facility at TRIUMF, Proc. 15th IEEE International Pulsed Power Conference, Monterey, USA, June 13–17, 2005, pp. 178–181.
- [40] M.J. Barnes, G.D. Wait, A 25 kV, 75 kHz, kicker for measurement of muon lifetime, *IEEE Transactions on Plasma Science*, Vol. 32, No. 5, October 2004, pp. 1932–1944.
- [41] F. Arntz, M. Gaudreau, M. Kempkes, A. Krasnykh, A. Kardo-Sysoev, A kicker driver for the international linear collider, Proc. 22nd Particle Accelerator Conference (PAC'07), Albuquerque, New Mexico, USA, 25–29 June 2007, pp. 2972–2974.
- [42] F. Arntz, J. Casey, M. Kempkes, N. Butler, M. Gaudreau, Solid-state high voltage pulse power in the 10-100 nanosecond regime, Proc. 10th European Particle Accelerator Conference (EPAC'06), Edinburgh, Scotland, June 26–30, 2006, pp. 3134–3136.
- [43] M.J. Barnes, T. Fowler, G. Ravida, Design, Testing and operation of the modulator for the CTF3 tail clipper kicker, to be publ. in proceedings of 23rd Particle Accelerator Conference (PAC'09), Vancouver, Canada, May 4–8, 2009.
- [44] M.J. Barnes, T. Fowler, G. Ravida, A. Ueda, Design of the modulator for the CTF3 tail clipper kicker, Proc. 22nd Particle Accelerator Conference (PAC'07), Albuquerque, New Mexico, USA, June 25–29, 2007, pp. 2185–2187.
- [45] A. Wolski, Critical R&D issues for the ILC damping rings and new test facilities,
<http://www.cockcroft.ac.uk/papers/Cockcroft-07-14.pdf>.
- [46] T. Naito et al., Development of strip-line kicker system for ILC damping ring, Proc. 22nd Particle Accelerator Conference (PAC'07), Albuquerque, New Mexico, USA, June 25–29, 2007, pp. 2772–2774.
- [47] D. Alesini, S. Guiducci, F. Marcellini, P. Raimondi, Fast injection kickers for Daphne collider and ILC damping rings, DAPHNE Technical Note, INFN - LNF, Accelerator Division. Note I-17, June 6, 2006.
- [48] G. Geschonke, A. Ghigo, CTF3 Design Report, CERN/PS 2002-008 (RF).

- [49] M.J. Barnes, G.D. Wait, Optimization of speed-up network component values for the 30Ω resistively terminated prototype kicker magnet, Proc. 15th Particle Accelerator Conference (PAC'93), Washington DC, USA, May 17–20, 1993, pp. 1330–1332.

Bibliography

- Eds. O. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojoic, J. Poole, P. Proudlock, LHC Design Report, volume I, The LHC main ring, chapter 16, Injection system, <http://ab-div.web.cern.ch/ab-div/Publications/LHC-DesignReport.html>, CERN, Geneva, 2004, CERN-2004-003.
- Eds. O. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojoic, J. Poole, P. Proudlock, LHC Design Report, volume I, The LHC main ring, chapter 17, Beam dumping system, <http://ab-div.web.cern.ch/ab-div/Publications/LHC-DesignReport.html>, CERN, Geneva, 2004, CERN-2004-003.
- Eds. M. Benedikt, P. Collier, V. Mertens, J. Poole, K. Schindl, LHC Design Report, volume III, The LHC injector chain, <http://ab-div.web.cern.ch/ab-div/Publications/LHC-DesignReport.html>, CERN, Geneva, 2004, CERN-2004-003.
- P. Smith, *Transient Electronics: Pulsed Circuit Technology*, John Wiley & Sons, ISBN: 047197773X, 2002.
- L. Ducimetière, D. Fiander, Commutation losses of a multigap high voltage thyratron, Proc. Power Modulator Symposium, San Diego, USA, June 26–28, 1990, pp. 248–253.
- G.D. Wait, M.J. Barnes, K.D. Metzmacher, L. Sermeus, The application of saturating inductors for improving the performance of the CERN PS Kicker Systems, Proc. 17th Particle Accelerator Conference (PAC'97), Vancouver, Canada, May 12–16, 1997, pp. 1328–1330.
- K.D. Metzmacher, L. Sermeus, The PSB injection and recombination kicker systems for LHC, CERNPS/CA/Note 2000-004.
- K.D. Metzmacher, L. Sermeus, The PS injection kicker KFA45 performance for LHC, CERNPS/PO/Note 2002-015 (Tech.).
- L.I. Donley, J.C. Dooling, G.E. McMichael, Using ferrite as a fast switch for improving rise time of IPNS extraction kicker, Proc. 20th Particle Accelerator Conference (PAC'03), Portland, Oregon, USA, May 12–16, 2003, pp. 1159–1161.
- W. Zhang et al., Design, development, and construction of SNS extraction fast kicker system, Proc. 20th Particle Accelerator Conference (PAC'03), Portland, Oregon, USA, May 12–16, 2003, pp. 550–552.
- G. Nassibian, Travelling Wave kicker magnets with sharp rise and less overshoot, Proc. 8th Particle Accelerator Conference (PAC'79), *IEEE Transactions on Nuclear Science*, Vol. NS-26, No. 3, pp. 4018–4020.
- M.J. Barnes, G.D. Wait, R. Barnes, Estimation of negative mutual coupling between adjacent cells of a magnetic kicker, TRIUMF Design Note TRI-DN-89-K87.
- C. Jensen, B. Hanna, R. Reilly, A fast injection kicker magnet for the Tevatron, Proc. 19th Particle Accelerator Conference (PAC'01), Chicago, USA, June 18–22, 2001, pp. 3720–3722.
- T. Kawakubo, E. Nakamura, S. Murasugi, Low beam-coupling impedance kicker magnet system generating a high magnetic field with a fast rise time, Proc. 7th European Particle Accelerator Conference (EPAC'00), Vienna, Austria, June 26–30, 2000, pp. 2246–2248.
- P.E. Faugeras, E. Frick, C.G. Harrison, H. Kuhn, V. Rodel, G.H. Schröder, J.P. Zanasco, The SPS fast pulsed magnet systems, Proc. 12th Modulator Symposium, New York, USA, Feb. 4–5, 1976 (CERN-SPS-BT-76-1).
- J. Wang, Design and test results of kicker units for the positron accumulator ring at the APS, Proc. 16th Particle Accelerator Conference (PAC'95), Dallas, Texas, USA, May 1–5, 1995, pp. 1245–1247.

- M. Giovannozzi et al., Design and tests of a low-loss multi-turn ejection for the CERN PS, 39th ICFA Advanced Beam Dynamics Workshop: High Intensity High Brightness Hadron Beams (HB2006), Tsukuba, Japan, May 29 – June 2, 2006, pp. 192–196.
- M.J. Barnes, E.W. Blackmore, G.D. Wait, J. Lemire-Elmore (UBC), B. Rablah, G. Leyh, M. Nguyen, C. Pappas, Analysis of high power IGBT short circuit failures, *IEEE Transactions on Plasma Science*, Vol. 33, No. 4, August 2005, pp. 1252–1261.
- T.S. Mattison, R.L. Cassel, A.R. Donaldson, G. Gross, Fast and reliable kicker magnets for the SLC damping rings, Proc. 16th Particle Accelerator Conference (PAC'95), Dallas, Texas, USA, May 1–5, 1995, pp. 1915–1917.
- D. Qunell, C. Jensen, D. Tinsley, Kicker system for 8 GeV proton injection, Proc. 17th Particle Accelerator Conference (PAC'97), Vancouver, Canada, May 12–16, 1997, pp. 1287–1289.
- A. Ueda, T. Ushiku, T. Mitsuhashi, Construction of travelling wave kicker magnet and pulse power supply for the KEK-photon factory storage ring, Proc. 19th Particle Accelerator Conference (PAC'01), Chicago, USA, June 18–22, 2001, pp. 4050–4052.
- H. Hahn, N. Tsoupas, J.E. Tuozzolo, The RHIC injection kicker, Proc. 17th Particle Accelerator Conference (PAC'97), Vancouver, Canada, May 12–16, 1997, pp. 213–215.
- H. Hahn, W. Fischer, V.I. Ptitsyn, and J.E. Tuozzolo, All-ferrite RHIC injection kicker, Proc. 19th Particle Accelerator Conference (PAC'01), Chicago, USA, June 18–22, 2001, pp. 3705–3707.
- H. Hahn, W. Fischer, Y.K. Semertzidis, D.S. Warburton, Up-graded RHIC injection kicker system, Proc. 20th Particle Accelerator Conference (PAC'03), Portland, Oregon, USA, May 12–16, 2003, pp. 1625–1627.
- C.C. Jensen, G.E. Krafczyk, NuMi Proton Kicker Extraction System, Proc. 21st Particle Accelerator Conference (PAC'05), Knoxville, Tennessee, USA, May 16–20, 2005, pp. 692–694.
- M.J. Barnes, Optimize SPICE models to accurately simulate frequency-dependent impedances, *Personal Engineering & Instrumentation News*, December 1996.
- M.J. Barnes, G.D. Wait, Application of ELEKTRA to the prediction of inductance distribution for a travelling wave kicker magnet, Vector Fields North American User Group Meeting, San Jose, California, October 17, 1996.
- M.J. Barnes, G.D. Wait, Solid state switch alternatives for klystron modulators, Proc. 1995 Modulator/Klystron Workshop, Stanford, California, USA, October 9–11, 1995.
- T. Oki, The bridged-T network lumped kicker: A novel fast magnetic kicker system for a compact synchrotron, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, Vol. 607, Issue 3, 21 August 2009, pp. 489–497.
- W. Jiang, Fast High-Voltage Switching Using Stacked MOSFETs, *IEEE Transactions on Dielectrics and Electrical Insulation*, Vol. 14, Issue 4, Aug. 2007, pp. 947–950.

Injection and extraction magnets: septa

M.J. Barnes, J. Borburgh, B. Goddard, M. Hourican
CERN, Geneva, Switzerland

Abstract

An accelerator has limited dynamic range: a chain of accelerators is required to reach high energy. A combination of septa and kicker magnets is frequently used to inject and extract beam from each stage. The kicker magnets typically produce rectangular field pulses with fast rise- and/or fall-times, however the field strength is relatively low. To compensate for their relatively low field strength, the kicker magnets are generally combined with electromagnetic septa. The septa provide relatively strong field strength but are either DC or slow pulsed. This paper discusses injection and extraction systems with particular emphasis on the hardware required for the septa.

1 Introduction

An accelerator has limited dynamic range: a chain of accelerators is required to reach high energy. Thus beam transfer into (injection) and out of (extraction) an accelerator is required. The design of the injection and extraction systems aims to achieve the following:

- minimize beam loss,
- place the newly injected or extracted particles onto the correct trajectory, with the correct phase space parameters.

A combination of septa and kicker magnets is frequently used for injection and extraction. Septa can be electrostatic or magnetic: they provide slower field rise- and fall-times, but stronger field than kicker magnets. Some septa are designed to be operated with DC. Kicker magnets provide fast field rise- and fall-times, but relatively weak fields.

This paper of the CERN Accelerator School discusses the processes of injection and extraction as well as the hardware associated with the septa. The hardware associated with the kicker magnets is covered in the paper *Injection and extraction magnets: kicker magnets*.

In general, a septum (plural: septa) is a partition that separates two cavities or spaces. In a particle accelerator a septum is a device which separates two field regions. Important features of septa are an ideally homogeneous (electric or magnetic) field in one region, for deflecting beam, and a low fringe field (ideally zero magnetic and electric field) next to the septum so as not to affect the circulating beam. Hence a septum provides a space separation of circulating and injected/extracted beam. By contrast a kicker magnet provides time selection (separation) of beam to be injected/extracted.

2 Injection

Injection is the process of transferring a particle beam into a circular accelerator or accumulator ring, at the appropriate time, while minimizing beam loss and placing the injected particles onto the correct trajectory, with the correct phase space parameters. Injection methods include

- single-turn (fast) hadron injection,

- multi-turn hadron injection,
- charge-exchange H⁻ injection,
- lepton injection.

2.1 Single-turn (fast) injection

Figure 1 shows an example of fast single-turn injection in one plane. The injected beam passes through the homogeneous field region (gap) of the septum: circulating beam is in the field-free region of the septum (i.e., space separation of injected and circulating beam). The septum deflects the injected beam onto the closed orbit at the centre of the kicker magnet; the kicker magnet compensates the remaining angle. The septum and kicker are either side of a quadrupole (defocusing in the injection plane) which provides some of the required deflection and minimizes the required strength of the kicker magnet.

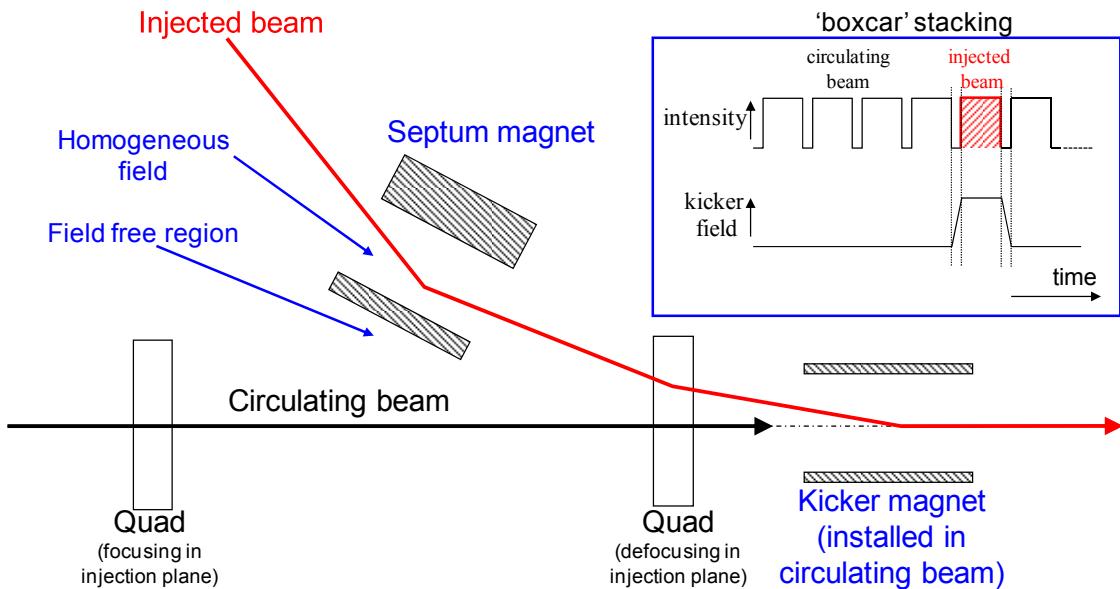


Fig. 1: Fast single-turn injection in one plane

The injected beam passes through the high field region of the septum once only and thus the degree of field homogeneity is not as critical as for a dipole magnet installed in an accelerator ring. However, the circulating beam passes through the ‘field-free’ region many times, so that achieving a very low level of stray field in the ‘field-free region’ is important.

The kicker magnet is installed in the accelerator and hence the circulating beam is in the aperture of the kicker. Thus the kicker field must rise from zero to full field in the time interval between the circulating beam and the start of the injected beam (Fig. 1, top right) and fall from full field to zero field in the time interval between the end of the injected beam and the subsequent circulating beam (Fig. 1, top right). The kicker magnet is discussed in more detail in the proceedings of this CERN Accelerator School, in the paper *Injection and extraction magnets: kicker magnets*.

Figure 2 shows an example of fast single-turn injection in two planes: a Lambertson septum (see Section 5.2.4) is used for a two-plane injection scheme. The injected beam passes through the homogeneous field region of the septum: circulating beam is in the field-free region of the septum. In the example shown in Fig. 2 the septum deflects the beam horizontally to the centre of the kicker magnet; the kicker magnet deflects the beam vertically onto the closed orbit. The septum and kicker

are either side of an F-quadrupole (horizontally focusing and vertically defocusing) to minimize the required strength of the kicker magnet.

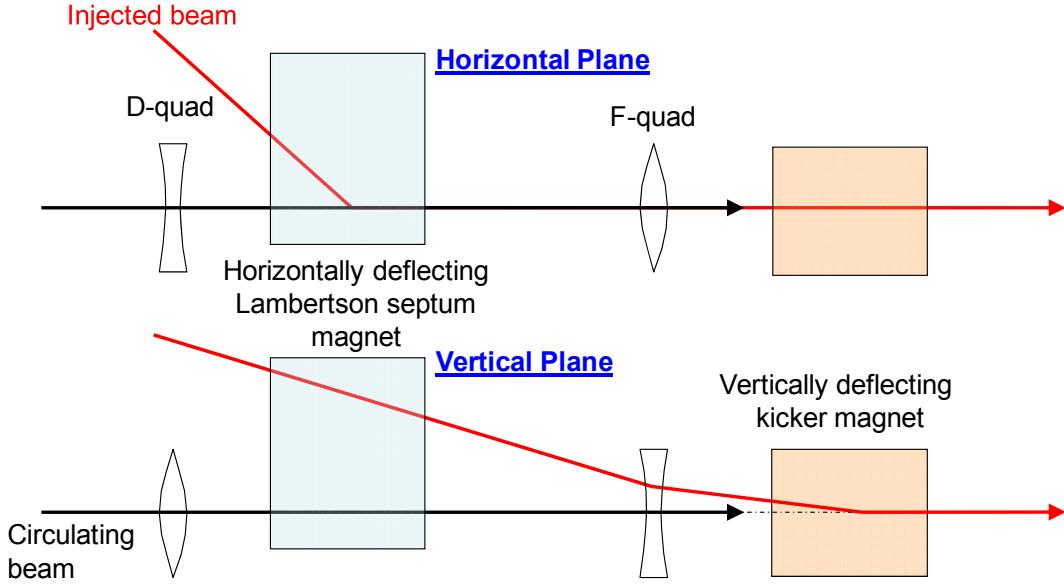


Fig. 2: Fast single-turn injection in two planes

2.2 Multi-turn hadron injection

A simple multi-turn injection employs a programmed orbit bumper and a septum. The orbit bumper usually bumps the beam in the horizontal plane because the horizontal acceptance is larger than the vertical acceptance in a conventional accelerator ring [1]. For hadrons the beam density at injection can be limited either by space charge effects or by the injector capacity. If the charge density cannot be increased, the horizontal phase space can sometimes be filled to increase the overall injected intensity: however, this requires the condition that the acceptance of the receiving machine be larger than the delivered beam emittance [2].

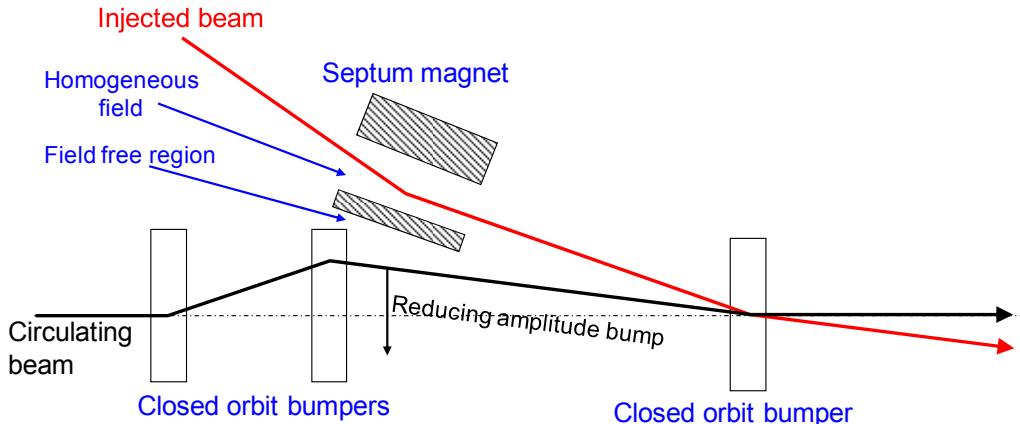


Fig. 3: Multi-turn hadron injection

Figure 3 shows an example of multi-turn hadron injection: no kicker magnet is required. The orbit bump is reduced with time so that the early beam occupies the central region of the horizontal acceptance and the later beam the periphery of the acceptance: this technique is known as phase space painting. At the end of the injection the beam bump is reduced to zero.

2.3 Charge-exchange H⁻ injection

Multi-turn injection is essential to accumulate high intensity. Disadvantages inherent in using an injection septum include [2]

- septum thickness of several millimetres reduces aperture,
- beam losses resulting from circulating beam hitting the septum render it radioactive,
- number of injected turns limited to 10–20.

Charge-exchange injection provides an elegant alternative method of injection. A uniform transverse phase space density is painted by modifying a closed orbit bump and steering injected beam [1, 2] (Fig. 4 and Fig. 5). The conversion from H⁻ ion to p⁺ means that the protons can be accumulated into already-occupied phase space, which allows very high densities to be achieved with relatively low losses.

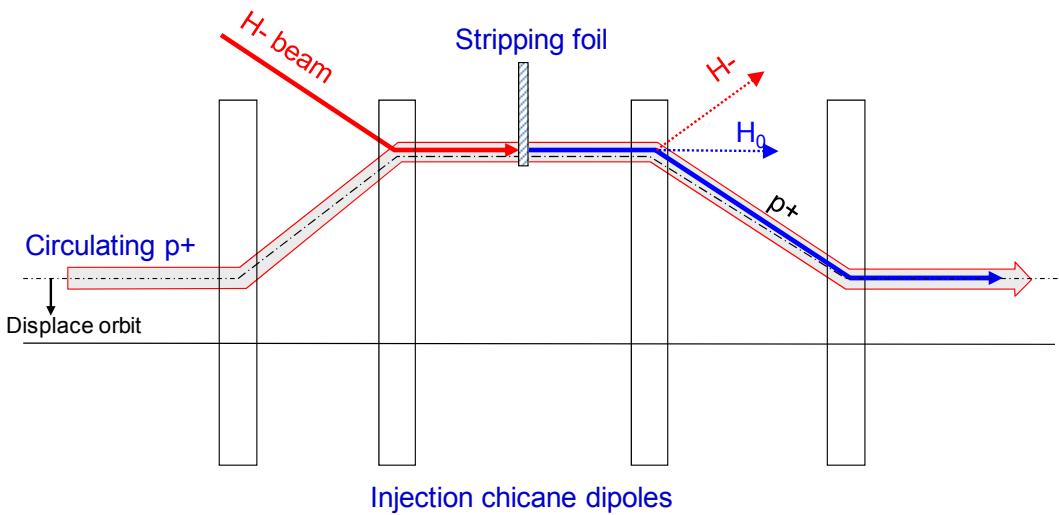


Fig. 4: Charge exchange – start of injection process

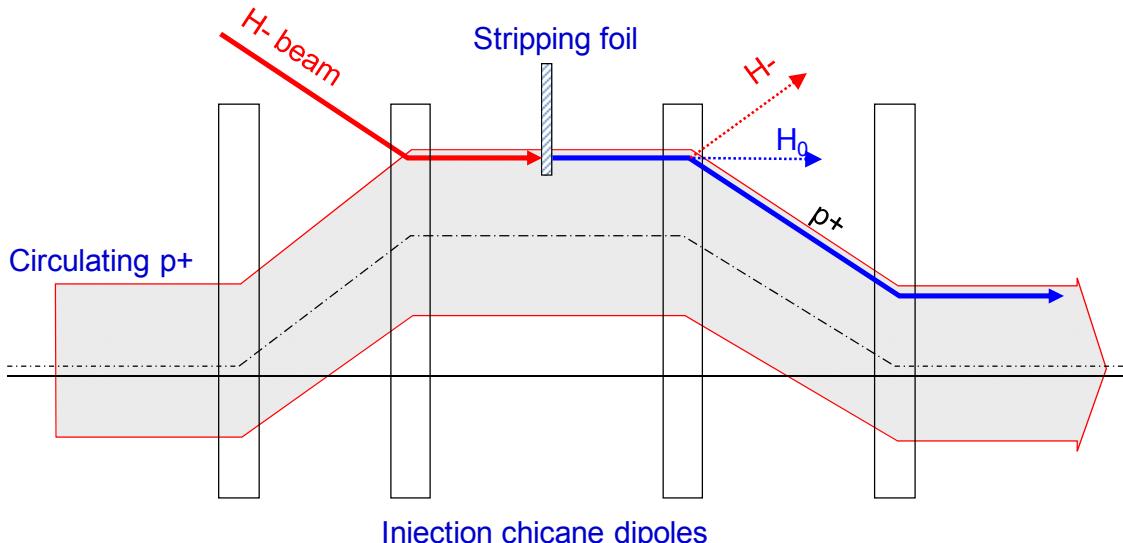


Fig. 5: Charge exchange – end of injection process

2.4 Lepton injection

Single-turn injection can be used as for hadrons; however, the lepton motion is strongly damped, which is not the case for proton or ion injection [2]. For lepton injection beam is injected at an angle with respect to the closed orbit (Fig. 6) and the injected beam performs damped betatron oscillations about the closed orbit [2].

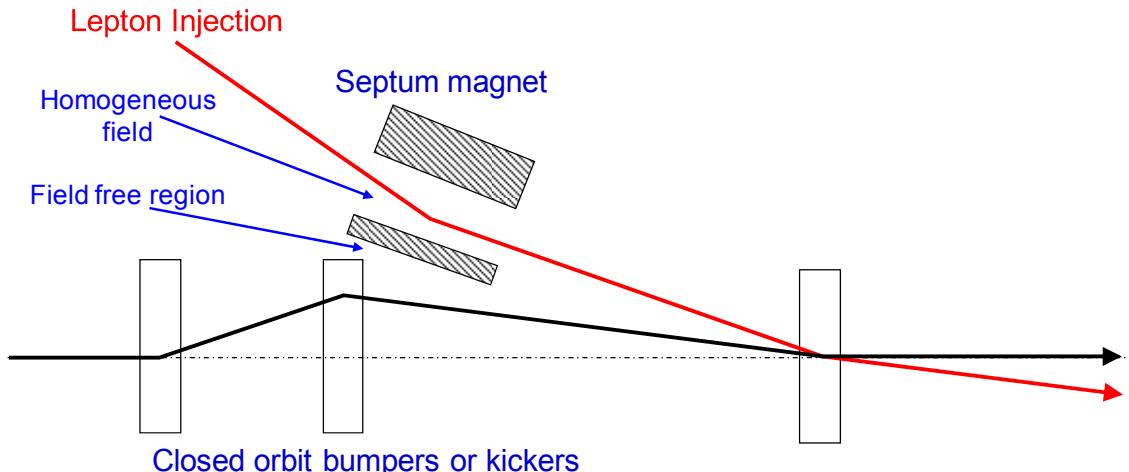


Fig. 6: Damped betatron lepton injection

3 Extraction

Extraction is the process of removing a particle beam from an accelerator to a transfer line or a beam dump, at the appropriate time, while minimizing beam loss and placing the extracted particles onto the correct trajectory, with the correct phase space parameters. Extraction methods include

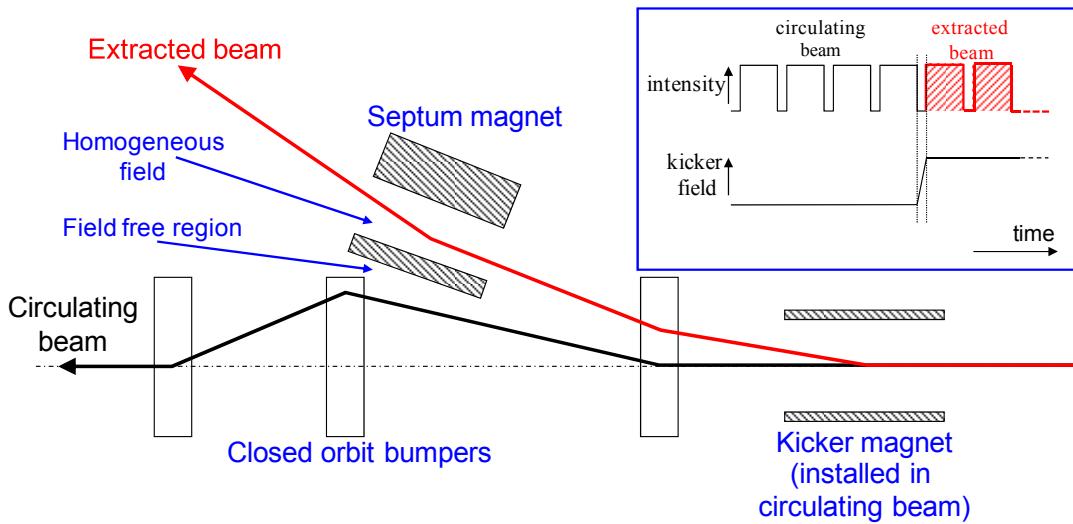
- single-turn (fast) extraction,
- non-resonant multi-turn extraction,
- resonant multi-turn (slow) extraction.

Extraction usually occurs at higher energy than injection, hence stronger elements (larger $\int B \cdot dl$) are required. At high energies many kicker and septum modules may be needed. To reduce the required strength of the kicker magnet, the beam can be moved near to a septum by a closed orbit bump.

3.1 Single-turn (fast) extraction

Figure 7 shows an example of fast single-turn extraction in one plane. The kicker magnet deflects the entire beam into the septum gap in a single turn (kicker magnet provides time selection [separation] of beam to be extracted). The septum deflects the entire kicked beam into the transfer line (septum provides space separation of circulating and extracted beam). The extracted beam passes through the homogeneous field region of the septum: the circulating beam, prior to extraction, is in the field-free region of the septum. A closed orbit bump can be applied to bring the circulating beam near to the septum to minimize the required strength of the kicker magnet.

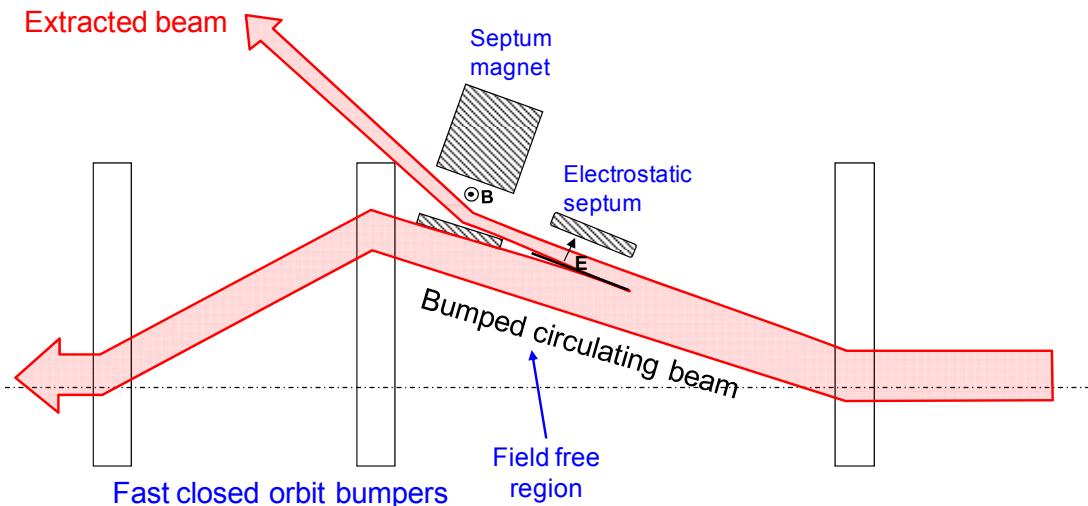
As for injection, the extracted beam passes through the homogeneous field region of the septum once only and thus the degree of homogeneity is not as critical as for a dipole magnet installed in an accelerator ring. However, the circulating beam passes through the ‘field-free’ region many times, so that achieving a very low level of stray field in the ‘field-free region’ is important.

**Fig. 7:** Fast single-turn extraction in one plane

The kicker magnet is installed in the accelerator and hence the circulating beam is in the aperture of the kicker. Thus the kicker field must rise from zero to full field in a beam-free time interval deliberately created in the circulating beam (Fig. 7, top right). The entire beam is generally extracted and hence fast fall-time is typically not required: however, sometimes, bunch-by-bunch transfers are made and then the field of the kicker magnets must have fast rise- and fall-times [3]. The kicker magnet is discussed in more detail in the proceedings of this CAS, in the paper *Injection and extraction magnets: kicker magnets*.

3.2 Non-resonant multi-turn extraction

Some filling schemes require a beam to be injected in several turns to a larger machine. Non-resonant multi-turn extraction (over a few turns) was used, for example, for filling the SPS, at CERN, with beam extracted from the PS, with high-intensity proton beams ($>2.5 \cdot 10^{13}$ protons) [2]. The process is shown schematically in Fig. 8: a fast bumper deflects the beam onto the septum and part of the beam is ‘shaved’ off at each turn. The beam is extracted in a few turns, with the machine tune rotating the beam. This is intrinsically a high-loss process — and hence a thin septum is essential.

**Fig. 8:** Non-resonant multi-turn extraction

3.3 Resonant multi-turn (slow) extraction

Resonant multi-turn extraction is generally used for delivering beam to experiments; the extraction process can be spread over a time-interval from milliseconds to hours. Non-linear fields (slow bumpers) excite betatron resonances which drive the beam slowly across the septum (Fig. 9): this is often a third-order resonance [4].

Sextupole fields distort the circular normalized-phase-space particle trajectories: a stable area is defined, delimited by unstable fixed points [2]. Sextupole families are arranged to produce suitable phase space orientation of the stable triangle at a thin electrostatic septum. The stable area can be reduced by increasing the sextupole strength or, more easily, by approaching machine tune Q_h to a resonant $1/3$ integer tune [2].

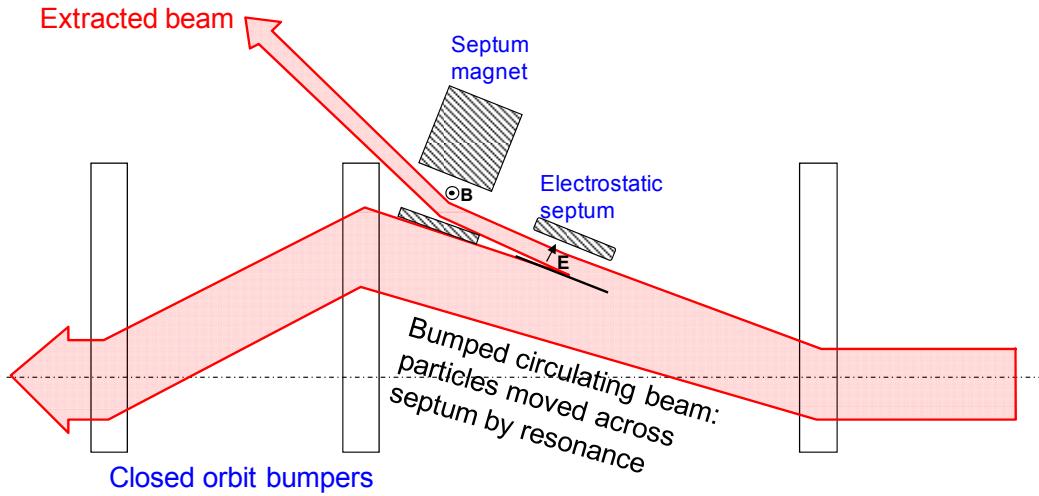


Fig. 9: Resonant multi-turn (slow) extraction

4 Deflection of beam

The force used to direct a charged particle beam is known as the Lorentz force. The Lorentz force is given by Eq. (1):

$$F = q \left[E + (v \times B) \right], \quad (1)$$

where

- F is the force (N),
- E is the electric field (V/m),
- B is the magnetic field (T),
- q is the electric charge of the particle (C),
- v is the instantaneous velocity of the particle (m/s),
- \times is the vector cross product.

The deflection of a charged particle beam in a magnetic field is given by Eq. (2) [5]:

$$\theta_{B,x} = \left[\frac{0.2998}{p} \right] \cdot \int_{z_0}^{z_1} |B_y| dz = \left[\frac{0.2998 \cdot l_{eff}}{p} \right] \cdot |B_y|, \quad (2)$$

where

- B_y is the magnetic flux density in the y -direction (T),
 p is the beam momentum (GeV/c),
 l_{eff} is the effective length of the magnet [usually different from the mechanical length, due to fringe fields at the ends of the magnet] (m), and
 $\theta_{B,x}$ is the deflection angle, in the x -direction, due to magnetic field B_y (radians).

The deflection of a charged particle beam in an electric field is given by Eq. (3) [5]:

$$\theta_{E,x} = \tan^{-1} \left[\frac{1}{(p \cdot 10^9) \cdot \beta} \cdot \int_{z_0}^{z_1} |E_x| dz \right] = \tan^{-1} \left[\frac{|E_x| \cdot l_{\text{eff}}}{(p \cdot 10^9) \cdot \beta} \right] = \tan^{-1} \left[\frac{|V| \cdot l_{\text{eff}}}{d \cdot (p \cdot 10^9) \cdot \beta} \right], \quad (3)$$

where

- V is the potential difference between plates (V),
 d is the separation of the plates (m),
 E_x is the electric field in the x -direction (V/m),
 β is a unit-less quantity that specifies the fraction of the speed of light at which the particles travel (v/c), and
 $\theta_{E,x}$ is the deflection angle, in the x -direction, due to electric field E_x (radians).

5 Septa

Two main types of septa exist, namely electrostatic septa and magnetic septa:

- an electrostatic septum is a DC electrostatic device with very thin (typically $\leq 100 \mu\text{m}$) separation between the zero field and high field regions,
- a magnetic septum is either a pulsed or DC dipole magnet with a thin (typically 2 mm to 20 mm) separation between the zero-field and high-field regions.

One of the significant challenges associated with the design of a septum is to achieve a low leakage field next to the septum, to avoid affecting the circulating beam, and the required field homogeneity in the gap for deflecting beam.

In order to minimize losses during the extraction process, the goal is to construct a magnetic or electrostatic septum with as thin a septum as possible. The thinnest septa are of the electrostatic type: beam is deflected by accelerating the beam perpendicular to the initial beam direction using an electric field. Using a septum which is as thin as possible increases extraction efficiency, reduces activation of the equipment and minimizes the strengths required for other extraction elements such as kicker magnets and/or preceding septa.

5.1 Electrostatic septum

To achieve a slow-extraction efficiency of greater than 98%, the effective thickness of the first septum unit must be $\leq 100 \mu\text{m}$ [3]. This may be realised by a very carefully aligned electrostatic septum: the septum can be a foil or an array of wires.

Figure 10 shows an electrostatic septum with a foil septum: the thin septum foil results in small interaction with beam. The orbiting beam passes through the hollow foil support, which is a field-free region. The extracted beam passes just on the other side of the septum foil (high, homogeneous, field region). Electrostatic septa use vacuum as an insulator, between septum foil/wires and electrode, and are therefore in a vacuum tank. To allow precise matching of the septum position with the circulation

beam trajectory, there is often a displacement system which allows parallel and angular movement with respect to the circulating beam. The foil (or wire array — see below) is tensioned: this helps to prevent any sagging under the heat load resulting from collisions of intercepted beam particles [3].

The power supply for an electrostatic septum is typically a DC Cockroft–Walton type high-voltage generator [4].

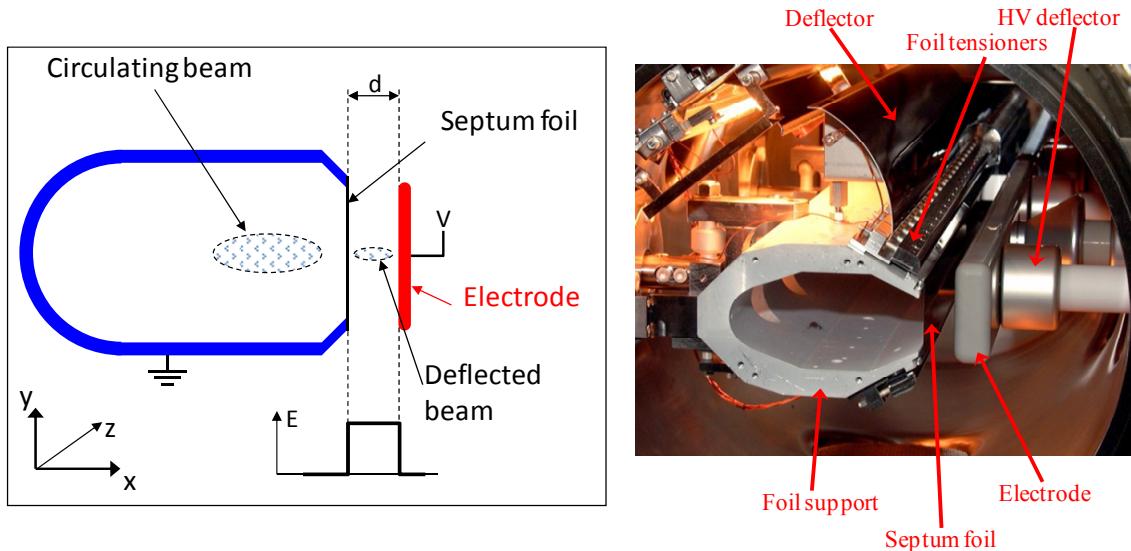


Fig. 10: Electrostatic septum with a foil septum

Typical technical data for an electrostatic septum include

- electrode length in the range 500 mm to 3000 mm;
- gap width variable in the range 10 mm to 35 mm;
- septum thickness of $\leq 100 \mu\text{m}$;
- vacuum in the range 10^{-9} mbar to 10^{-12} mbar ;
- voltage up to 300 kV;
- electric field strength up to 10 MV/m;
- septum foil of molybdenum (or tungsten-rhenium wires);
- electrode made of anodized aluminium, stainless steel or, for extremely low vacuum applications, titanium;
- some electrostatic septa are bakeable up to 300°C to achieve vacuum in the 10^{-12} mbar range (not applicable to an aluminium electrode).

A bake-out system is required for Ultra-High Vacuum (UHV) applications. In Europe UHV is generally defined to be between 10^{-6} mbar and 10^{-12} mbar .

Conditioning and preparation of surfaces exposed to High Voltage (HV) is a significant challenge.

An adjustable foil position and gap width are useful for

- precisely adjusting position ($\sim 100 \mu\text{m}$) and angle of the very thin septum foil to the beam position;

- selective conditioning of septum (e.g., increase gap width and increase voltage to condition feedthroughs, etc);
- permitting compensation for ‘as built errors’ in other equipment in the injection or extraction region.

Variants of the electrostatic septum include

- Diagonal foil: with the remote displacement at 30° from the horizontal plane. This design allows for a longitudinal painting injection scheme.
- An array of wires (Fig. 11), instead of a foil, with a diameter of ~60 µm per wire. A wire array septum allows some of the field to penetrate into the circulating beam region: the degree of penetration depends on the wire diameter and spacing. Residual gas can be ionized by the circulating beam, and the ions created can cross through the wire array into the high-field area and provoke HV breakdowns. To deal with these effects clearing electrodes, also known as ion traps, are installed (Fig. 11). These electrodes provide a vertical electric field in the circulating beam area: this electric field results in the ions being captured. At CERN, one ion trap is typically at a voltage of -3 kV and the other ion trap is typically at -7 kV: the net negative potential helps to compensate for leakage field, through the wire array, from the (negative) electrode.

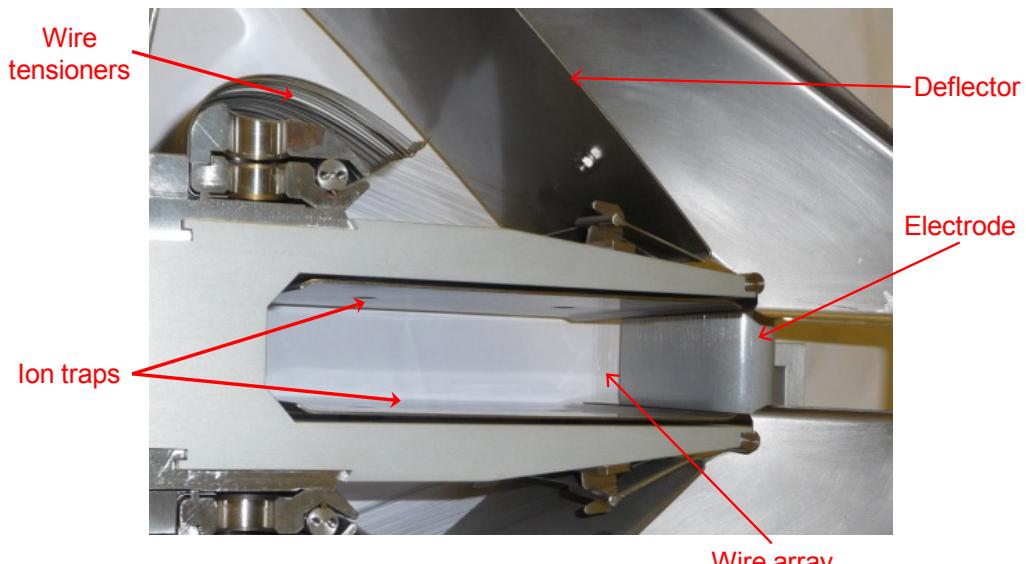


Fig. 11: Electrostatic septum with a wire array septum

5.2 Magnetic septum

The main difference between a dipole magnet and a magnetic septum is that the magnetic septum has a field-free region and a homogeneous dipole field region, separated by a relatively thin septum, whereas a dipole magnet has only a homogeneous field region. As a consequence of the relatively thin septum there is often a high current density in the septum conductor. There are several varieties of magnetic septa:

- direct-drive DC septum magnet,
- direct-drive pulsed septum magnet,
- eddy-current septum,
- Lambertson septum.

5.2.1 Direct-drive DC septum magnet

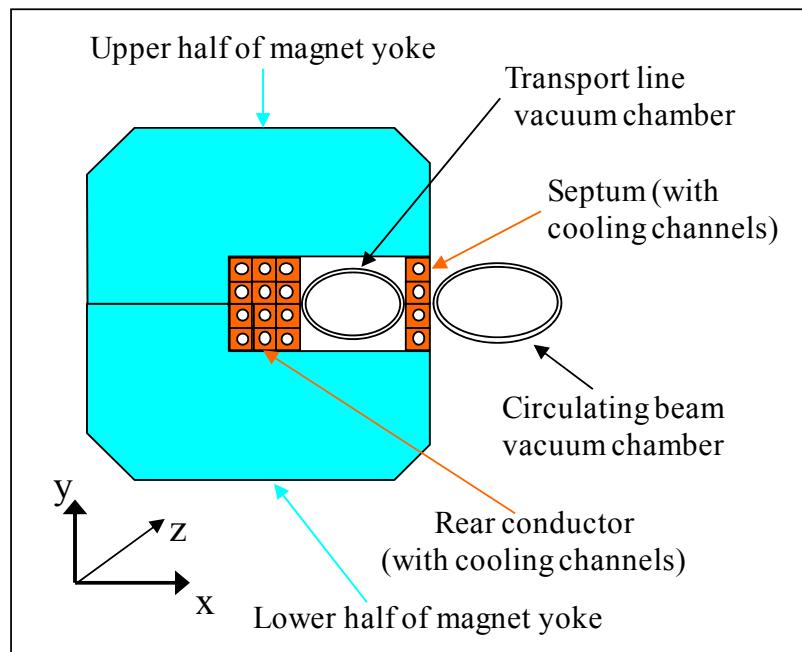


Fig. 12: Direct-drive DC septum magnet

Figure 12 shows a direct-drive DC septum magnet. The septum conductor is typically 6 mm to 20 mm thick: the current density in the septum conductor can be as high as 85 A/mm^2 . The beam to be deflected passes through the gap of the septum magnet (high, homogeneous, field region): the circulating beam is on the other side of the septum conductor. A magnetic screen may be used to further reduce the leakage field into the circulating beam region.

A DC septum magnet is often used outside vacuum: in this case the coil and the magnet yoke can be split in two, an upper and a lower part, to allow the magnet to be ‘clamped’ over the vacuum chamber. The magnet is usually constructed with a multi-turn (series) coil, so as to reduce the current needed. However, the current required is between 0.5 kA and 4 kA and the DC septum magnet has a power consumption of up to 100 kW! Thus cooling of a DC septum is a significant issue.

Typical technical data for a direct-drive DC septum magnet are

- magnetic length per magnet yoke in the range 400 mm to 1200 mm;
- gap height of 25 mm to 60 mm;
- septum thickness of 6 mm to 20 mm;
- outside vacuum;
- laminated steel yoke;
- multi-turn coil, with water cooling circuits (flow rate: 12 l/min. to 60 l/min.);
- current in the range 0.5 kA to 4 kA;
- power supplied by controllable rectifier;
- power consumption: up to 100 kW!

The effects of insulation between turns of the conductor are discussed in Section 5.3.1.

5.2.2 Direct-drive pulsed septum magnet

Figure 13 shows a direct-drive pulsed septum magnet. This type of magnet is often used under vacuum to minimize the distance between circulating and deflected beam. The coil is generally constructed as a single-turn, so as to minimize magnet self-inductance. To allow precise matching of the septum position with the circulation beam trajectory, the magnet is also often fitted with a remote displacement system.

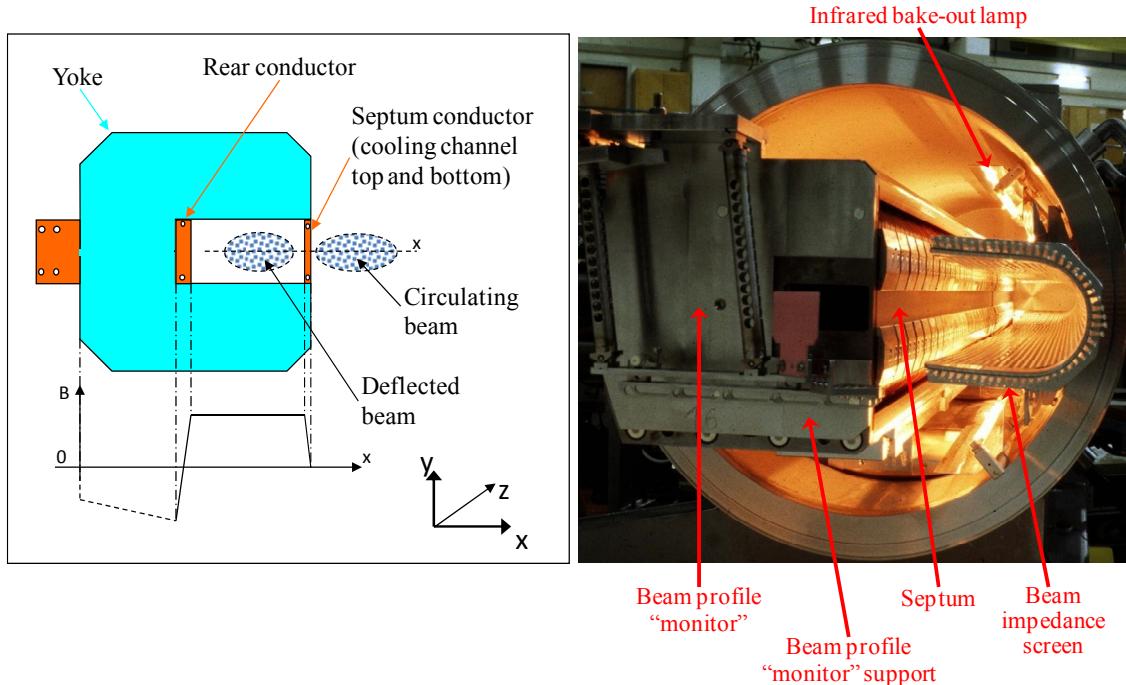


Fig. 13: Direct-drive pulsed septum magnet

Typical technical data for a direct-drive pulsed septum magnet are

- magnetic length per magnet yoke in the range 300 mm to 1200 mm;
- gap height of 18 mm to 60 mm;
- septum thickness of 3 mm to 20 mm;
- vacuum of $\sim 10^{-9}$ mbar;
- steel yoke constructed from 0.35 mm to 1.5 mm thick laminations;
- single-turn coil, with water cooling circuits (flow rate: 1 l/min. to 80 l/min.);
- bake-able up to 200°C;
- current in the range 7 kA to 40 kA, half-sine with a half-period duration of ~ 3 ms;
- power supplied by capacitor discharge. The flat top of the current is improved with a third harmonic circuit and active filters — (rectifier circuit used for up to 6 s ‘pulse’);
- a transformer is used between power supply and magnet.

High-intensity accelerators are very sensitive to longitudinal and transverse beam coupling impedance. The ‘beam impedance screen’, shown in Fig. 13, provides a continuous path for the image current of the circulating beam. Beam coupling impedance is discussed further in the proceedings of this CAS, in Section 4.5 of the paper *Injection and extraction magnets: kicker magnets*.

The pulsed septum magnet is powered with a half sine wave current with a half-period duration of typically 3 ms: the peak current density in the septum conductor is up to 300 A/mm^2 . Figure 14 shows a simplified schematic for a power supply [6, 7] for powering a pulsed septum magnet; a third harmonic circuit is used to obtain a better flat top current than is given by a basic sinusoidal discharge current:

- The capacitors are accurately charged to the required voltage.
- The ‘third harmonic circuit’ generates a current which is superimposed upon (adds to) the discharge current from the ‘fundamental circuit’.
- A transformer is used to allow the use of standard 2 kV capacitors on the primary and to give the required high current on the secondary. The transformer turns-ratio $N1:N2$ (Fig. 14) is typically in the range 4:1 to 50:1 [7].
- An active filter circuit (not shown) can be used to obtain a stability of flat top current of 10^{-4} over a time of 500 μs [7].

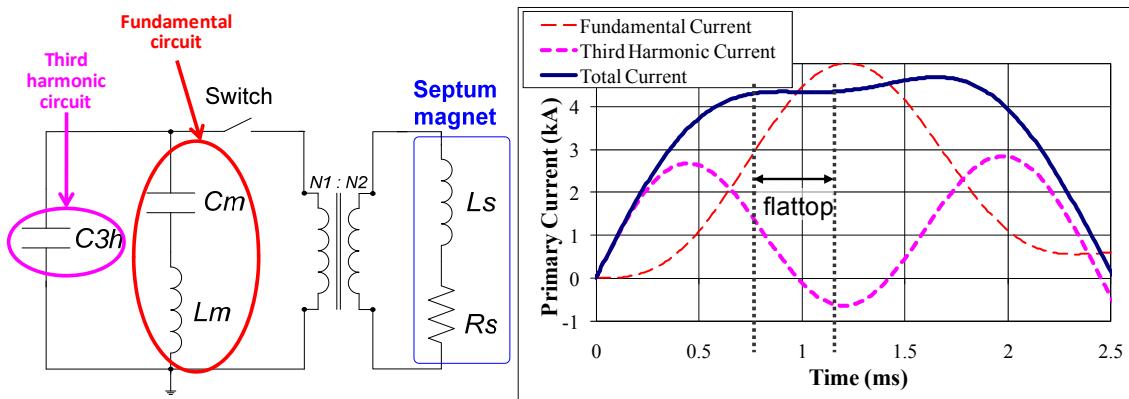


Fig. 14: Third harmonic circuit and example waveforms for a pulsed magnetic septum

5.2.3 Eddy-current septum

An eddy-current septum [8] (Fig. 15) is powered with a half or full sine wave current with a period of typically 50 μs . The coil is generally constructed as a single-turn, so as to minimize magnet self-inductance. The coil is situated around the back leg of the C-shaped yoke (Fig. 15), and therefore coil dimensions are generally not critical. When the magnet is pulsed, the magnetic field induces eddy-currents in the septum, counteracting the fringe field created. The septum conductor can be made thinner than for the direct drive septum, but cooling circuits may be needed at the edges to cool the septum.

The field in the septum gap as function of time follows the coil current. The electrical resistance of the septum is kept low: once the septum current is flowing, it takes quite some time to decay away. The Left Hand Side (LHS) of Fig. 15 shows an eddy-current septum without a return box and magnetic screen: a typical maximum leakage field would be 10% of the gap field. To reduce further the fringe field of the eddy-current septum a copper box (return box) can be placed around the septum magnet: this is shown in the Right Hand Side (RHS) of Fig. 15. In addition a magnetic screen can be added next to the septum conductor. These modifications permit the fringe field, seen by the circulating beam, to be reduced to below 0.01% of the gap field at all times and places [9].

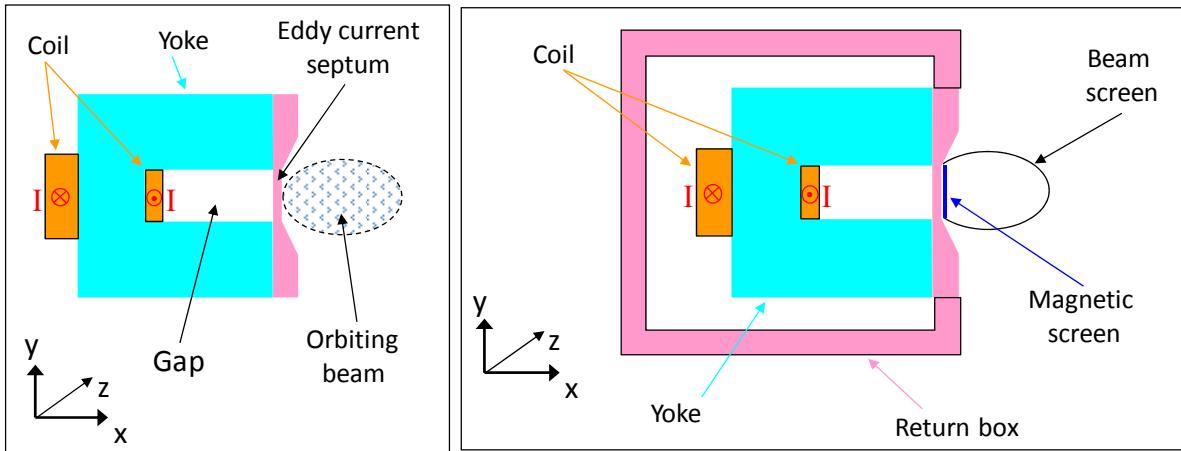


Fig. 15: Eddy-current septum: without return box or magnetic screen (LHS), with return box and magnetic screen (RHS)

Typical technical data for an eddy-current septum are

- magnetic length per magnet yoke in the range 400 mm to 800 mm;
- gap height of 10 mm to 30 mm;
- septum thickness of 1 mm to 3 mm;
- vacuum of $\sim 10^{-9}$ mbar, or out of vacuum;
- steel yoke with 0.1 mm to 0.35 mm thick laminations;
- single-turn coil, with water cooling circuits (flow rate: 1 l/min. to 10 l/min.);
- current of ~ 10 kA peak;
- fast pulsed with 50 μ s period;
- powered with a capacitor discharge: half-sine or full-sine wave.

5.2.4 *Lambertson septum*

A Lambertson iron-septum is generally a fairly rugged device [10]. The conductors are enclosed in the steel yoke, ‘well away’ from the beam. The Lambertson septum used in the LHC injection beam line is shown in Fig. 16: this septum is a DC device but pulsed Lambertson septa also exist. For the injection into the LHC, the transfer line from the SPS passes through the gap of the Lambertson septum [10]. There is a thin steel yoke between the gap of the Lambertson septum and septum hole, containing circulating LHC beam in a beam pipe (Fig. 16) — however, sufficient steel is required to avoid saturation. As shown in Fig. 16, the septum deflects beam horizontally to the right; the downstream LHC injection kicker magnets deflect the beam vertically into the septum hole, onto the central orbit of the circulating beam. To minimize field in the septum hole and beam hole (Fig. 16), containing circulating beams, an additional screen is used around each LHC beam pipe.

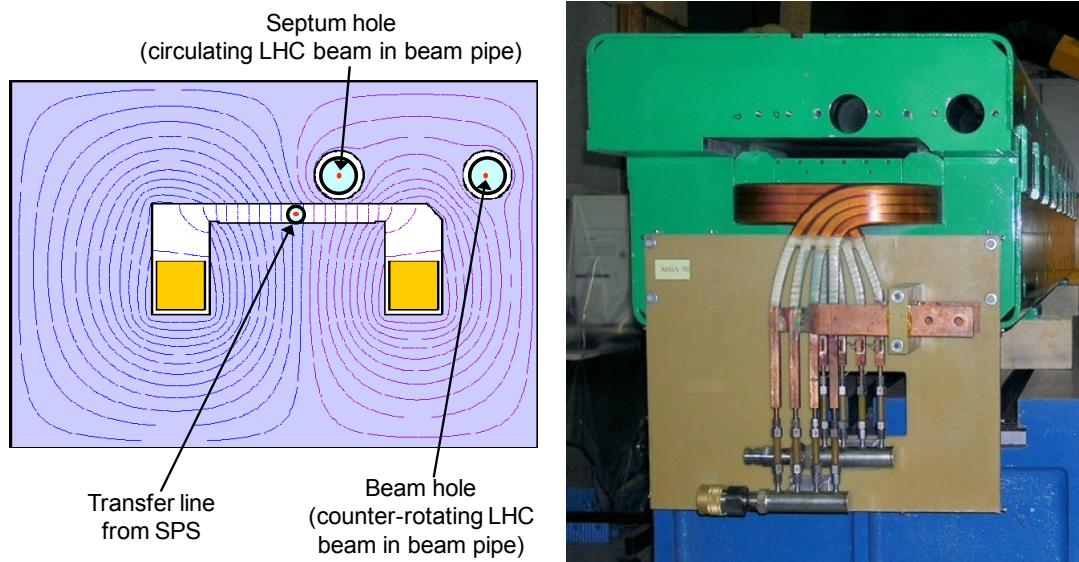


Fig. 16: Lambertson septum used in LHC injection beam-line

5.3 Practical considerations

5.3.1 *Insulation and cooling of septum magnet*

A single-turn coil may be inserted (with minimum clearance) into the gap at the outer end of a C-type yoke (Fig. 17). If the core permeability is high and the current sheet has a nearly uniform current density, there is little leakage flux outside the gap, except at the magnet ends, and excellent field homogeneity is achieved in the aperture.

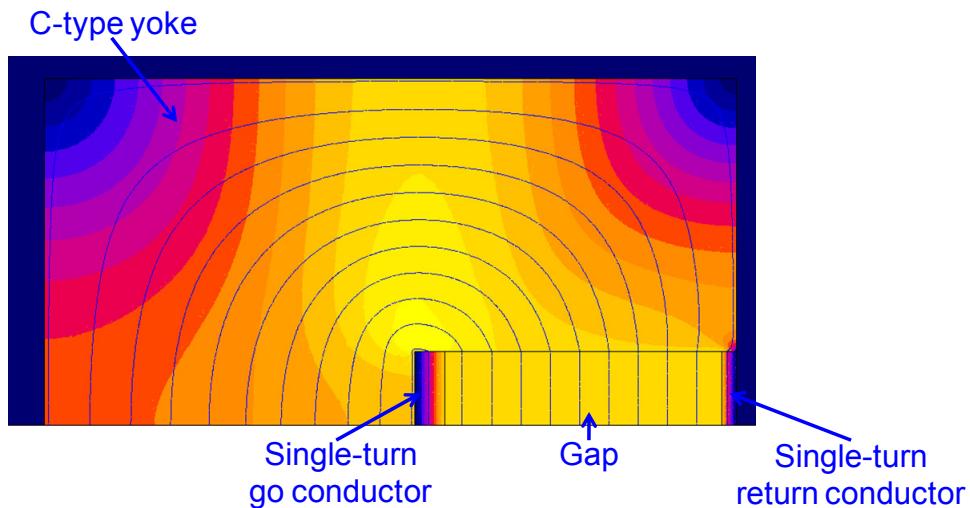


Fig. 17: ‘Idealized’ septum magnet; minimum clearance between single-turn coil and C-type yoke

However, in reality, there is a small space between the coil and yoke which is not occupied by conductor. In addition, a multi-turn septum has, by necessity, insulation between the individual windings, thus the current density is not uniform. Further the presence of cooling channels (e.g., Fig. 12) affects the uniformity of the current density. The overall effect of insulation and cooling is to increase the leakage field which, if other measures are not taken, could be up to $\sim 2\%$ of the gap field: in addition the field inhomogeneity in the gap can be up to $\pm 2\%$.

In a thin edge-cooled septum the high temperature gradient changes the local resistivity of the septum conductor and thus the uniformity of the current density: to achieve a more uniform current density, the septum is profiled to compensate for the resistivity change resulting from the temperature gradient. In addition a significant increase in the temperature of the coil and power connections can lead to load changes as seen by the power supply: thus suitable regulation of the power supply is required.

A small septum conductor thickness results in high current density, high thermal loads and high mechanical stress on the coil. Thus cooling is a major design consideration for a septum magnet. The flow characteristics (laminar/turbulent/mixed) in the cooling tube are dependent on the Reynolds number (R_e):

$$R_e = \left(\frac{\rho U_m D}{\mu} \right), \quad (4)$$

where

ρ is the density of the cooling fluid ($\text{kg}\cdot\text{m}^{-3}$),
 U_m is the mean velocity of the cooling fluid (m/s),
 D is the diameter of the cooling tube (m), and
 μ is the dynamic viscosity of the cooling fluid ($\text{N}\cdot\text{s}\cdot\text{m}^{-2}$).

Excessive cooling flow rate will lead to cavitation and erosion. Thus the cooling circuit must be designed for optimum flow conditions: turbulent flow results in best heat exchange, but laminar flow results in low erosion and cavitation. An optimum flow-condition typically corresponds to a turbulent flow with water speeds up to 10 m/s [11].

5.3.2 Vacuum considerations of magnetic septum

To reduce septum thickness, as seen by the beam (apparent septum thickness), complex, thin-walled, vacuum chambers can be used around which an outside vacuum magnetic septum can be clamped (Fig. 18). Complicated (difficult to manufacture) and UHV compatible (material quality) vacuum chambers are often required for injection/extraction points.

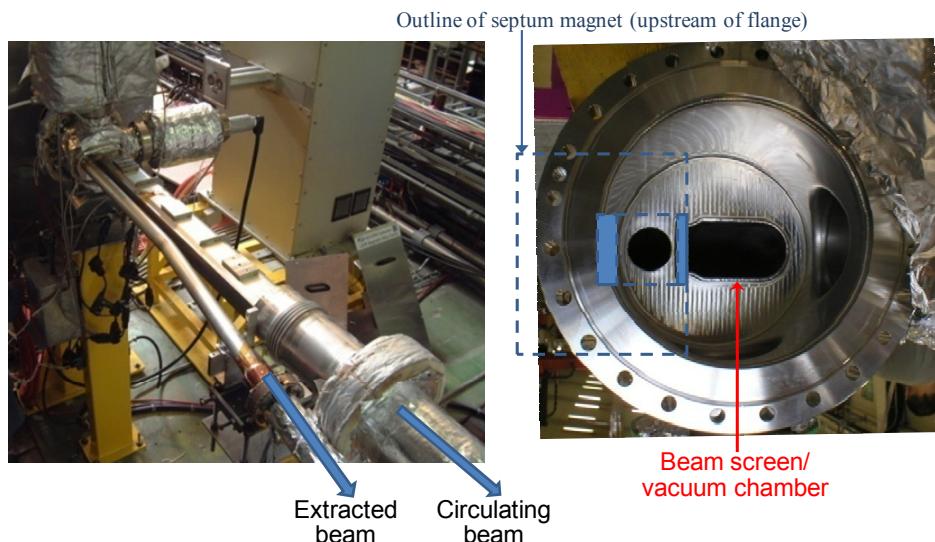


Fig. 18: To reduce apparent septum thickness, complex, thin-walled, vacuum chambers can be used (LHS), around which the septum magnet can be clamped (RHS)

To reduce apparent septum thickness even further, the magnetic septum can be put under vacuum. To reach UHV, pumping is required and bake-out may be necessary which requires suitable mechanical design and all the relevant heating equipment. In this case under vacuum heaters, heating jackets, reflectors, ceramic insulators, etc. are required. In some cases, where pressure is critical, a Non-Evaporable Getter (NEG) coating may have to be applied to the chambers, which requires activation systems.

5.3.3 Forces on a magnetic septum

The mechanical forces on the septum conductor can be significant and are normally at a maximum on the mid-axis of the septum conductor. High cycle numbers can lead to fatigue problems in thin copper cross-section. The force on the conductor (F_b) is given by Eq. (5):

$$F_b = (BI_s l_s / 2), \quad (5)$$

where

- B is the flux-density in the gap (T),
- I_s is the current in the septum conductor (A), and
- l_s is the length of the septum conductor (m).

The force on the conductor can exceed 10 kN: such force results in deflection of the conductor, of up to 40 μm , which can result in fatigue failure for a pulsed septum. The coil fixation, for a pulsed septum, is designed to be flexible and therefore uses springs (Fig. 19). The springs can be made of beryllium copper alloy and are inserted at regular intervals along the length of the coil: the spring is in contact with the septum via a lever which is then clamped in a slot in the magnet yoke (Fig. 19).

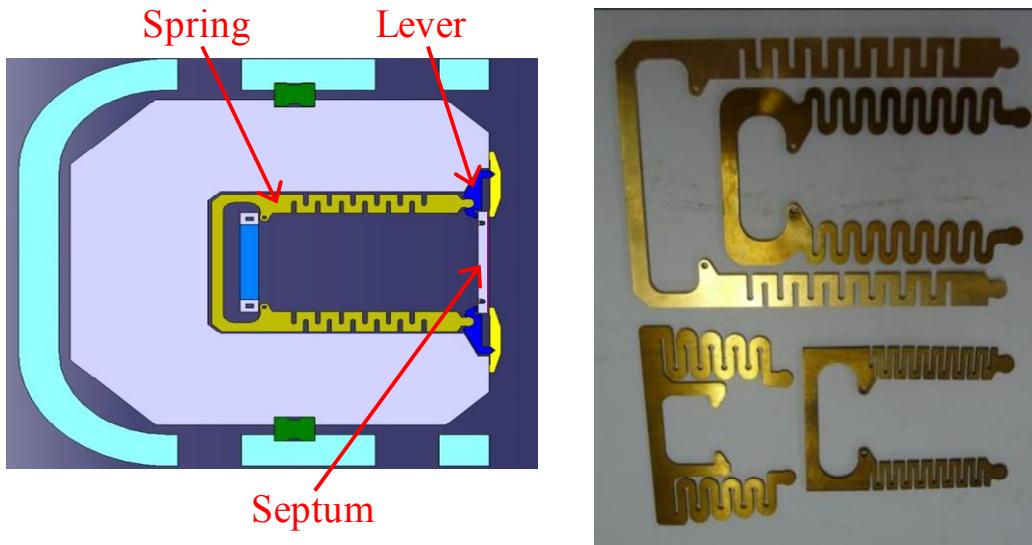


Fig. 19: Installation of springs (LHS) and examples of springs (RHS)

Acknowledgements

The authors wish to acknowledge the contributions of B. Balhan, J-M. Cravero, M. Gyr, and T. Masson to this article.

References

- [1] G.H. Rees, Injection, CAS - CERN Accelerator School: 5th General accelerator physics course, Jyväskylä, Finland, 1992, CERN 94-01 v 2.
- [2] B. Goddard, Transfer lines, Introduction to Accelerator Physics 2009, February 23–27, 2009, La Villa du Lac, Divonne, France, <http://indico.cern.ch/conferenceDisplay.py?confId=43703>
- [3] G.H. Rees, Extraction, CAS - CERN Accelerator School: 5th General accelerator physics course, Jyväskylä, Finland, 1992, CERN 94-01 v 2.
- [4] O. Bruning, CERN complex, Introduction to Accelerator Physics 2009, February 23–27, 2009, La Villa du Lac, Divonne, France, <http://indico.cern.ch/conferenceDisplay.py?confId=43703>.
- [5] C. Bovet, R. Gouiran, I. Gumowski, K.H. Reich, A selection of formulae and data useful for the design of A.G. synchrotrons, CERN/MPSSI/Int. DL/70/4.
- [6] J-M. Cravero and J.P. Royer, The new pulsed power converter for the septum magnet in the PS straight section 42, CERN PS/PO/ Note 97-03, 1997.
- [7] J.P. Royer, High current with precision flat-top capacitor discharge power converters for pulsed septum magnets, CERN/PS 95-13 (PO), 1995.
- [8] M.J. Barnes, B. Balhan, J. Borburgh, T. Fowler, B. Goddard, W.J.M. Weterings, A. Ueda, Development of an eddy current septum for Linac4, Proc. 11th European Particle Accelerator Conference (EPAC'08), Genoa, Italy, June 23–27, 2008, pp. 1434–1436.
- [9] P. Lebasque et al., Eddy current septum magnets for booster injection and extraction and storage ring injection at synchrotron soleil, Proc. 10th European Particle Accelerator Conference (EPAC'06), Edinburgh, Scotland, June 26–30, 2006, pp. 3511–3513.
- [10] S. Bidon et al., Steel septum magnets for the LHC beam injection and extraction, Proc. 8th European Particle Accelerator Conference (EPAC'02), Paris, June 3–8, 2002, pp. 2514–2516.
- [11] R.L. Keizer, Calculation of DC operated septum magnets cooling problems, CERN SI/Int. MAE/71-1.

Bibliography

- J. Borburgh, M. Crescenti, M. Hourican, T. Masson, Design and construction of the LEIR extraction septum, *IEEE Trans. on Applied Superconductivity*, Vol. 16, No. 2, June 2006.
- J. Borburgh, B. Balhan, T. Fowler, M. Hourican, W.J.M. Weterings, Septa and distributor developments for H- injection into the booster from Linac4, Proc. 11th European Particle Accelerator Conference (EPAC'08), Genoa, Italy, June 23–27, 2008, pp. 2338–2340.
- Y. Yonemure et al., Beam extraction of the POP FFAG with a massless septum, Proc. 20th Particle Accelerator Conference (PAC'03), Portland, Oregon, USA, May 12–16, 2003, pp. 1679–1681.
- S. Fartoukh, A semi-analytical method to generate an arbitrary 2D magnetic field and determine the associated current distribution, LHC Project Report 1012, CERN.
- Y. Iwashita, A. Noda, Massless septum with hybrid magnet, Proc. 6th European Particle Accelerator Conference (EPAC'98), Stockholm, Sweden, June 22–26, 1998, pp. 2109–2110.
- I. Sakai et al., Operation of an opposite field septum for the J-Parc main ring injection, Proc. 10th European Particle Accelerator Conference (EPAC'06), Edinburgh, Scotland, June 26–30, 2006, pp. 1750–1752.
- M. Benedikt, P. Collier, V. Mertens, J. Poole, K. Schindl (Eds.), LHC Design Report, volume III, The LHC injector chain, <http://ab-div.web.cern.ch/ab-div/Publications/LHC-DesignReport.html>, CERN, Geneva, 2004, CERN-2004-003.
- E. Dallago, G. Venchi, S. Rossi, M. Pullia, T. Fowler, M.B. Larsen, The power supply for the beam chopper magnets of a medical synchrotron, 37th IEEE Power Electronics Specialists Conference, 2006, (PESC '06), Jeju, Korea, June 18–22, 2006.

Permanent magnets including undulators and wigglers

J. Bahrdt

Helmholtzzentrum für Materialien und Energie, Berlin, Germany

Abstract

After a few historic remarks on magnetic materials we introduce the basic definitions related to permanent magnets. The magnetic properties of the most common materials are reviewed and the production processes are described. Measurement techniques for the characterization of macroscopic and microscopic properties of permanent magnets are presented. Field simulation techniques for permanent magnet devices are discussed. Today, permanent magnets are used in many fields. This article concentrates on the applications of permanent magnets in accelerators starting from dipoles and quadrupoles on to wigglers and undulators.

1 History

Permanent magnets were already mentioned around 600 BC by Thales of Miletus. He attributed magnets a soul because he observed that they attract small pieces of iron. People were always impressed by permanent magnet phenomena as documented in an old anecdote: A long time ago a shoemaker walked through the beautiful countryside of Greece. After a while he was surprised that his shoes started to fall apart. Finally, he noticed mysterious stones on his path that pulled out the iron nails from his sandals.

Magnetic materials were discovered two and a half thousand years ago in the Greek area Magnesia, which gave this remarkable material its name — magnetite ($\text{Fe}^{\text{II}}(\text{Fe}^{\text{III}})_2\text{O}_4$). In 200 BC the first scientific device called Si Nan (translation: pointing south) was built in China. A magnetic ‘spoon’ spins on a polished surface and once it comes to rest it points southwards. It is not known whether this device ever worked reliably as a compass. The first European compass was mentioned in 1200 AD. A piece of magnetite sitting on a wooden piece in a bowl of water aligns after a while in the north–south direction. Around 1600 William Gilbert described how to magnetize iron by mechanical deformation such as forging or drawing in the north–south direction. Also cooling down a red-hot iron bar may freeze the earth magnetic field. In 1750 the first ferrites were fabricated by Gowan Knight. He used the sintering technique which is still an essential step in today’s magnet production. In 1815 Hans Oersted discovered that a current-carrying wire produces a magnet field. Following this invention it took only six years to build the first electromagnet that was capable of magnetizing steel (William Sturgeon). In 1867 a German handbook was published that describes the fabrication of magnetic materials from non-magnetic components and vice versa.

The twentieth century saw a rapid development of various types of magnetic materials which was always driven by the demand for higher remanence, higher stability with respect to reverse fields and temperature, and last but not least by a cost effective production which implied the availability of the constituents of the material. There are only a few chemical elements that show ferromagnetic properties. Those are the transition metals Fe, Co, Ni with Curie temperatures of several 100°C and the lanthanides Eu, Gd, Tb, Dy, Ho, Er, Tm with Curie temperatures below room temperature. All relevant permanent magnets are alloys made of a large variety of components. Over the last 100 years the energy product of permanent magnets increased by several orders of magnitude (Fig. 1).

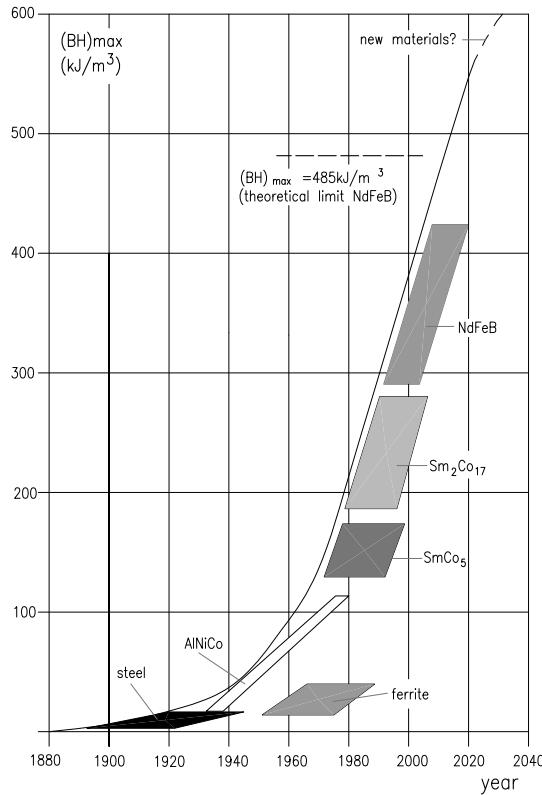


Fig. 1: Development of the energy product of various magnet types over the last century [1]

Before coming back in more detail later, we give here a brief overview of the big steps in magnet technology in the twentieth century. In 1916 it was observed that the coercivity of normal steel could be enhanced with Co additions. AlNiCo, an alloy made of Al, Ni and Co, played an important role for several decades. The grades AlNiCo 3 and 5 go back to 1931 and 1938, respectively. In 1938 the production techniques of ferrites were significantly improved in Japan. From 1945 onwards permanent magnets became comparable with electromagnets in performance and cost. New AlNiCo grades with improved performance, AlNiCo 8 and 9, were invented in 1956. Until 1970 AlNiCo was the prevalent permanent magnet material. In 1970 the first rare earth alloy, SmCo₅, was produced. Though rare earths are generally more abundant than copper or lead they are not concentrated in big mines and it is difficult to extract the rare earths from the ore. Most of the material is located in China and was not always accessible. Therefore, the research on magnets without rare earths was still ongoing, resulting in the production of FeCrCo in 1971. Owing to the Co crisis in the late 1970s the magnet suppliers started to look for alternatives with less Co content. Hard ferrites were produced all the time since the constituents are plentiful and non strategic. In 1981 Sm₂(Co,Cu,Fe,Zr)₁₇ was invented. Pure Sm is expensive, and with the elaboration of a Ca-reduction process Sm-oxides could be used as well, thus reducing the cost significantly. With the invention of Nd₂Fe₁₄B in 1983 a high performing material became available which does not require Co at all. Fe and B are plentiful and Nd is a factor of ten more frequent than Sm. In the following years the Nd₂Fe₁₄B production increased exponentially (Fig. 2) and the price per kilogram dropped one order of magnitude (Fig. 3). It is worth mentioning that the USA stopped production in 2004. In Europe only one supplier is left. In 2001 China surpassed Japan in the production rate though most of the material is still used in China. Comparing the tonnage dedicated for export China and Japan are comparable. A trend is observable that magnet suppliers in industrialized countries specialize in downstream products with a higher added value or in high-performance magnets for special applications.

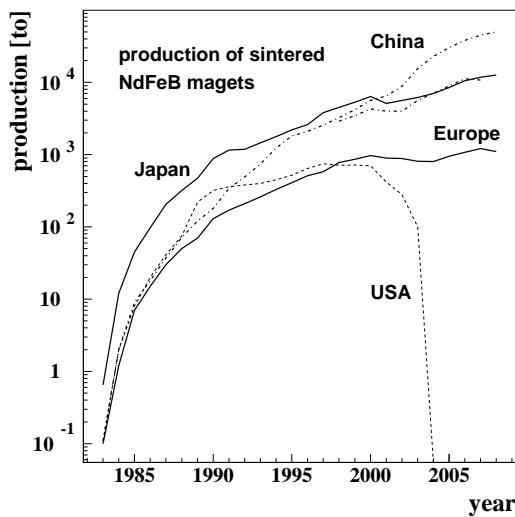


Fig. 2: Production rate of $\text{Nd}_2\text{Fe}_{14}\text{B}$ [2]. For China the total magnet production and the production dedicated for export is plotted

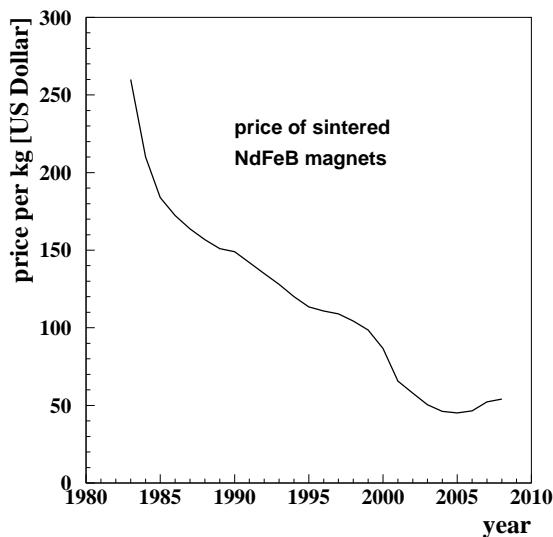


Fig. 3: Cost development of $\text{Nd}_2\text{Fe}_{14}\text{B}$ magnets [2]

2 Permanent magnet applications

Today, permanent magnets are used in many areas. A systematic classification of the applications related to the underlying physical law is given by R. Parker (see bibliography).

- *Coulomb law*: compass, magnetic bearing, magnetic coupling, fixing tool for machining, transportation line, conveyer, hysteresis device, small MRI system for medical applications.
- *Faraday law*: dynamo, generators based on wind or water energy, microphone, eddy current speedometer.

- *Lorentz force law*: loud-speaker, servo motor, voice coil motor (hard disk drive), device where the Lorentz force acts on free electrons such as: sputter facility, ion getter pump, accelerator magnet including undulator and wiggler, Halbach type dipole and higher order multipole.

The third class of applications will be discussed later in more detail. Figure 4 shows the development of applications from 1999 to 2003. Today, industrialized countries use about 50% of the magnets in voice coil motors. Table 1 gives an overview of permanent magnet applications in China.

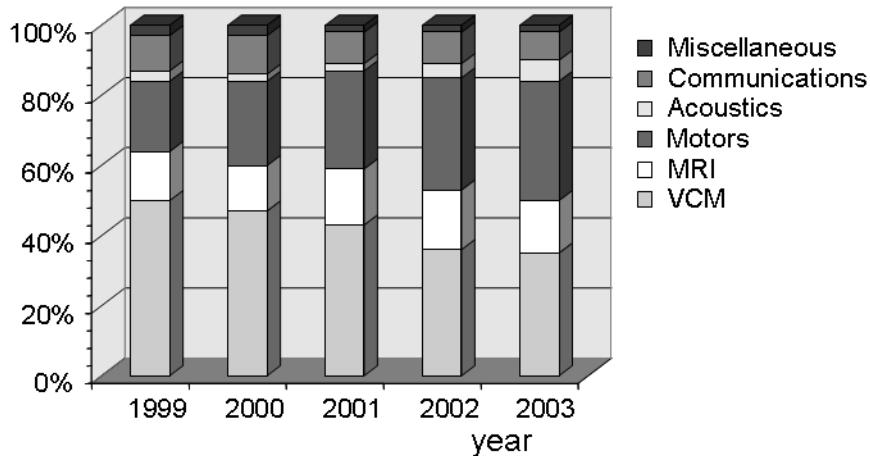


Fig. 4: Applications of permanent magnets worldwide in 2004 [3]

Table 1: Magnet applications in China in 2007

High-tech products (t)	Low-tech products (t)		
MRI	1800	Loud speaker	11280
VCM	1300	Separator	3610
CD-pickup	2515	Magnetizer	900
DVD / CD-ROM	4060		
Mobile phone	3160		
Cordless tool	3180		
Electric bike	5860		

There are several advantages of permanent magnets as compared to electromagnets: i) Machines such as motors can be built more compactly with permanent magnets. A permanent magnet array can be scaled in all three dimensions maintaining the magnetic field level at the centre. This is possible since permanent magnets can be described as blocks that carry infinite thin layers of surface currents with a constant surface current density of the order of 10 kA/cm (see Section 10). Scaling an electromagnet to smaller dimensions while keeping the field constant requires a linear enhancement of the current density. The technical limit for water-cooled coils is about 500 A/cm². ii) In principle, infinitely high fields can be produced. Let us imagine a magnet configuration that produces a field B_0 . Scaling this configuration by a factor of two in all three dimensions and adding it to the old configuration doubles the total field. Of course, the available space and price limit this procedure rather soon. Nevertheless, there are many devices whose fields exceed the remanent field by large factors, e.g., Iwahita built a 3.9 T dipole magnet [4]. iii) The power consumption is zero and, thus, there are no cooling problems. iv) Permanent magnet devices are failsafe as long as they are operated below the Curie temperature.

3 Basic definitions

In this article we use Gaussian units. The macroscopic property of magnetic material is described by the dependence of magnetic induction and magnetization on the external field. The zero crossing of the induction or magnetization between the second and third quadrant is called coercive force (the absolute value of the external field as plotted in Fig. 5). The location of this crossing can be significantly different for the induction (H_c) and the magnetization (H_{cj}). The BH -dependency is reproduced if the material is periodically driven to complete magnetization in one direction and complete magnetization in the opposite direction by tuning the external field. A non-magnetized block shows a different dependency which is called the initial magnetization curve. The local dependence between B and H is described by the permeability. Apart from the usual permeability $\mu = dB/dH$ which describes the linear part in the first and second quadrant we differentiate between the initial, the differential or maximum, and the reversal or recoil permeability (Fig. 6). The reversal permeability describes the properties in the non-linear part close to the knee. For small field variations the BH -curve exhibits a small loop around a straight line which is approximately parallel to the slope of the BH -curve above the knee. Nd₂Fe₁₄B magnets have a different permeability parallel and perpendicular to the easy axis of about $\mu_{par} = 1.04$ and $\mu_{perp} = 1.17$. Depending on the working point of the magnet, i.e., the operating point on the hysteresis loop, the remanence can be lower than $B_r = B(H=0)$ by a few per cent.

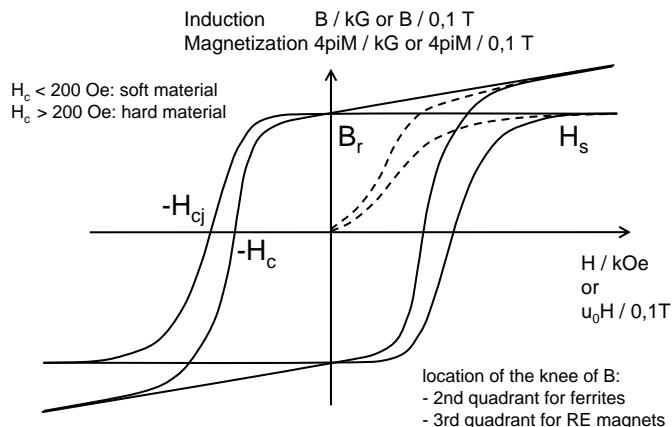


Fig. 5: Hysteresis loop of a magnetic material

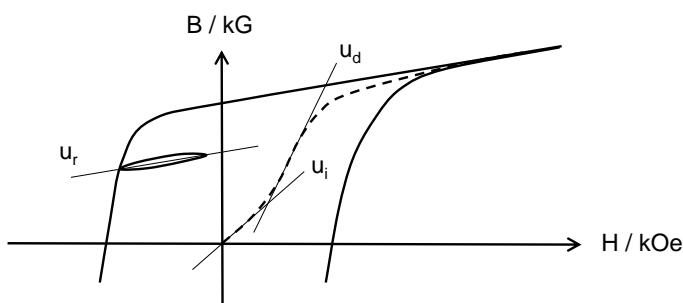


Fig. 6: Initial permeability μ_i , differential or maximum permeability μ_d , and reversal or recoil permeability μ_r

We call a material hard magnetic if $H_c > 200$ Oe which is a bit arbitrary. The performance of a magnet is related to the maximum energy product. It is equal to the largest rectangle which can be plotted under the BH -curve (Fig. 7). Though modern magnets have a nearly rectangular BH -curve we have always

$$(BH)_{\max} \leq B_r^2 / \mu .$$

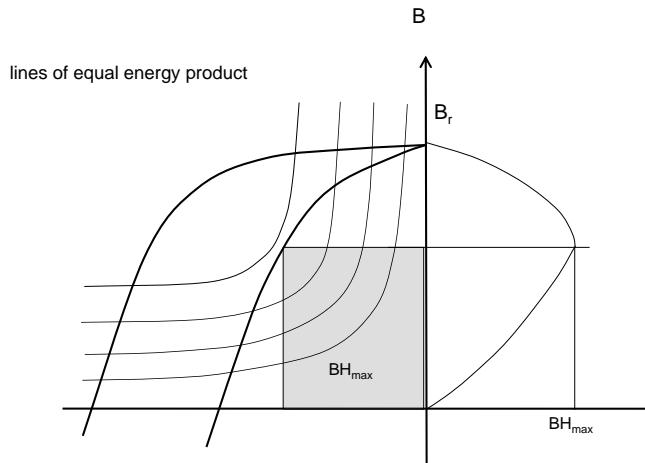


Fig. 7: Two graphic representations of the maximum energy product

The efficiency of an electric circuit is given by the conductivity = 1 / resistance = current / voltage. In analogy the efficiency of a magnetic circuit is written as: permeance = 1 / reluctance = flux / magnetomotive force difference. The magnetomotive force is the magnetic potential as produced by currents or magnetized samples. The permeance of a volume V is defined by

$$P = \frac{\iint \vec{B} \cdot d\vec{s}}{\int \vec{H} \cdot d\vec{l}}$$

where the integral over B is taken at the plane B and the line integral over H is taken between the planes A and C (Fig. 8)

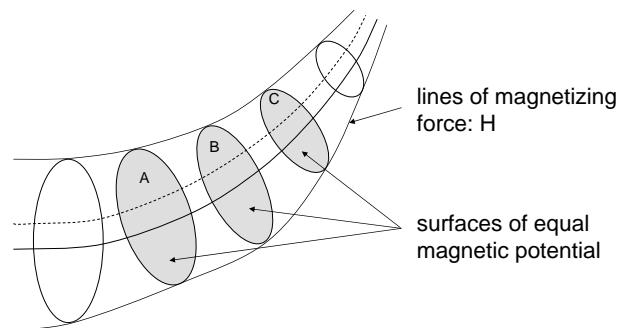


Fig. 8: For the definition of the permeance see text

Inside a magnetized body demagnetizing forces proportional to the magnetization are present. The strength is given by the demagnetization factor D which varies between 0 and 4π depending on the shape of the sample (small for pencil-like samples with the easy axis pointing in the long direction and large for flat samples). The working point of a material is given by B_d and H_d where B_d / H_d is the coefficient of demagnetization or unit permeance. The line connecting the origin and the working point is called the load line. In a magnetic design the working point has to be chosen such that it is well above the knee for all possible operation conditions. Knowing the demagnetization factor the working point can be derived according to:

$$\begin{aligned}\vec{B} &= \vec{H} + 4\pi \cdot \vec{M} \\ \vec{H}_d &= -D \cdot \vec{M} \\ \vec{B}_d &= \vec{H}_d - \frac{4\pi}{D} \vec{H}_d \\ \frac{\vec{B}_d}{\vec{H}_d} &= 1 - \frac{4\pi}{D}\end{aligned}$$

In Cartesian coordinates we always have: $D_x + D_y + D_z = 4\pi$. For specific geometries the literature provides tables and approximated expressions for the evaluation of demagnetization factors D (see the book by R. Parker in the bibliography). More complicated geometries have to be evaluated numerically or measured. Depending on the measurement method one gets either magnetometric or fluxmetric (ballistic) demagnetization factors. In the first case a long coil is placed along the complete sample yielding averaged values for the whole sample, in the second case a flat coil is placed around the centre cross section yielding a value averaged over this area. The magnetometric values are always larger than the fluxmetric values.

Analytic expressions of demagnetization factors are available for a generalized ellipsoid and for a rectangular prism. Osborn [5] gives values for a generalized ellipsoid with the semi-axes $a \geq b \geq c$:

$$\begin{aligned}D_x &= \frac{4\pi \cos(\varphi) \cos(\vartheta)}{\sin^3(\vartheta) \sin^2(\alpha)} (F(k, \vartheta) - E(k, \vartheta)) \\ D_y &= \frac{4\pi \cos(\varphi) \cos(\vartheta)}{\sin^3(\vartheta) \sin^2(\alpha) \cos^2(\alpha)} \left(E(k, \vartheta) - F(k, \vartheta) \cos^2(\alpha) - \frac{\sin^2(\alpha) \sin(\vartheta) \cos(\vartheta)}{\cos(\varphi)} \right) \\ D_z &= \frac{4\pi \cos(\varphi) \cos(\vartheta)}{\sin^3(\vartheta) \cos^2(\alpha)} \left(\frac{\sin(\vartheta) \cos(\varphi)}{\cos(\vartheta)} - E(k, \vartheta) \right)\end{aligned}$$

$$\cos(\vartheta) = c/a$$

$$\cos(\varphi) = b/a$$

$$\sin(\alpha) = \sin(\varphi) / \sin(\vartheta) = k$$

F and E are elliptical integrals of the first and second kind with k = modulus and θ = amplitude. It has to be emphasized that these values are exact and they are constant over the whole volume. Special cases are the sphere with $D_x = D_y = D_z = 4\pi/3$, the infinite long circular cylinder with $D_{par} = 0$ and $D_{perp} = 2\pi$ or an infinite wide plane with $D_{in-plane} = 0$ and $D_{perp-plane} = 4\pi$.

Averaged demagnetization factors for a parallel prism are derived by Aharoni [6]. In reality the factors vary over the volume.

$$\begin{aligned}
D_z / 4 &= \frac{b^2 - c^2}{2bc} \ln\left(\frac{sabc - a}{sabc + a}\right) + \frac{a^2 - c^2}{2ac} \ln\left(\frac{sabc - b}{sabc + b}\right) + \frac{b}{2c} \ln\left(\frac{sab + a}{sab - a}\right) + \frac{a}{2c} \ln\left(\frac{sab + b}{sab - b}\right) \\
&+ \frac{c}{2a} \ln\left(\frac{sbc - b}{sbc + b}\right) + \frac{c}{2b} \ln\left(\frac{sac - a}{sac + a}\right) + 2 \arctan\left(\frac{ab}{c \cdot sabc}\right) + \frac{a^3 + b^3 - 2c^3}{3abc} \\
&+ \frac{a^2 + b^2 - 2c^2}{3abc} sabc + \frac{c}{ab} (sac + sbc) - \frac{sab^3 + sbc^3 + sac^3}{3abc} \\
sabc &= \sqrt{a^2 + b^2 + c^2} \\
sab &= \sqrt{a^2 + b^2} \\
sac &= \sqrt{a^2 + c^2} \\
sbc &= \sqrt{b^2 + c^2}
\end{aligned}$$

In analogy D_x and D_y can be derived. Special cases are a cube with $D_x = D_y = D_z = 4\pi/3$ and an infinite long rectangular cylinder with

$$D_{par} = 0$$

$$D_{perp} / 4 = \frac{1 - p^2}{2p} \ln(1 + p^2) + p \cdot \ln(p) + 2 \cdot \arctan(1/p)$$

$$p = c/a$$

For more complicated geometries or if the demagnetization distribution over a rectangular block is needed, analytic expressions do not exist and the demagnetization factors have to be evaluated numerically. Figure 9 shows the variation of the demagnetization over rectangular blocks with different shapes.

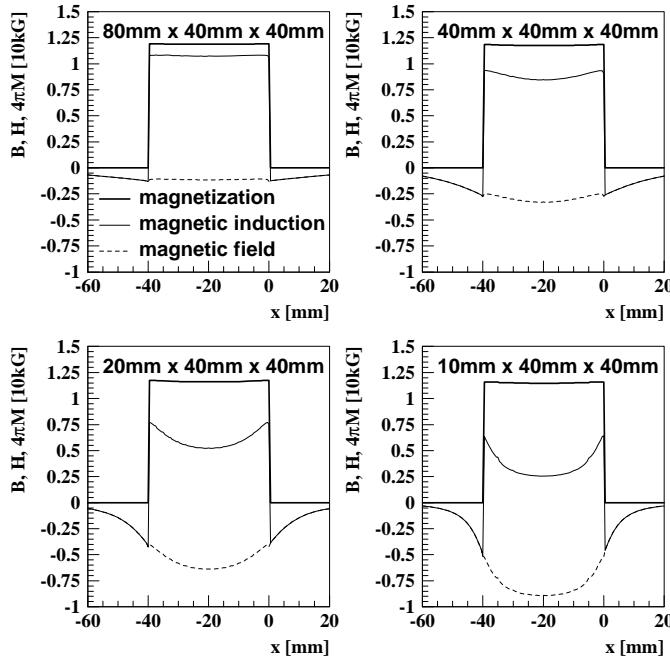


Fig. 9: Magnetic induction, magnetic field, and magnetization for four blocks with different geometries. The length in the easy axis direction is varied as indicated in the graph

4 Permanent magnet types

In this section the magnetic properties of the most common magnet types are summarized. This overview will give only a taste of the diversity of magnet properties. The magnet material has to be chosen carefully according to the demands: Some applications need high remanence, others require high coercivity or high temperature stability. Some applications do not require high performance and only the price counts. The properties of high-end grades are listed in Tables 2–9.

Permanent magnets are either of type I or type II (Fig. 10). Type I magnets have a high leakage flux leaving the magnet at the sides. The energy stored in these leakage fields is not usable. The permeability of these materials is large and H_{cj} is usually smaller than B_r . Typical examples are 35%Co Fe or AlNiCo. Type II magnets have a low leakage flux, the permeability is close to one and H_{cj} is much larger than B_r . Rare earths based magnets as well as hard ferrites belong to this species. Sintered magnets are either isostatic pressed (IP), transversally pressed (TP), or axially pressed (AP).

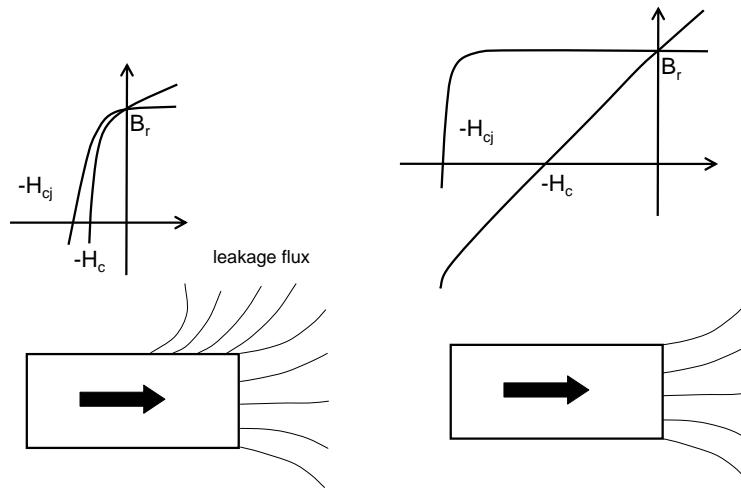


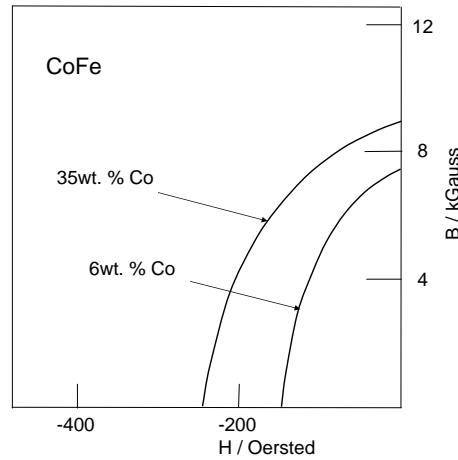
Fig. 10: Type I magnets (left) and type II magnets (right)

4.1 Carbon steel or martensitic steel

Pure carbon steel (up to one per cent carbon) has a rather low energy product which can be significantly enhanced with the addition of Co (Fig. 11). The performance can further be improved with other non-magnetic ingredients. Internal strain and lattice imperfections can also have a positive effect on the performance.

Table 2: Magnetic properties of carbon steel

Grade	Remanence (kG)	Coercivity H_{cj} (kOe)	Energy product (MGoe)
3.5 Cr	9.8	0.05	0.22
36.0 Co	9.6	0.24	0.94

**Fig. 11:** Carbon steel

4.2 AlNiCo

The magnets consist of an alloy of Al, Ni, Co and also Fe, Cu, Ti. The remanence is pretty high whereas the coercivity is low which has to be taken into account in the magnetic design. Some grades can be operated at temperatures up to 550°C. The energy product can be enhanced when an anisotropy is deliberately introduced, e.g., by cooling the blocks in a magnetic field. The material is extremely difficult to machine and very brittle which requires a near to finish production. Better mechanical properties have been achieved with sintered magnets. The temperature coefficients are as low as -0.02% for B_r and -0.02% up to 0.01% for H_{cj} .

Table 3: Typical magnetic properties of a few AlNiCo grades

Grade	Remanence (kG)	Coercivity H_{cj} (kOe)	Energy product (MGoe)
AlNiCo 5 cast	13.5	0.74	7.5 0
AlNiCo 9 cast	10.6	1.5	10.0
AlNiCo 5 sintered	11.2	0.61	4.4
AlNiCo 8 sintered	8.0	1.60	4.5

4.3 FeCoCr

The material properties are similar to those of AlNiCo 5 but this magnet requires less Co. Furthermore, it has a higher ductility and it can be oriented by mechanical deformation.

Table 4: Magnetic properties of FeCoCr

Grade	Remanence (kG)	Coercivity H_{cj} (kOe)	Energy product (MGoe)
FeCoCr	13.0	0.55	5.0

4.4 MnCaI

This material does not need Co at all. It has a higher ductility than AlNiCo. The magnet can be oriented in a warm extrusion process. The extrusion is, however, rather expensive.

Table 5: Magnetic properties of MnCaI

Grade	Remanence (kG)	Coercivity H_{cj} (kOe)	Energy product (MGOe)
MnCaI cast	3.0	0.95	1.0
MnCaI cast & exruded	6.0	2.5	7.0

4.5 Hard ferrites

The chemical composition is $MO_6(F_2O_3)$ or $MFe_{12}O_{19}$ with $M = Ba, Sr$ or Pb . The magnets are sintered. They are either isotropic or oriented. The remanence is low but the coercivity is pretty high. The material has large temperature coefficients of -0.2% for the remanence and +0.1 up to 0.5% for the coercivity. H_c and H_{cj} are similar and the knee is located in the second quadrant.

Table 6: Magnetic properties of hard ferrites

Grade	Remanence (kG)	Coercivity (kOe)	Energy product (MGOe)
Hard ferrite	4.0	3.6 (4.0), $H_c (H_{cj})$	4.0

4.6 SmCo₅

Samarium based rare earth magnets have a high energy product. The high coercive grade is the favourite material in the presence of strong reverse fields or in an ionizing radiation environment. The maximum operation temperature is 250°C. The temperature coefficients are as low as -0.045% / deg. and -0.22% / deg. (H_{cj}).

Table 7: Magnetic properties of SmCo₅

Grade	Typ. remanence (kG)	Min. coercivity H_{cj} (kOe)	Typ. energy product (MGOe)
IP	10.1	12.5	25.0
IP	9.0	25.0	20.0

4.7 Sm₂Co₁₇

The material has a higher remanence than SmCo₅. The maximum operation temperature is 350–500°C depending on the grade. The temperature coefficients are lower as compared to SmCo₅: -0.035% / deg. and -0.12% / deg. (H_{cj}).

In comparison with Nd₂Fe₁₄B magnets, SmCo₅ and Sm₂Co₁₇ magnets are brittle and the handling is delicate. The temperature coefficients are lower as compared to Nd₂Fe₁₄B which is advantageous in temperature sensitive applications.

Table 8: Magnetic properties of $\text{Sm}_2\text{Co}_{17}$

Grade	Typ. remanence (kG)	Min. coercivity H_{cj} (kOe)	Typ. energy product (MGoe)
IP	11.0	20.0	28.0
TP	10.7	20.0	27.0
AP	10.4	20.0	25.0

4.8 $\text{Nd}_2\text{Fe}_{14}\text{B}$

This material has the highest remanence among all rare earth based permanent magnets. The temperature gradients are higher than in Sm based magnets: -0.09 to -0.11% / deg for B_r and -0.45 to -0.6% for H_{cj} . Depending on the amount of added dysprosium the coercivity can be enhanced significantly sacrificing remanence. The maximum operation temperature depends on the grade and varies between 80°C for material of highest remanence and 230°C for material with highest coercivity.

Table 9: Magnetic properties of $\text{Nd}_2\text{Fe}_{14}\text{B}$

Grade	Typ. remanence (kG)	Min. coercivity H_{cj} (kOe)	Typ. energy product (MGoe)
IP	14.7	11	53
TP	14.1	14	48
AP	13.4	14	43

5 Temperature dependence

5.1 Compensation of temperature dependent effects

Critical permanent magnet applications need a compensation of the temperature dependent remanence changes. This can be accomplished with various strategies:

- Curie alloys consisting of Ni and Fe have a negative temperature gradient of the permeability. They can be used as temperature sensitive flux shunts. With rising temperature they shortcut less flux, and thus compensate for the remanence loss of the permanent magnets. These materials are used, for example, in speedometers. Other applications are accelerator magnets. In the 344 permanent magnet gradient dipoles of the 8.9 GeV antiproton recycler ring at Fermilab, thin Ni-Fe bars shortcut part of the magnetic material (strontium ferrite) between the Fe pole tips and the iron yokes [7].
- The next Brazilian light source, LNLS II, will be based on 48 permanent magnet gradient dipoles made of hard ferrites [8]. The gap of the magnets is passively readjusted with temperature.
- Compared to $\text{Nd}_2\text{Fe}_{14}\text{B}$ material SmCo_5 has a small remanence temperature coefficient. It can further be reduced by mixing the material with other compounds with a positive temperature gradient such as ErCo_5 and/or GdCo_5 .

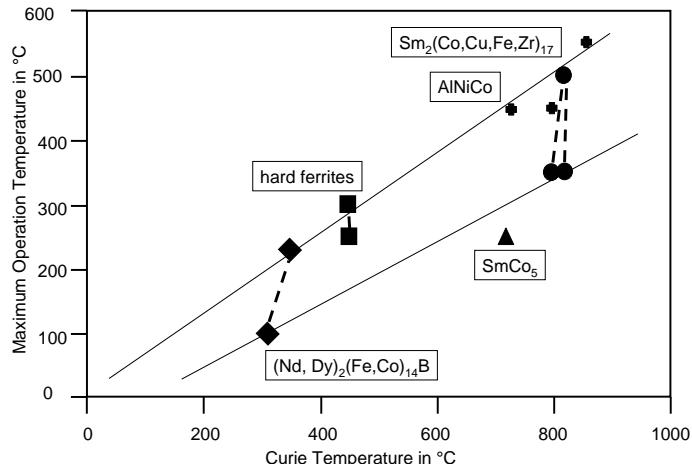
5.2 Curie temperature

The temperature range for a safe operation depends on the Curie temperature. Above the Curie temperature the remanence and coercivity drop to zero and the material becomes paramagnetic. The Curie temperatures of some magnetic materials are summarized in Table 10.

Table 10: Curie temperatures of typical elements and permanent magnets

Grade	Curie temperature (°C)
Iron	770
Cobalt	1130
Ni	358
$\text{Nd}_2\text{Fe}_{14}\text{B}$	310
SmCo_5 , $\text{Sm}_2\text{Co}_{17}$	700–800
35% Co steel	890
CrFeCo	630
AlNiCo	850
Hard ferrites	400

The coercivity of $\text{Nd}_2\text{Fe}_{17}\text{B}$ magnets can be enhanced with the addition of Dy (Fig. 12). On the other hand, the Dy reduces the remanence. The maximum operation temperature of $\text{Sm}_2\text{Co}_{17}$ grades can be raised with a specific tempering procedure. In both cases the Curie temperature of the material grows a bit but much less than the coercivity. Optimized permanent magnets can be used up to about 75% of the Curie temperature.

**Fig. 12:** Temperature stability versus Curie temperature of various magnet materials [9]. The dotted lines connect different grades of the same magnet material (for details see text)

Only recently, Hitachi described a vapour deposition and diffusion process which enhances the coercivity of $\text{Nd}_2\text{Fe}_{14}\text{B}$ magnet by about 4 kOe without sacrificing remanence. Alternatively, the remanence can be enhanced by 400 G without losing in coercivity [10]. The magnets are exposed to dysprosium vapour. The Dy atoms diffuse along the grain boundaries into the bulk material without penetrating into the grains. The penetration depth is only a few millimetres which limits this method to thin magnets. This material is of particular interest for in-vacuum undulators where the magnets are close to a several GeV electron beam and the demagnetization stability is an important issue.

5.3 Reversible demagnetization

Operating permanent magnets above the Curie temperature causes a complete demagnetization. The crystal structure remains unchanged, and hence the magnets can be remagnetized recovering full

performance. Since the demagnetization factor varies over the magnet volume, certain parts demagnetize earlier than more stable regions. For critical applications the magnet supplier deliberately ages the blocks, in order to avoid a remanence loss over the years. For this purpose the magnets can either be heated well above the final operation temperature or they can be stabilized in a reverse field.

Synchrotron radiation light sources are based on periodic permanent magnet structures, so-called undulators, which are passed by relativistic electrons emitting photons in the VUV to X-ray regime (see Sections 12–14). If the magnets are hit by off-axis electrons they may get demagnetized. Demagnetization has been reported at the ESRF [11], the APS [12–13] (see Fig. 13) and DESY [14]. Though the remanence losses reached 10% the structures could be repaired in these cases by remagnetizing the blocks. If, however, the ionizing radiation changes the crystal structures, irreversible losses may occur [15].

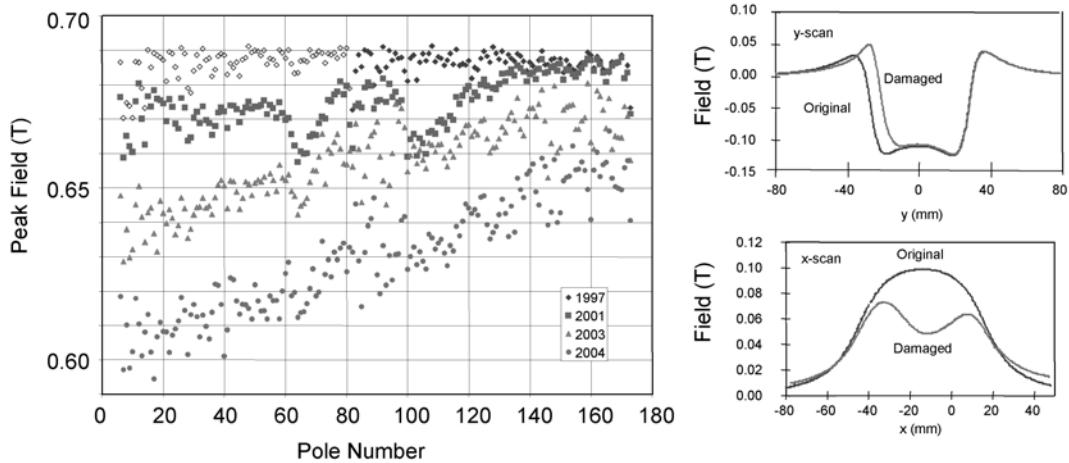


Fig. 13: History of sector 3 undulator at the APS (left) and single block Hall probe scans (right). The damage is localized at the surface close to the electron beam (courtesy of L. Moog, APS)

6 Permanent magnet fabrication

6.1 Sintered magnets

Sumitomo developed the fabrication process for sintered rare earth magnets. The various ingredients are mixed in the desired percentages, melted under vacuum conditions and cast to macroscopic pieces. The pieces are crushed and then milled in several steps down to particle sizes of a few micrometres. The powder is highly reactive and has to be processed under inert gas conditions. The powder is pressed in the presence of a high field which aligns the magnetic domains. The pressed pieces are sintered at temperatures around 850°C which melts the surfaces of the grains (so-called liquid phase sintering). The quality of the magnet depends on the type of pressing.

Highest remanence is achieved with isostatic pressing. Here, the powder is poured into long rubber tubes which are placed into a liquid environment. The powder is compressed with equal pressure from all directions. The magnetic field is applied with a long coil surrounding the tube. These magnets have rather large dipole orientation errors whereas the homogeneity is pretty good. Alternatively, the powder can be pressed in a die. Depending on the geometry, the pressing direction can be parallel (axial pressing) or perpendicular to the field lines (transverse pressing). The latter geometry yields a higher remanence but it can not be applied for block geometries with high aspect ratios. Die pressed magnets have very similar magnetic properties over the whole batch since they all

see the same magnetic environment during the pressing. The dipole errors are small but the inhomogeneities can be larger than in isostatic pressed magnets. Die pressing is preferred for large batches because the die can be shaped close to the final geometry and the machining time afterwards can be minimized (near net-shape production). More details on the development and the characteristics of sintered Nd₂Fe₁₄B magnets are given in Ref. [16].

6.2 Melt spun magnets

In 1984 General Motors developed another fabrication method for rare earth permanent magnets. The melted ingredients are poured into an induction heated container under Ar atmosphere. A liquid alloy jet is quenched on a water cooled spinning wheel forming a 300 µm microcrystalline ribbon (Fig. 14). This material is further processed following one of three procedures:

- Magnequench I: The ribbons are bonded to form a solid block which can further be machined. The material is isotropic.
- Magnequench II: The ribbons are pressed under high temperature. The material is isotropic.
- Magnequench III: Starting with Magnequench II material the blocks are deformed under high temperature resulting in an anisotropic grade. This material has the highest energy product among the three grades.

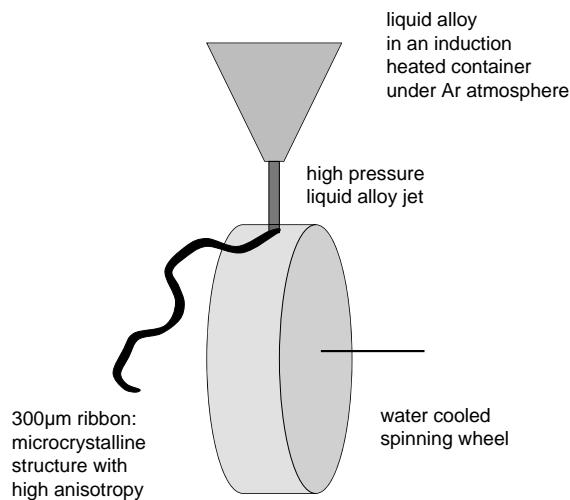


Fig. 14: Magnet fabrication following the Magnequench technique

7 Measurement techniques for macroscopic properties

The magnetic hysteresis can be measured either in a closed loop or in an open loop geometry. At room temperature or higher temperatures a hysteresis graph can be used (Fig. 15). Here, the sample is clamped between two iron yokes which short-cut the magnetic flux. Powering a coil that is wound around the iron yoke the external field can be varied. The parameters B and H can be directly measured with pickup coils. Large samples (a few cubic centimetres) can be characterized which minimizes the impact of surface effects.

At low temperatures another method has to be used, i.e., a vibrating sample spectrometer: In a He cryostate a small sample vibrates back and forth inducing a voltage in a pickup coil. The signal is proportional to the dipole moment of the sample. Knowing the sample geometry, this value can be

converted to the remanence. Usually, the samples must be small (a few cubic millimetres) and the surface effects have to be taken into account for the interpretation of the results. Furthermore, a precise calibration of the setup is required which includes the demagnetization factor of the sample. Ellipsoidal samples with a constant demagnetization factor are preferable but difficult to fabricate and align.

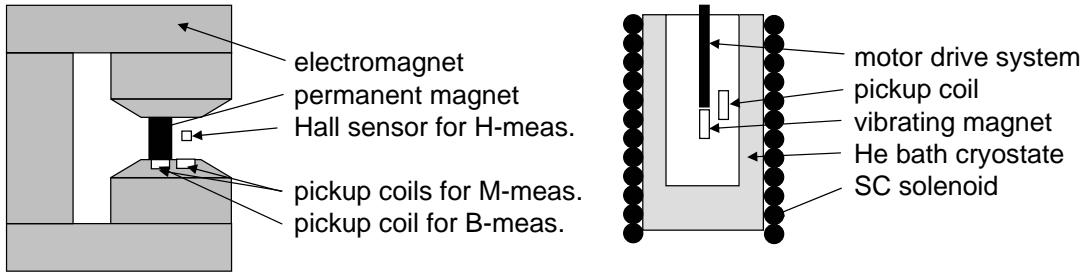


Fig. 15: Hysteresis graph (left) and vibrating sample magnetometer (right)

Usually, the dipole moment is measured in a Helmholtz coil with high accuracy. The geometry is optimized such that the measurement is insensitive to a displacement of the magnet and the magnet block size. The magnetizing force of two coils with winding N and current I is

$$H = \frac{2NI}{a} \left[\left(1 + \frac{(d/2 - x)^2}{a^2} \right)^{-1.5} + \left(1 + \frac{(d/2 + x^2)}{a^2} \right)^{-1.5} \right]$$

where d = distance of the coils and a = radius of coil.

In the Helmholtz geometry, $d = a$, the quadratic terms disappear. An automated rotation of the magnet around all three axes and an averaging over several turns yields rms reproducibilities better than 0.07% for the main component and better than 0.04° for the easy axis orientation errors with respect to a reference surface (values of the BESSY II system).

The dipole moment can also be measured with a fluxgate sensor located at a large distance (typically 1 m) from the sample. In the far field only the dipole component contributes and higher orders can be neglected. The signals are small and the magnet has to be flipped to get rid of environmental fields. This method is adequate for the main component but less accurate for the minor components.

Close to the magnet surface (i.e., a few millimetres for magnets with dimensions of a few centimetres) the magnet fields can not be described by the dipole moment only. Higher order multipoles have to be included. These higher order terms can be measured with a set-up where the magnet is moved with respect to a fixed wire. The rms reproducibility of the BESSY II system, which uses a single fixed wire, is summarized in Table 11. In other laboratories alternative fluxmetric methods are applied: at SPRING-8 a multifilament rotating coil close to the magnet surface is used [17]. At the ESRF, modules of several magnets are continuously moved at a speed of 60 mm/s passing a fixed multifilament wire (20 single wires) [18]. The modules include an equal number of magnets with the easy axis pointing to the wire and into opposite direction. Thus the modules are magnetically compensated and the field integrals are below 100 G cm. The rms error of a module measurement is 1 G cm.

The information on dipole errors and magnet field inhomogeneities can be used to sort the magnets such that the fabrication errors cancel in the assembled structure.

Table 11: Reproducibility of field integral measurements of single magnet blocks (BESSY II system). The blocks have a distance of 5 mm to the wire. The orientation of the easy axis is given with respect to the wire

Easy axis orientation	Absolute error (rms values)	Relative error (rms values)
Parallel	$3 \cdot 10^{-4}$ T mm	$6 \cdot 10^{-3}$
Perpendicular	$1.5 \cdot 10^{-3}$ T mm	$2.5 \cdot 10^{-4}$

The capacity of the furnace crucible defines the size of a batch of magnets. Within a batch the magnet properties are similar, whereas the dipole moment variations between batches can be as large as a few per cent. Magnets from the same batch can be sorted efficiently resulting in high-performance structures. This strategy is limited to typical batch sizes of 1–2 tonnes. For larger undulators with a total weight of the magnet material of 20 tonnes and more (e.g., undulator length >100 m for FEL structures) other strategies have to be employed. One of them is a sophisticated mixing scheme of the different powder batches before pressing the blocks.

8 Microscopic properties

8.1 Microscopic structure of RE permanent magnets

Rare earth (RE) based permanent magnets are either sintered or melt spun. They consist of monocrystalline grains with a diameter of a few micrometres which are embedded in a RE-enriched matrix. In the case of Nd₂Fe₁₄B this matrix contains ingredients such as Nd, Co, Cu, Al, Ga, Dy and Nd-oxides. The crystal structures of RE-based magnets are described in detail in Ref. [19]. The unit cell of SmCo₅ is hexagonal. Sm₂Co₁₇ has a rhombohedral unit cell and Nd₂Fe₁₇B is tetragonal. The monocrystalline areas are formed when the melted phase is cooled down. They remain unchanged during the crushing and milling process and they are oriented with respect to the external field during the pressing. At typical sintering temperatures of about 850°C the matrix enclosing the monocrystalline particles melts whereas the crystallites remain unchanged. This liquid phase sintering yields rather dense products. The theoretical limit of the energy product is given by

$$(BH)_{\max} = B_r^2 / \mu$$

$$B_r(20^\circ\text{C}) = B_{r-sat}(20^\circ\text{C}) \cdot \frac{\rho}{\rho_0} \cdot (1 - V_{nonmagnetic}) \cdot f_\phi ,$$

where

$$f_\phi = \cos(\phi)$$

and

$$\phi = \arctan \left(2 \frac{B_{r-perp}}{B_{r-par}} \right) .$$

Typical values for Nd₂Fe₁₄B magnets are as follows [20]. Owing to the liquid phase sintering the density is close to the single crystal density: $\rho/\rho_0 > 0.985$. An optimized pressing process yields alignment coefficients as high as $f_\phi > 0.98$. The vacuum induction furnace and the consequent inert gas processing keep the amount of impurities (Nd oxides and others) below 2.5 weight per cent. The occurrence of other RE constituents is also below 2.5 weight per cent. Thus the amount of

nonmagnetic material is less than 5%. With these parameters an energy product of 59 MGOe has been achieved which is close to the theoretical limit of 63 MGOe.

It is worth noting that $\text{Nd}_2\text{Fe}_{14}\text{B}$ magnets are sensitive to hydrogen decrepitation: $\text{Nd} + \text{H}_2\text{O} \gg \text{NdOH} + \text{H}$, $\text{H} + \text{Nd} \gg \text{NdH}$. The hydrogen decrepitation can have fatal consequences for magnets operated in a sensitive environment (e.g., undulators of an accelerator). During magnet fabrication special care is needed to cope with these effects. Appropriate chemical constituents between the grain boundaries minimize the hydrogen decrepitation and an appropriate surface passivation or coating (Al, TiN and others) protects the magnets in a humid environment. On the other hand, the decrepitation process may be interesting for a final decomposition of permanent magnet material and a RE recovery [21, 22].

8.2 Coercivity

The potential existence of Bloch walls in a magnetic particle depends on the size of the particle. Energy considerations show that Bloch walls can not exist below a certain grain size. For example, this critical size is 0.01 μm for Fe and 1 μm for Ba ferrites. Below these limits the particles behave like a single domain. For larger grains Bloch walls may show up. The typical grain size of a RE-magnet of a few micrometres is a bit larger than the single domain size. Usually, the magnetization vector rotates in the plane of the boundary between two magnetic domains. Exceptions are thin films where the rotation occurs perpendicular to the boundary plane (Néel wall).

The coercivity can be enhanced by various methods:

- intentional addition of imperfections which impede the Bloch wall movement in large grains (e.g, carbides in steel magnets);
- preparation of single domain grains which can be switched only completely requiring high fields;
- implementation or enhancement of the magnet anisotropy (shape or crystal anisotropy).

Oriented AlNiCo 5 is an example of a magnet with shape anisotropy. When cooling down the molten components a spinodal decomposition into the magnetic phase FeCo and the less magnetic phase FeNiAl occurs. The spinodal decomposition as described by the Cahn–Hilliard equation is a diffusion process which results in a periodic and crystallographic oriented structure of the phases. The size of the segregated phases and the width of the boundaries (it is described by a tanh function) is a function of time.

When AlNiCo 5 is cooled down in the presence of a magnetic field the two phases get oriented and a strong shape anisotropy evolves [23]. As a consequence the energy product along the easy axis is about a factor of ten larger than in the perpendicular direction.

RE-based permanent magnets exhibit a large crystal anisotropy. These magnets are either of the nucleation type or the pinning type. Nucleation type magnets are SmCo_5 , $\text{Nd}_2\text{Fe}_{14}\text{B}$, and ferrites. Pinning type magnets are $\text{Sm}(\text{Co},\text{Fe},\text{Cu},\text{Hf})_7$, SmCo_5+Cu precipitation, and $\text{Sm}_2\text{Co}_{17}+\text{SmCo}_5$ precipitation of the size of domain wall thickness.

The Bloch walls of nucleation type magnets move easily within one grain and are stopped only at the grain boundaries. After heating, many domain walls exist within each grain. They are easily pushed out of the grain bulk to the boundaries with rather low fields (high initial permeability). Then, the domain walls are fixed to the grain boundaries. Once the magnet is fully magnetized, high reverse fields are needed to switch the domains. Most grains switch the magnetization in a single step without forming new Bloch walls.

Pinning type magnets have pinning centres within the grains which impede the Bloch wall motion. These can be impurities or precipitations. The two magnet types show a different initial magnetization after complete demagnetization by heating as plotted in Fig. 16.

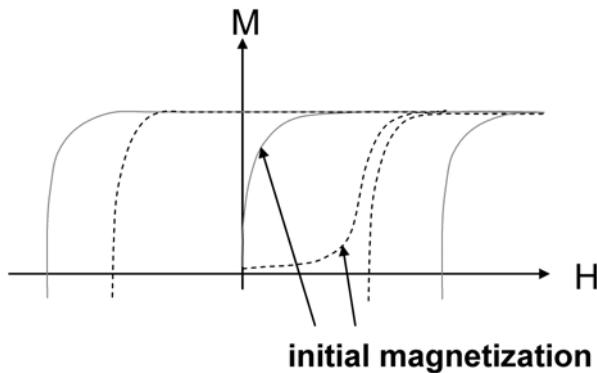


Fig. 16: Initial magnetization of nucleation type magnets (solid lines) and pinning type magnets (dashed lines)

Partial replacement of neodymium with dysprosium enhances the crystal anisotropy and, hence, the coercivity. Simultaneously, the remanence is reduced. $\text{Nd}_2\text{Fe}_{14}\text{B}$ magnet suppliers provide various grades of one material which differ in the Dy content. Depending on the specific application the appropriate material can be chosen. Dysprosium is rather expensive, so much effort has been spent in the maximization of the coercivity without using additional Dy. A correlation between the coercivity and the grain size was discussed by Mager in 1952 [24]. Systematic studies show a decrease of the coercivity with growing grain size within the range of the final grain size of 3.8 to 7.6 μm (Table 12):

Table 12: Coercivity dependence on grain size [25]

Grain size before sintering (μm)	Final grain size (μm)	H_{cj} (20°C) kA/m	H_{cj} (100°C) kA/m
1.9	3.8	1178	581
2.2	4.3	1162	573
2.6	4.9	1090	525
3.0	6.0	971	462
3.5	7.6	883	414

The experimental data can be fitted with $H_{cj}(20^\circ\text{C}) \propto (\text{final_grain_size})^{-0.44}$. The grain size grows during sintering according to $R(t) = k \cdot t^{1/n}$.

The parameter n depends on the material. For pure metals we find $n = 2-4$. In sintered $\text{Nd}_2\text{Fe}_{14}\text{B}$ the grain growth rate depends on the B concentration. It is about $n = 16-20$ for B concentrations below 5.7 at. % and decreases to $n = 7.5$ for B concentrations above 5.7 at. %. Also the RE-rich constituents have an impact on the grain growth rate. For magnets with RE-rich constituents < 4 wt. % we have $n = 30-40$. For higher fractions of RE-rich constituents n decreases to $n = 10$ [26]. The sintering time has to be adjusted appropriately, to get an optimum grain size of 3–5 μm and to avoid giant grain growth. Usually, the grain size is measured according to the standard methods as defined in ASTEM E112.

Another figure of merit during sintering is the number of corners per grain (looking in the easy axis direction). Owing to the crystal structure of Nd₂Fe₁₄B six corners indicate an unperturbed crystal structure. Hence the number of six-corner grains should be maximized [26].

The coercivity depends also on the alignment factor f_φ . With increasing f_φ the remanence grows but the coercivity diminishes (Fig. 17) [20].

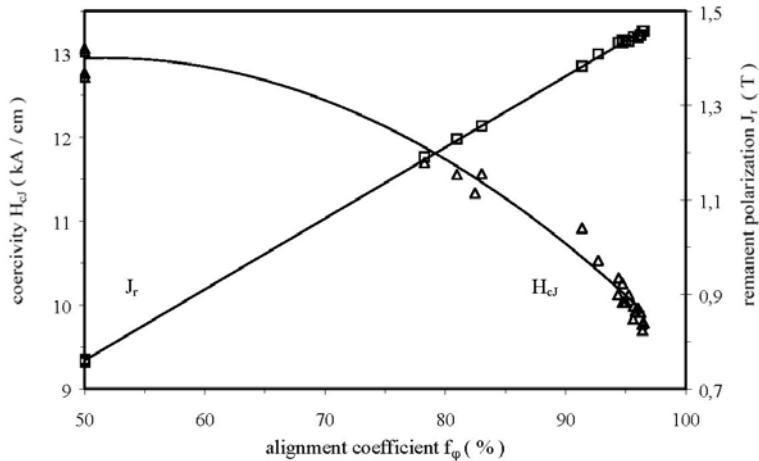


Fig. 17: Remanence and coercivity versus alignment factor [20]

The coercivity is a function of the direction of the applied external field [27]. It increases with the angle between the field and the easy axis. For angles smaller than 45° the dependence is roughly described by

$$H_{cj} \propto \frac{1}{\cos(\theta)}.$$

In certain cases this enhanced coercivity can be used in a magnetic design.

9 Observation of magnetic domains

An effective development of magnetic materials is based on detailed microscopic information of the material. A large variety of techniques is used for the observation of magnetic domains. In the following we can give only a brief overview of the most common methods and a few new methods. For more details we refer to the book by Schaefer (see bibliography).

Bitter elaborated a simple procedure to visualize magnetic domain boundaries by strewing ferromagnetic powder onto a magnetic surface. The ferromagnetic particles move to the areas of strong field gradients which are equivalent to the domain boundaries. The pictures are called Bitter patterns. Today, ferrofluids are used which are colloidal suspensions containing small ferromagnetic particles of the size of a few tens of nanometres. The resolution of this method is 100 nm. It is restricted to stationary measurements.

Fast processes can be studied by making use of various magneto-optical effects. All magneto-optical effects are described with a generalized permittivity tensor. For a cubic crystal it has the form

$$\vec{\epsilon} = \epsilon \begin{pmatrix} 1 & -iQ_v m_3 & iQ_v m_2 \\ iQ_v m_3 & 1 & -iQ_v m_1 \\ -iQ_v m_2 & iQ_v m_1 & 1 \end{pmatrix} + \begin{pmatrix} B_1 m_1^2 & B_2 m_1 m_2 & B_2 m_1 m_3 \\ B_2 m_1 m_2 & B_1 m_2^2 & B_2 m_2 m_3 \\ B_2 m_1 m_3 & B_2 m_2 m_3 & B_1 m_3^2 \end{pmatrix}$$

Similarly, a magnetic permeability tensor can be set up. The matrix elements, however, are two orders of magnitude smaller and are usually neglected. Inserting this tensor in the Fresnel equations, all of the magneto-optical effects can be described quantitatively. In magneto-optical spectroscopy a magnetic sample is irradiated with linearly polarized light. The rotation of the polarization vector in transmission geometry is called the Faraday effect and the rotation in reflection geometry is called the Kerr effect. The linear polarized light introduces vibrations of the charged particles in the sample. In the presence of a magnetic field the moving charges experience a Lorentz force. The modified vibration introduces perpendicular electric field components in the reflected or transmitted beam. Figure 18 shows the geometries of all magneto-optical effects.

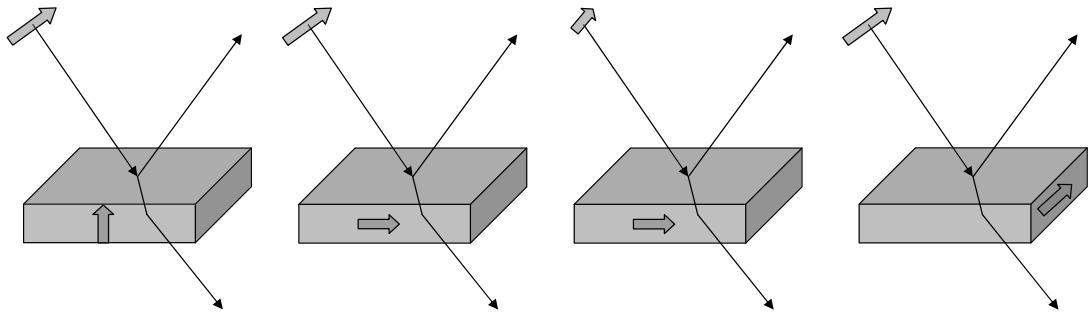


Fig. 18: Magneto-optical effects. From left to right: (i) Polar magnetization, in plane polarization: clockwise rotation of polarization in reflection and transmission. (ii) Longitudinal magnetization, in plane polarization: anticlockwise rotation of polarization in reflection and transmission. (iii) Longitudinal magnetization, polarization perpendicular to plane of incidence: clockwise rotation of polarization in reflection and anticlockwise rotation in transmission. (iv) Transverse magnetization, in plane polarization: in reflection no rotation but amplitude variation, no effect in transmission.

All these methods are not element specific. With the development of synchrotron radiation light sources the element specific investigation of magnetic samples in the soft X-ray regime has evolved to a large research field where the 2p–3d transitions of Fe, Co, Ni and the 3d–4f transitions of the rare earth metals are of particular interest. The samples are irradiated with circular polarized photons as produced by circular undulators in a storage ring. The different absorption coefficient of right and left circular polarized light is used to identify the magnetic domains (XMCD: X-ray Magnetic Circular Dichroism). Photoelectrons emitted by the sample can be used in a photoelectron microscope (PEEM) to visualize the local distribution of the domains (Fig. 19). It has to be emphasized that this type of experiment goes far beyond the usual Kerr effect measurement with visible light since the investigation of layered magnetic samples requires tuneable monochromatic light of variable polarization in the soft X-ray regime which is only available at modern synchrotron radiation light sources.

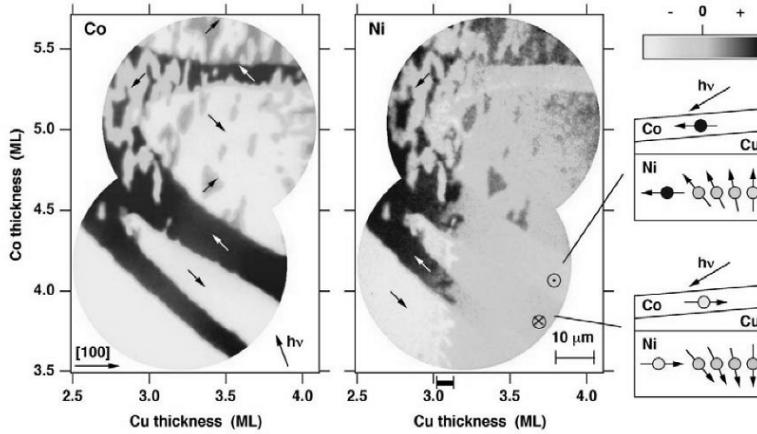


Fig. 19: Measurement of a system consisting of a Co layer on a Ni substrate separated by a wedged Cu layer making use of a photoelectron microscope at BESSY II. Here the exchange coupling between the Co and the Ni layer is studied in dependence of the thickness of the nonmagnetic Cu layer. At large distances the Ni atoms orient vertically, at small distances the Ni atoms tilt in the direction of the Co magnetic moments [28].

Magnetic domains can be imaged also with a soft X-ray holographic technique. A pinhole in the radiation cone of a circularly polarizing undulator prepares a coherent photon beam which illuminates the sample. Simultaneously, a small pinhole (100–350 nm) is irradiated with the same beam. Both transmitted beams are superimposed on a CCD camera. With the geometry of the reference pinhole known, the hologram can be inverted yielding the real structure of the magnetic domains of the sample [29]. No lenses, mirrors or zone plates are needed for this technique thus stressing the potential of this method. So far, a resolution of 50 nm has been demonstrated (Fig. 20).

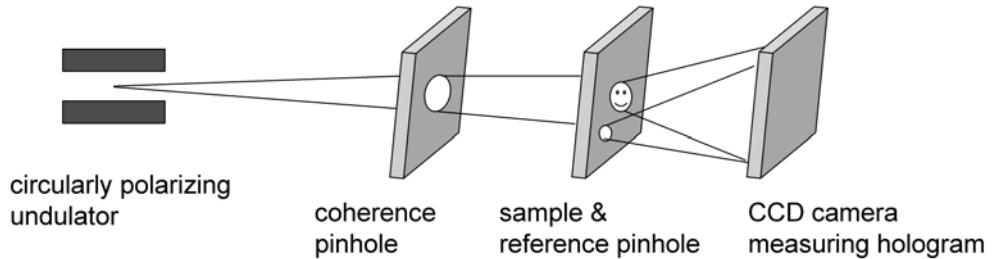


Fig. 20: Soft-X-ray holography using the circularly polarized undulator radiation. The magnetic sample is irradiated with right- and left-handed circular polarized light and the varying absorption strength depicts the orientation of the magnetization [29].

The bulk of thick magnetic samples can not be studied with photon beams on account of the limited penetration depth. Neutrons penetrate thick samples in the centimetre range. Neutron decoherence imaging is suited to detect magnetic domains of such samples [30]. The neutrons of a nuclear reactor hit a source grating. Each slit of the grating represents a coherent line source whereas the different line sources are mutually incoherent. A proper choice of the grating line density provides a coherent superposition of the fringes of all line sources in the image plane. After a few metres of free propagation the neutrons are scattered at the magnetic domains of a sample. A phase grating

behind the sample imprints a periodic phase modulation onto the wavefront. During further propagation the phase modulation transforms into a density modulation (Talbot imaging) which can be detected with a sliding detection grating in front of a detector. An undistorted neutron beam produces a periodic intensity modulation. The modulation amplitude decreases for a distorted beam and gives information on the scattering process (Fig. 21). The resolution demonstrated so far is 50–100 μm .

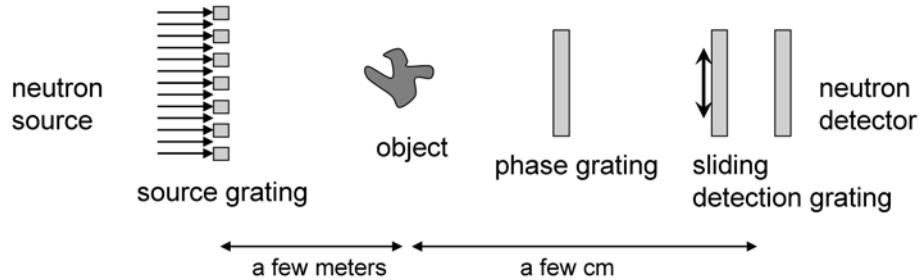


Fig. 21: Neutron decoherence imaging [30]

In plane magnetic domains thin layers can be detected with a transmission electron microscope acting as a Lorentz force microscope. Several 100 MeV electrons hit the magnetic film with magnetization vectors oriented perpendicular to the electron beam. The Lorentz force deflects the electrons and from their distribution behind the sample the magnetic domains can be reconstructed (Fig. 22). The resolution of this method is 10 nm. If the domains are oriented perpendicular to the film the sample has to be tilted to produce a deflection of the electrons. The resolution is reduced in these cases.

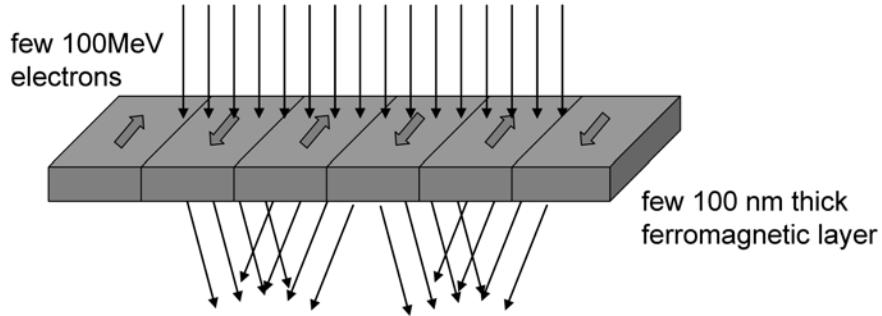


Fig. 22: Lorentz force microscope. For other domain geometries a tilting of the sample may be necessary

10 Simulation methods for permanent magnet devices

There are many codes on the market which solve Maxwell's equations in the presence of permanent magnets and current carrying wires, e.g., Refs. [31, 32]. In the following we concentrate on an algorithm which is very efficient for pure permanent magnet structures but also applicable to geometries including iron though it is significantly slower in these cases. The algorithm is widely used for the design of undulators and wigglers. The algorithm is implemented in the code RADIA [33, 34] which is freely available from the ESRF [35].

10.1 3D fields

Magnetic fields of pure permanent magnet structures can be simulated with an accuracy of a few per cent using the current sheet or charge sheet equivalent model (CSEM). These methods assume a permeability of one, hence the fields of individual blocks can be linearly superimposed. A magnet block is represented by current sheets at the magnet surfaces parallel to the easy axis or charge sheets at the surface perpendicular to the easy axis. The current sheets or charge sheets are assumed to be infinitely thin (Fig. 23).

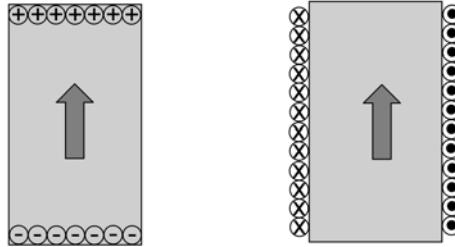


Fig. 23: Charge sheet (left) and current sheet (right) equivalent method

The magnetic induction of a current carrying wire is evaluated from the equation of Biot–Savart:

$$\vec{B}(\vec{r}_0) = \frac{1}{c} \int I d\vec{l} \times \frac{\vec{r}_0 - \vec{r}'}{|\vec{r}_0 - \vec{r}'|^3}.$$

In analogy in the current sheet equivalent method the magnetic induction is derived via integration over the surface current density $\vec{j}_M = \vec{M} \times \vec{n}$:

$$\vec{B}(\vec{r}_0) = \int (\bar{\nabla} \times \vec{M}(r')) \times \frac{\vec{r}_0 - \vec{r}'}{|\vec{r}_0 - \vec{r}'|^3} dV' = \oint_{\text{surface}} \frac{\vec{j}_M}{c} \times \frac{\vec{r}_0 - \vec{r}'}{|\vec{r}_0 - \vec{r}'|^3} dS'.$$

Following the charge sheet equivalent model a scalar potential is established by integrating over the surface charge density $\sigma_M = \vec{n} \cdot \vec{M}(\vec{r}')$ at the pole faces. The derivative of the scalar potential yields the magnetic field.

$$\Phi(\vec{r}_0) = - \int \frac{\bar{\nabla}' \cdot \vec{M}(r')}{|\vec{r}_0 - \vec{r}'|} dV' = \oint_{\text{surface}} \frac{\sigma_M \cdot dS'}{|\vec{r}_0 - \vec{r}'|}$$

$$\vec{H}(\vec{r}_0) = -\text{grad}(\Phi(\vec{r}_0))$$

For rectangular blocks and a homogeneous magnetization the integration can be done analytically. In this specific case the magnetic induction is given by

$$\vec{B}(\vec{r}_0) = \bar{\bar{Q}}(\vec{r}_0) \cdot \vec{M}$$

$$Q_{xx} = \sum_{ijk=1}^2 (-1)^{i+j+k+1} \arctan \left(\frac{y_j z_k}{x_i \sqrt{x_i^2 + y_j^2 + z_k^2}} \right)$$

$$Q_{xy} = \ln \left(\prod_{ijk=1}^2 \left(z_k + \sqrt{x_i^2 + y_j^2 + z_k^2} \right)^{(-1)^{i+j+k}} \right)$$

$$x_{1,2}(y_{1,2}, z_{1,2}) = x_c(y_c, z_c) - x_0(y_0, z_0) \pm w_{x(y,z)} / 2$$

and similar for the other Q_{ij} . The parameters x_c, y_c, z_c define the centre of the magnet, w_x, w_y, w_z are the magnet dimensions, and x_0, y_0, z_0 are the coordinates of the observation point.

Similarly, the magnetic induction B and the field integral of any (plane) polygon can be evaluated [33, 34]. Based on these expressions the magnetic induction and field integrals of an arbitrary polyhedron can be evaluated:

$$\begin{aligned}\vec{B}(\vec{r}_0) &= \overline{\overline{Q}}(\vec{r}_0) \cdot \vec{M} \\ \overline{\overline{Q}}(\vec{r}_0) &= \iint_{\text{surface}} \frac{(\vec{r}_0 - \vec{r}') \otimes \vec{n}'_{\text{surface}}}{|\vec{r}_0 - \vec{r}'|^3} d\vec{r}' \\ \vec{I}(r_0, \vec{v}) &= \int_{-\infty}^{\infty} \vec{B}(\vec{r}_0 + \vec{v}) dl = \overline{\overline{G}}(\vec{r}_0, \vec{v}) \cdot \vec{M} \\ \overline{\overline{G}}(\vec{r}_0, \vec{v}) &= 2 \cdot \iint_{\text{surface}} \frac{[(\vec{r}' - \vec{r}_0) \times \vec{v}] \times \vec{v} \otimes \vec{n}'_{\text{surface}}}{|(\vec{r}' - \vec{r}_0) \times \vec{v}|^2} d\vec{r}'\end{aligned}$$

Q and G are 3×3 matrices describing the geometric shape of the magnetic cell, \otimes denotes a dyadic product and \vec{v} is the integration direction. These equations are based on the assumption of a permeability of $\mu = 1$. Fields evaluated under this assumption are a few per cent higher than in reality. A higher accuracy is achieved with a realistic susceptibility $\chi \neq 0$ with $\mu = 1 + \chi$. The permeability of RE based permanent magnets is higher than one and it is different when parallel to or perpendicular to the easy axis. Typical values are $\mu_{\text{par}} = 1.06$, $\mu_{\text{perp}} = 1.17$. The values depend slightly on the fabrication procedure and the magnet grade. For example $\mu_{\text{par}} = 1.05$ for axially pressed and $\mu_{\text{par}} = 1.03$ for isostatic pressed magnets. μ_{par} shows no correlation with coercivity whereas μ_{perp} decreases with increasing coercivity (1.17 for $H_{cj} = 18$ kOe and 1.12 for $H_{cj} = 32$ kOe [27]).

To include a realistic permeability, iterative simulation strategies must be employed. The magnet configuration has to be segmented into individual cells where the cell size depends on the desired accuracy. Then, the geometry factors Q and G for the cells are evaluated. In a first run the magnetic induction and magnetic field H_i at the centre of each cell is evaluated assuming a magnetization of $M_0 = M(H_i = 0)$ in each cell. In the following iterations the parallel and perpendicular magnetization is corrected and new values for the magnetic induction and magnetic field are derived:

$$\begin{aligned}\vec{B}_i &= \sum_{\substack{k=1 \\ k \neq i}}^N \overline{\overline{Q}}_{k,i} \cdot \vec{M}_k + \overline{\overline{Q}}_{ii} \cdot \vec{M}_i \\ \vec{H}_i &= \vec{B}_i - 4\pi \cdot \vec{M}_i \\ M_{i-\text{par}} &= \frac{1}{4\pi} B_r + (\mu_{\text{par}} - 1) \cdot H_{i-\text{par}} \\ M_{i-\text{perp}} &= (\mu_{\text{perp}} - 1) \cdot H_{i-\text{perp}}\end{aligned}$$

A linear dependence of the magnetization on the magnetic field is assumed:

$$\begin{aligned}M_{\text{par}}(H_{\text{par}}) &= M_r + \chi_{\text{par}} H_{\text{par}} \\ M_{\text{perp}}(H_{\text{perp}}) &= \chi_{\text{perp}} H_{\text{perp}}\end{aligned}$$

For reverse fields H approaching H_{cj} the magnetization does not change linearly anymore with H and irreversible losses may occur. These cases can be simulated with a non-linear approach for the

magnetization which includes also the temperature [36]. The temperature dependence of M_r , H_{cj} and χ_{perp} can be parametrized as

$$\begin{aligned} M_r(T) &= M_r(T_0) \cdot (1 + a_1(T - T_0) + a_2(T - T_0)^2 + \dots) \\ H_{cj}(T) &= H_{cj}(T_0) \cdot (1 + b_1(T - T_0) + b_2(T - T_0)^2 + \dots) \\ \chi_{perp}(T) &= \chi_{perp}(T_0) \cdot (1 + a_1(T - T_0) + a_2(T - T_0)^2 + \dots) \end{aligned}$$

where a_i and b_i are extracted from the data sheet of the magnet supplier. The non-linear behaviour of the magnetization for a given temperature T is expressed by

$$M(H, T) = \alpha(T) \sum_{i=1}^3 M_{si} \tanh\left(\frac{\chi_i}{M_{si}}(H - H_{cj}(T))\right).$$

M_{si} , χ_i are derived from a fit of the $M(H)$ curve from the magnet supplier at $T = T_0$ and $\alpha(T)$ is determined from

$$M(H = 0, T) = M_r(T).$$

This model has been tested at a real undulator structure and excellent agreement between simulation and measurement has been found [36].

It is worth noting that the operation of a permanent magnet in the third quadrant (still above the knee) does not imply that the magnet does not contribute to the field.

10.2 Complex notation of 2D field

The magnet geometry can be approximated with a two-dimensional model if the magnet is long compared to the end sections. Then, it is convenient to express the magnetic induction in complex notation:

$$\begin{aligned} \vec{B}^*(\vec{z}_0) &= B_x - iB_y \\ \vec{z}_0 &= x_0 + iy_0 = r_0 \cdot e^{i\phi_0} \end{aligned}$$

The complex conjugate \vec{B}^* is used instead of \vec{B} because \vec{B}^* is an analytic function whereas \vec{B} is not. Extremely useful tools for a magnet field optimization such as conformal mapping can be applied in this case (see next section). In complex notation the field of a current flowing into the plane is given by

$$\vec{B}^*(\vec{z}_0) = a \int \frac{\vec{j}_z}{\vec{z}_0 - \vec{z}} \cdot dx \cdot dy.$$

Any iron-free permanent magnet distribution can be expressed as [37]:

$$\begin{aligned} \vec{B}^*(\vec{z}_0) &= b \int \frac{\vec{B}_r}{(\vec{z}_0 - \vec{z})^2} \cdot dx \cdot dy, \\ \vec{B}_r &= B_{rx} + iB_{ry}. \end{aligned}$$

This equation expresses an important rule: If the easy axis vector of all magnets of a complex two-dimensional configuration is rotated by the same angle ($+i\phi$) this adds a factor of $e^{i\phi}$ on the right-hand side. In consequence, the total field vector at a given point rotates by $(-\phi)$: This behaviour is known as the **easy axis rotation theorem**.

Any source-free two-dimensional field distribution can be expanded in terms of multipoles. Using the complex notation we have a compact form with a_n and b_n being the regular and skew multipoles.

$$\vec{B}^*(\vec{z}) = \sum_{n=1}^{\infty} (a_n - ib_n) \vec{z}^{n-1} .$$

10.3 Conformal mapping

The technique of conformal mapping can be extremely useful in solving Maxwell's equations in complicated 2D-magnet arrangements [38–40]. Using a conformal map the complete configuration is transformed to a simpler geometry where Maxwell's equations are solved. Then, the results are transformed back to the old variables. One might argue that the technique of conformal mapping is no longer necessary when high performance multiprocessor computers are cheap and available everywhere. This is, however, only half the truth. For many problems the solution in transformed geometry provides higher accuracy for less effort, and, even more important, the conformal mapping technique provides a deeper insight into the properties of magnet systems. This understanding helps in a fast and efficient design.

Conformal maps preserve both angles and shapes of infinitesimal small figures, whereas the size of these figures usually changes. The intersection angle between any two curves in the original and the transformed geometry is identical.

Any analytic function represents a conformal map in the regions of non-zero derivatives. Let $F = \vec{w}(\vec{z})$ be a complex function with $\vec{w} = u(x, y) + iv(x, y)$ where u and v are real functions. F is analytic if and only if the Cauchy–Riemann (C–R) relations are fulfilled:

$$\begin{aligned}\frac{\partial u}{\partial x} &= \frac{\partial v}{\partial y} \\ \frac{\partial v}{\partial x} &= -\frac{\partial u}{\partial y}\end{aligned}$$

u and v also obey the Laplace equation which is proven by differentiating the C–R equations:

$$\begin{aligned}\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} &= 0 \\ \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} &= 0\end{aligned}$$

Within a source-free region (no currents, no magnetic charges) the complex magnetic field strength $\vec{H}^* = H_x - iH_y$ satisfies the C–R relations which are identical to Maxwell's equations in two dimensions. \vec{H}^* is an analytic function and an integration yields the complex potential F :

$$\begin{aligned}F &= A(x, y) + iV(x, y) \\ \vec{H}^*(\vec{z}) &= i \frac{\partial F}{\partial \vec{z}}\end{aligned}$$

$V(x, y)$ is the scalar potential and $A(x, y)$ the z component of the vector potential in the usual sense, where A_y and A_x can be set to zero without losing generality (2-dimensional fields). As an example we will derive the conformal map that transforms a multipole into dipole geometry: For F being the complex potential of \vec{H}^* we have

$$H^*(\vec{z}) = i \frac{\partial F}{\partial \vec{z}} = i \frac{\partial F}{\partial \vec{w}} \frac{\partial \vec{w}}{\partial \vec{z}} = H^*(\vec{w}) \cdot \frac{\partial \vec{w}}{\partial \vec{z}} .$$

The field strength of an ideal regular octupole is given by $\vec{H}^*(\vec{z}) = o \cdot \vec{z}^3$ whereas a dipole is described by $\vec{H}^*(\vec{w}) = \text{const}$. Thus we have $\vec{w}' = (o \cdot \vec{z}^3) / \text{const}$ and the conformal map that transforms the octupole to dipole geometry is given by $\vec{w}(\vec{z}) = (o \cdot \vec{z}^4) / (4 \cdot \text{const})$.

11 Permanent magnet multipoles

Accelerators are built of dipole magnets for the deflection of the relativistic particles, quadrupoles for a strong focusing of the electron beam, and sextupoles and higher order multipoles which cope with non-linear effects. Usually, the magnets are constructed as electromagnets since they provide flexibility in terms of the machine parameters. There are, however, cases where the flexibility is not needed and permanent magnets are the better choice. In the previous section we presented the fields of ideal multipoles. Now, we will discuss the layout of permanent magnet multipoles.

11.1 Fixed-strength multipoles

Halbach proposed segmented permanent magnet multipoles which approximate ideal multipoles to an accuracy limited by the number of magnets M to be used per period [37]:

$$\begin{aligned} \vec{B}^*(\vec{z}) &= \vec{B}_r \sum_{n=0}^{\infty} \left(\frac{\vec{z}}{r_1} \right)^{n-1} \frac{n}{n-1} \left(1 - \left(\frac{r_1}{r_2} \right)^{n-1} \right) K_n \\ K_n &= \cos^n(\varepsilon\pi/M) \frac{\sin(n\varepsilon\pi/M)}{n\pi/M} \\ n &= N + vM \\ \frac{n}{n-1} \left(1 - \left(\frac{r_1}{r_2} \right)^{n-1} \right)_{n=1} &= \ln(r_2/r_1) \end{aligned} .$$

Here, N is the order of the multipole ($N = 1$: dipole, $N = 2$: quadrupole etc.), v is the harmonic where $v = 0$ is the fundamental, r_1 and r_2 are the inner and outer radius, and ε is the stacking factor. $A = (N+1)2\pi/M$ is the relative angle of the easy axes of neighbouring magnet segments. A few examples are given in Fig. 24.

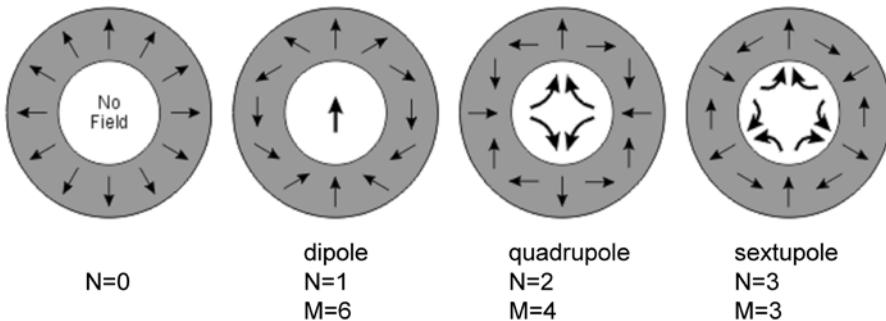


Fig. 24: Halbach-type multipoles. From left to right: 1: $N = 0$, at the centre of the disk the field is zero according to the easy axis rotation theorem. 2: Dipole ($N = 1, M = 6$). 3: Quadrupole ($N = 2, M = 4$). 4: Sextupole ($N = 3, M = 3$).

Halbach-type permanent magnet quadrupoles ($M = 6$) with gradients of 500 T/mm in a 6 mm, 20 mm inner/outer diameter have been built [41, 42]. The higher order content of the quadrupoles could be significantly reduced by adjusting the individual magnet segments. Halbach multipoles can be improved with respect to the peak field when a fraction of the permanent magnets at the pole tips is replaced with soft iron (Fig. 25). A field gradient of 300 T/mm in a 14 mm inner diameter has been realized [43].

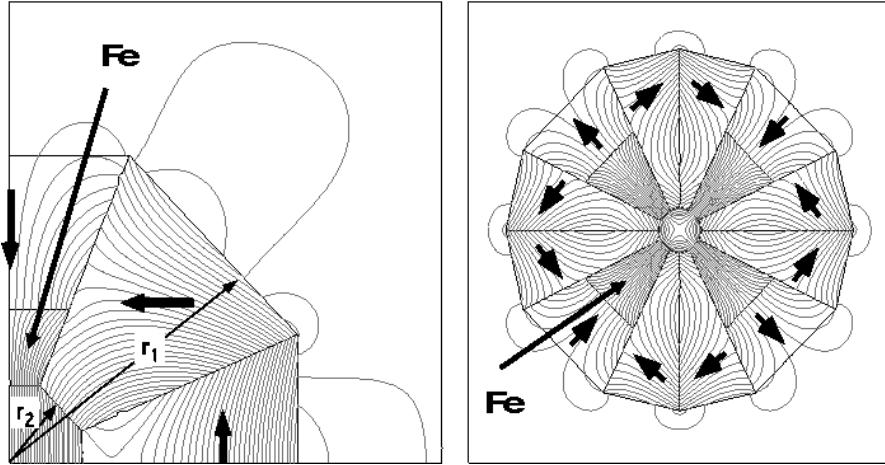


Fig. 25: Left: Modified Halbach dipole with iron pole tips. Right: Modified Halbach quadrupole (300 T/mm in the 14/100 mm inner/outer diameter) [43].

11.2 Variable strength multipoles

Permanent magnet quadrupoles are needed if the geometric constraints do not permit the use of bulky electromagnets such as at the final focus section of the ILC. The quadrupoles have to be placed close to the interaction point but must not block the other beam. A certain tuning range of the focusing strength is desirable which requires a specific design. Gluckstern proposed a continuously adjustable quadrupole consisting of five individual discs, each of them individually rotatable [44]. The linear effect of a rotated quadrupole is given by a symplectic 4×4 matrix with

$$M^T \Phi M = \Phi = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}$$

The off-diagonal elements characterize the coupling between the horizontal and vertical plane. With five individual discs with three rotation angles α and three quadrupole strength q (thicknesses) the coupling can be compensated and an arbitrary focusing can be realized: disc 1: α_1, q_1 ; disc 2: α_2, q_2 ; disc 3: α_3, q_3 ; disc 4: α_4, q_4 ; disc 5: α_5, q_5 (Gluckstern singlet, Fig. 26, left). Prototypes with inner/outer diameters of 12/36 mm, respectively, and a gradient of 140 T/m have been built [45, 46].

Another design is the binary stepwise permanent magnet quadrupole [47] consisting of two layers of quadrupoles. The inner quadrupole is fixed in strength. The outer quadrupole is made of a series of rings with relative thicknesses of powers of two. Thus, a discrete (bitwise) adjustment is possible via rotating specific rings by multiples of $\pi/2$. The resolution is limited by the strength of the thinnest ring (Fig. 26, right).

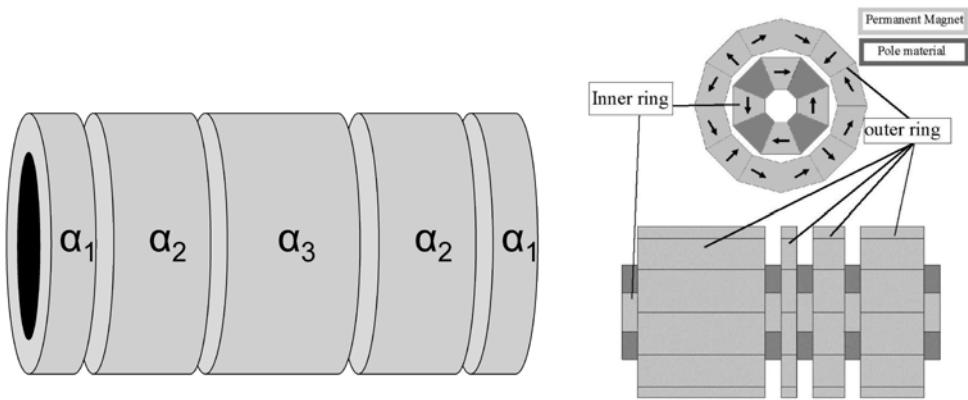


Fig. 26: Gluckstern singlet for continuous adjustment of quadrupole strength (left) [44] and binary stepwise adjustable quadrupole (right) [47]

12 Permanent magnet undulators and wigglers

Undulators and wigglers are periodic magnetic structures which force the electrons on a sinusoidal path. High field wigglers are used for electron beam damping to achieve low emittance in ultimate storage rings such as PETRA III [48] or NSLS II [49]. The main purpose of these devices, however, is the production of synchrotron radiation from the infrared to the hard X-ray region. First-generation synchrotron radiation light sources were storage rings which were shared between high-energy physics and synchrotron radiation users. Second-generation light sources were dedicated machines for the synchrotron radiation user community. Undulators and wigglers were still rarely used and bending magnets were the normal radiation sources. Third-generation synchrotron radiation light sources are explicitly built for the use of undulators and wigglers. The lattices have a high symmetry with several tens of straight sections to adapt as many undulators or wigglers as possible.

The brightness of photon beams as emitted by typical storage ring undulators is three to four orders of magnitude higher than that of dipole radiation (see Section 13). The photon energy is related to the undulator period length, and short-wavelength photons require short periods which can not be realized with electromagnets. Usually, permanent magnet undulators are chosen.

In 1973 Mallinson published a magnet design under the title: '*One-sided fluxes - A magnetic curiosity?*'. The article describes a linear array of magnets where the easy axis orientation is rotated between succeeding magnets. This array has a remarkable characteristic: At one side a significant amount of flux leaves the array whereas much less flux is detected at the opposite side. The high flux side is determined by the sense of rotation of the magnetization.

Halbach recognized the potential of this effect. He combined two of these arrays facing the high flux sides to each other. This array is called a pure permanent magnet undulator or Halbach I undulator [50] (Fig. 27).

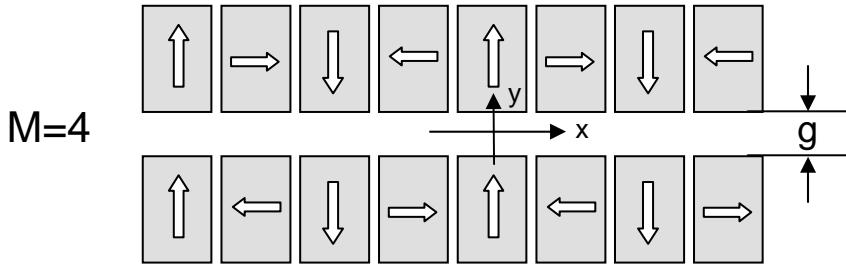


Fig. 27: Pure permanent magnet undulator (Halbach I)

The two-dimensional field of such an array can be expressed analytically:

$$\vec{B}^*(\vec{z}) = i2\vec{B}_r \sum_{v=0}^{\infty} \cos(nk\vec{z}) \cdot e^{-nkg/2} \cdot (1 - e^{-nkL}) \cdot \frac{\sin(n\varepsilon\pi/M)}{n\pi/M}$$

$$\vec{z} = x + i \cdot y$$

$$n = 1 + v \cdot M$$

$$k = 2\pi / \lambda_0$$

Here λ_0 is the period length, ε is the filling factor, g is the distance of the arrays (magnetic gap), M is the number of magnets per period, and n is the field harmonic. Note that the higher order field harmonic content is related to M . In most cases, four magnets per period are used. A larger M enhances the field only by a few per cent. Higher fields can be produced with hybrid undulators, so-called Halbach II devices [51], where soft iron material is used to concentrate the flux (Fig. 28). For high performance applications the iron poles can be replaced by CoFe pieces with a saturation magnetization of 2.4 T to boost the peak magnetic field further.

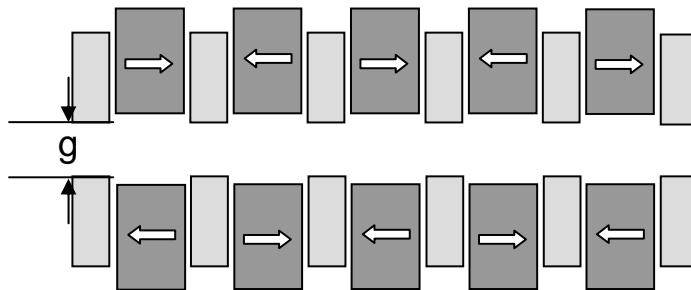


Fig. 28: Hybrid undulator (Halbach II)

There is no analytic field expression for a hybrid undulator and the field has to be parametrized based on numerical field simulations:

$$B_y \approx a \cdot \exp\left(-b \cdot \frac{g}{\lambda_0} + c \cdot \left(\frac{g}{\lambda_0}\right)^2\right).$$

For a hybrid undulator Elleaume derives the parameters $a = 3.69$, $b = 5.07$, $c = 1.52$ [52]. Parameters for other undulator designs are also presented in Ref. [52].

13 Spectral properties of undulators and wigglers

13.1 Ideal sources

The spectral properties of undulators and wigglers are described only briefly. For details we refer you to the books and CAS proceedings listed in the bibliography. The electromagnetic properties of relativistic particles are given by the Lienard–Wiechert potentials which are the solution of the inhomogeneous Maxwell equations (Gaussian units):

$$\Phi(\vec{x}, t) = \left[\frac{e}{(1 - \vec{\beta} \cdot \vec{n})R} \right]_{ret}$$

$$\vec{A}(\vec{x}, t) = \left[\frac{e\vec{\beta}}{(1 - \vec{\beta} \cdot \vec{n})R} \right]_{ret}$$

$\vec{\beta}(t_{ret})$ is the particle trajectory at the retarded time $t_{ret} = t - R(t_{ret})/c$, $\vec{n}(t_{ret})$ is the vector pointing from the particle to the observer, and $R(t_{ret})$ is the distance between the particle and the observer. The acceleration and velocity terms of the electric field are derived from these potentials and the pointing vector \vec{S} describes the emitted power:

$$\vec{E}^{acc}(t) = \frac{e}{c} \cdot \left[\left(\vec{n} \times \left[(\vec{n} - \vec{\beta}) \times \dot{\vec{\beta}} \right] \right) / \left(R \cdot (1 - \vec{\beta} \cdot \vec{n})^3 \right) \right]_{ret}$$

$$\vec{E}^{vel}(t) = e \cdot \left[(\vec{n} - \vec{\beta}) / \left(\gamma^2 \cdot R^2 \cdot (1 - \vec{\beta} \cdot \vec{n})^3 \right) \right]_{ret}$$

$$\vec{B} = \left[\vec{n} \times \vec{E} \right]_{ret}$$

$$\vec{S} = \frac{c}{4\pi} \vec{E} \times \vec{B}$$

γ is the Lorentz factor. A Fourier transform and a far field approximation deliver the spectral content of the radiation:

$$\frac{\partial^2 I}{\partial \omega \partial \Omega} = \frac{e^2}{4\pi^2 c} \left| \int_{-\infty}^{\infty} \left[\left(\vec{n} \times \left((\vec{n} - \vec{\beta}) \times \dot{\vec{\beta}} \right) \right) / (1 - \vec{\beta} \cdot \vec{n})^2 \right] e^{i\omega(t - \vec{n} \cdot \vec{r})} dt \right|^2$$

In a vertical dipole field a charged particle describes a horizontal circle with a radius ρ . The radiation is given by the analytic expression

$$\frac{\partial^2 I}{\partial \omega \partial \Omega} = \frac{3e^2}{4\pi^2 c} y^2 \gamma^2 (1 + (\theta\gamma)^2)^2 \left[K_{2/3}(\xi)^2 + \left((\theta\gamma)^2 / (1 + (\theta\gamma)^2) \right) \cdot K_{1/3}(\xi)^2 \right]$$

$$\xi = \frac{y}{2} (1 + (\gamma\theta_y)^2)^{3/2}$$

$$y = \omega / \omega_c$$

$$\omega_c = (3\gamma^3 c) / (2\rho)$$

The first and the second part in rectangular brackets describe the contributions from the horizontal and vertical electric field component, respectively. In plane the radiation is horizontally polarized. Off plane a circular component shows up which is opposite in sign above and below the midplane.

Wiggler and undulators are periodic magnetic structures where the particles employ an oscillating trajectory. The devices are characterized by the undulator parameter $K = 93.4 \cdot \lambda_0 \cdot B_0$ where λ_0 and B_0 are the period length and the maximum field, respectively. Devices with $K \gg 1$ are called wiggler. The emitted radiation can be evaluated as a sum of dipole spectra emitted at each individual pole. In undulators the K parameter and hence amplitudes of the oscillation are smaller and the radiation beams emitted at each of the poles overlap coherently. In planar undulators constructive interference is observed at the odd harmonics ($n = 1, 3, 5\dots$):

$$\lambda_n = \frac{\lambda_0}{n \cdot 2\gamma^2} \left(1 + K^2 / 2 + \gamma^2 \theta^2 \right)$$

Neglecting end pole effects, the undulator radiation in the far field close to the harmonics can be approximated with

$$\frac{\partial^2 I}{\partial \omega \partial \Omega} = \frac{e^2 \gamma^2 N^2}{c} \cdot F_n(K_x, K_y, \gamma\theta, \gamma\Phi) \cdot \frac{\sin^2(N\pi \cdot \Delta\omega / \omega_l(\theta))}{N^2 \sin^2(\pi \cdot \Delta\omega / \omega_l(\theta))}$$

where F is an infinite sum over Bessel functions.

13.2 Real sources

Wiggler have a low brightness and today they are used only for the production of high-energy photons which are not accessible with undulators. Wiggler are insensitive to field errors since the dipole spectra are spatially and spectrally broad and coherence effects can be ignored.

In contrast, undulators are based on a coherent overlap of the radiation contributions from the whole undulator. The spectral performance of an undulator is characterized by the phase error which describes the jitter in time between the electron beam and the emitted light:

$$\Delta\Phi = \frac{2\pi}{\beta\lambda(B\rho)^2} \cdot \int_0^z \left[\int_0^{z'} B_y^{fit} dz'' \cdot \int_0^{z'} B_y^{res} dz'' \right] \cdot dz' + 0.5 \cdot \frac{2\pi}{\beta\lambda(B\rho)^2} \cdot \int_0^z \left[\int_0^{z'} B_y^{res} dz'' \cdot \int_0^{z'} B_y^{res} dz'' \right] \cdot dz' .$$

Field errors produce phase errors which reduce the on-axis flux density of the odd harmonics approximately by $R = ((1 - \exp(-\sigma_\phi^2))M + \exp(-\sigma_\phi^2)M^2)/M^2$ where M is the number of poles [53]. Though the magnet quality of commercially available blocks has improved a lot during the last few years, it is still impossible to build high-quality permanent magnet undulators without applying specific techniques such as magnet sorting and undulator shimming (see Section 15). Phase errors below 2° can be achieved for devices with up to 100 periods. This limits the brightness reduction in the 15th harmonic to values well below 20%. Thus magnet field errors can be efficiently compensated and today the spectral performance of an undulator is only limited by the electron beam parameters, emittance, and energy spread.

14 Undulator and wiggler designs

There is a large variety of undulator and wiggler designs optimized for specific purposes. The device parameters can be grouped in four categories: tuning range, polarization, spectral purity, and on-axis-power density.

14.1 Tuning range: out-of-vacuum versus in-vacuum devices

In-vacuum undulators permit smaller period lengths than conventional devices for a fixed vertical vacuum aperture and K parameter and, therefore, they provide higher photon energies. Planar in-

vacuum devices rely on a mature technology and today many in-vacuum undulators are in operation all over the world [54].

As already mentioned, the negative temperature coefficients of the remanence and the coercivity of $\text{Nd}_2\text{Fe}_{14}\text{B}$ favour an operation at lower temperatures. Between 300 K and 150 K the remanence increases by 16%. The coercivity grows even more and thus another magnet grade with higher remanence and less coercivity can be used since the magnets gain stability at lower temperatures. Tanaka et al. proposed a cryogenic undulator to be operated at 150 K [55, 56]. Meanwhile, a cryogenic undulator is in operation at the ESRF [57–60], a 2 m device to be installed at the PSI [61] has been built at SPRING-8, and more devices are under consideration.

At 150 K $\text{Nd}_2\text{Fe}_{14}\text{B}$ features a spin reorientation [62, 63] and the remanence decreases below this temperature. By replacing the $\text{Nd}_2\text{Fe}_{14}\text{B}$ magnets with Pr-Fe-B magnets even lower temperatures can be used [55]. Temperatures below 80 K are of particular interest because textured dysprosium can be used as pole material which has an even higher saturation magnetization than CoFe. Furthermore, the temperature sensitivity of the magnetic performance decreases at lower temperatures. Recently, VAC in collaboration with Helmholtz-Zentrum Berlin (HZB) and Ludwig-Maximilian-Universitaet Muenchen (LMU) developed a new $(\text{Pr},\text{Nd})_2\text{Fe}_{14}\text{B}$ grade with an energy product of 65.3 MGoe at 85 K [64] (Fig. 29). A $(\text{Pr},\text{Nd})_2\text{Fe}_{14}\text{B}$ based undulator is currently under construction at HZB [65] and a short prototype Pr-Fe-B undulator has been measured recently in a vertical bath cryostat at the NSLS [66]. Tanaka et al. proposed to place passive high temperature superconducting loops around the pole tips of a cryogenic undulator. Closing the gap to 0 mm at a high temperature, cooling down and opening the gap again to the operation gap induces permanent currents in the loops which enhance the peak field further [67].

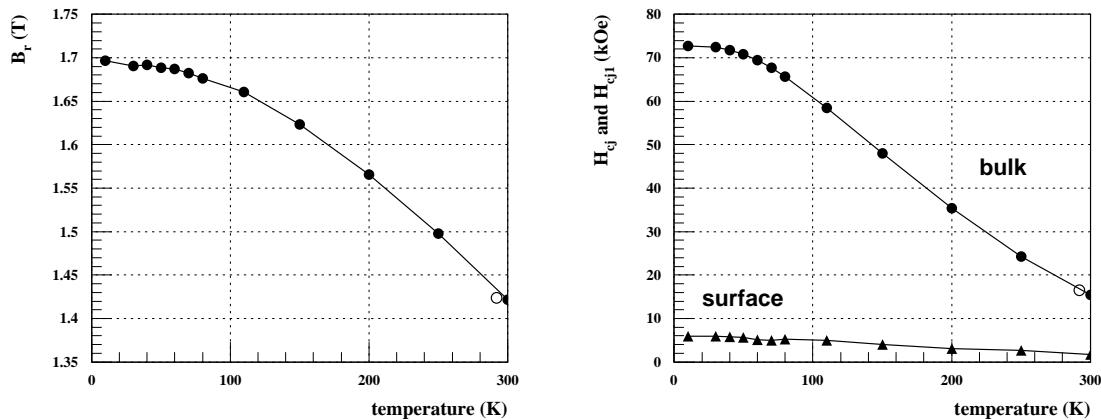


Fig. 29: Performance of $(\text{Pr},\text{Nd})_2\text{Fe}_{14}\text{B}$ material [64]

14.2 Variable polarization

The first wiggler for the production of linearly and elliptically polarized light, an asymmetric wiggler, was proposed in 1987 [68]. Positive and negative poles have different field amplitudes where the field integrals are compensated to zero. In the energy regime above the cut-off frequency of the weaker poles circularly polarized light is observed off plane just as in a bending magnet. The brightness of such sources is low due to the depth-of-field effect which increases with the angle of observation. However, high photon energies can be reached which are inaccessible with undulators. Asymmetric wigglers are being operated at several facilities [69–72]. At the ESRF an asymmetric wiggler with a peak field exceeding 3 T is in operation [73].

The brightness of an elliptical wiggler is significantly higher than that of an asymmetric wiggler since the on-axis beam is used [74]. Elliptical wigglers consist of individual magnet arrays for the horizontal and vertical magnetic fields where one of the field components is significantly larger than the other one. The electrons move on an elliptical trajectory producing elliptically polarized light on-axis. The helicity can be switched by moving one of the arrays longitudinally. Elliptical wigglers with variable helicity have been installed at the accumulator ring of the TRISTAN ring, at the Photon Factory [75, 76], and at SPRING-8 [77, 78].

In third-generation storage rings helical undulators are favoured on account of their higher degree of circular polarization and the higher brightness. Various helical undulator designs have been developed and implemented. The APPLE II design [79] provides the highest fields among all helical undulators and APPLE II undulators have become the work-horse in many light sources. The magnet arrays are split longitudinally and the individual magnet rows can be moved independently. In the elliptical mode (shifting two rows in the same direction) the horizontal and vertical fields are 90° out of phase whereas in the inclined mode (shifting two rows in opposite direction) they are in phase producing linearly polarized light at arbitrary angles. By moving three rows, any state of polarization can be produced and with an appropriate undulator setting polarizing effects of the beamline optical components in particular below 100 eV can be compensated [80, 81].

In the ESRF design [82] vertical and horizontal fields are produced by the upper and lower magnet girder, respectively. Any state of polarization can be realized and the good field region is large. However, the fields are lower than those of an APPLE II, the two beams have to be moved individually and the electron beam is steered vertically. The ELETTRA design [83] allows for a combined movement of both beams with one motor. This undulator provides only circularly polarized light (independent of gap setting), the fields are lower than in an APPLE II and the good field region is small. The SPRING-8 design [84] employs three magnet rows, the central one producing vertical fields and the side rows the horizontal fields. The good field region is large but the field levels are even lower than in the other designs.

14.3 Spectral purity: undulator periodicity

Usually, undulators have a periodic structure to achieve highest brightness at the fundamental and at higher harmonics. In the low-energy regime higher undulator harmonics may spoil the spectrum because they are efficiently transmitted through the monochromator together with the first harmonic. Helical devices produce only the first harmonic and higher harmonics are suppressed with the use of an appropriate pinhole. If the higher harmonics are needed and the spectral purity is essential, quasi-periodic structures as proposed by Sasaki et al. can help. The quasiperiodicity z_m is derived from the projection of the grid points of a quadratic [85, 86] or rectangular lattice [87] with a grid parameter ratio of r onto a tilted straight line with a tilt angle = α :

$$Z_m = \frac{d}{r \cdot \tan(\alpha)} \left(m + (r \cdot \tan(\alpha) - 1) \cdot \left\lfloor \frac{\tan(\alpha)}{r + \tan(\alpha)} m + 1 \right\rfloor \right).$$

In the first design (Halbach I undulator with $M = 2$) the quasiperiodicity was realized with two different air gaps between the poles [88]. In a Halbach I $M = 4$ design the quasiperiodicity can also be introduced by completely removing longitudinally magnetized blocks, replacing these blocks with blocks of a reduced height [89, 90] or by retracting them [91]. Specific higher harmonics can be maximized with an appropriate choice of the amount of magnet retraction or height reduction. A successful suppression of higher harmonics has been demonstrated by measurements [92, 93]. A quasiperiodic hybrid undulator (Halbach II) has been installed at BESSY II [94]. The quasiperiodic scheme has also been applied to an electromagnetic undulator [95].

14.4 On-axis power density

All helical undulators have a reduced on-axis power density. This is a comfortable side effect, though these rather expensive devices are built for another purpose. A device emitting linearly polarized light with a reduced on-axis power density is the figure-eight undulator [96–99]. It consists of independent magnet arrays for the horizontal and vertical fields, respectively. One array has half the period length of the other one. In this way the projected electron trajectory has the shape of an eight. Such devices have been built at SPRING-8 [100–102] and at ELETTRA [103–105]. If an asymmetry is added to the figure-eight motion of the electrons the right- and left-handed circularly polarized light components no longer cancel and elliptically polarized light is emitted [106–108].

In 1990 Tatchyn [109] proposed a variable-period undulator because of the better performance as compared to a fixed-period undulator. If the photon wavelength could be tuned by changing the period length while keeping the undulator parameter K constant ($K \approx 1$), highest brightness could be achieved without producing too much undesirable power in the higher harmonics. A variable-period undulator based on the staggered pole design [110] was proposed at the APS [111] but it was not built, finally, because of the mechanical complexity, even though the design has no permanent magnets. A variable-period ppm-undulator would be even more complicated. For an electromagnetic undulator a period doubling by rewiring has been demonstrated [112] and a superconducting undulator with the potential of period tripling has been proposed [113].

15 Shimming concepts for permanent magnet structures

Today's permanent magnet quality (dipole errors and magnetization homogeneity) is sufficient for most applications. There are, however, products which need a much better field performance and they require a sophisticated field tuning and optimization.

- Permanent magnet NMR spectrometers for medical applications require field homogeneities in the 10^{-6} regime which can be achieved only by shimming.
- A precise, smooth (low vibrations) and dynamic operation of a linear motor is essential for special applications. A homogeneous field quality can be achieved via magnet sorting.
- Sputtering facilities require a good field quality to provide homogeneous depositions. Magnet sorting and shimming help to reach the field accuracy.
- Permanent magnet based accelerator devices for third and next generation light sources such as multipoles or undulators need a sophisticated magnet sorting and shimming prior to installation.

In the following we concentrate on the shimming of undulators. The field optimization has to meet two targets: i) the spectral properties must not be deteriorated by magnet field errors, ii) the multipoles have to be sufficiently small to permit a transparent operation of the devices. The first goal can easily be achieved with well elaborated techniques whereas the second one needs more effort, in particular in the case of variably polarizing devices.

15.1 Static multipoles

Field errors of planar undulators are usually minimized (i.e., shimmed) with small soft iron sheets which are put on top of the magnets facing the electron beam. Trajectory errors are reduced with shims on transversely magnetized blocks whereas phase errors are minimized with shims on longitudinally magnetized blocks [114–117]. Another method developed by Pflueger varies the pole height for trajectory straightening [118].

Conventional shimming can not be applied in APPLE II undulators. Iron shims change their response when the magnet rows are phased and, furthermore, the air gap between the rows does not

permit the application of conventional shims at the centre close to the electron beam. A combination of several techniques has to be applied to achieve a field quality comparable to planar devices. In the following we describe the technique as applied to the BESSY II devices [119].

Prior to assembly the three dipole components are measured in an automated Helmholtz coil system and the magnet block inhomogeneities are measured in a stretched wire set-up. Based on these data the magnets are sorted with a simulated annealing algorithm where the figure of merit is a combination of the transverse field integral distribution and the phase error. An excellent agreement between the prediction of the field integrals based on single block measurements and Hall probe measurements of the complete device is observed (Fig. 30).

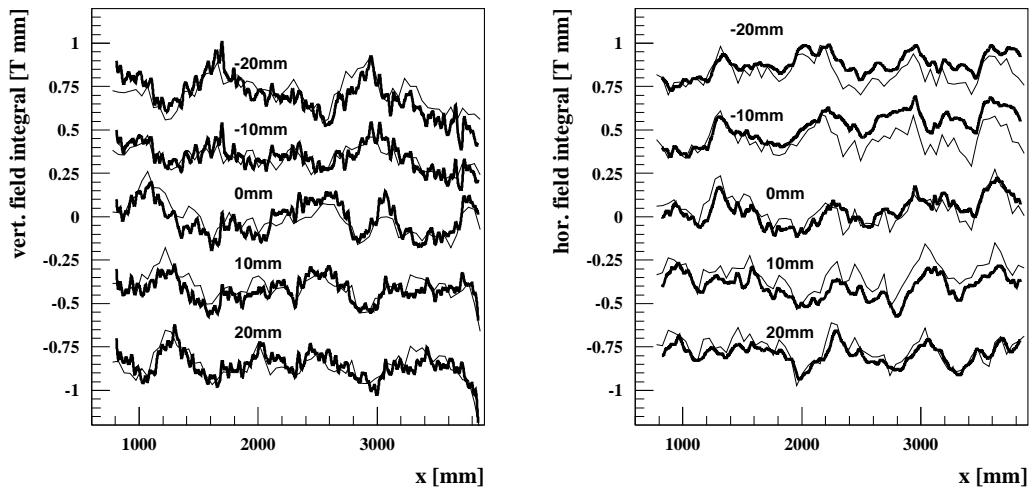


Fig. 30: Vertical (left) and horizontal (right) field integrals of the BESSY II UE49 APPLE undulator at various transverse positions. Thin lines: prediction from single block measurements. Thick lines: Hall probe scans of the complete device. The periodic part has been removed for a better visibility of the errors.

The pre-sorted and assembled structure is measured with a Hall probe, and trajectory and phase errors are removed via virtual shimming [120]. Here, the magnet blocks are moved transversely in the horizontal and vertical direction by up to 0.1 mm. Owing to the finite susceptibility of the magnetic material the field integrals change when phasing the magnet rows. This effect is minimized with iron shims which are placed on the vertically and longitudinally magnetized blocks. In the final step the phase independent field integrals are compensated with magic fingers at either end of the device. The magic fingers are arrays of small magnets with a cross-section of $4 \times 4 \text{ mm}^2$ and variable thickness in the longitudinal direction. The thicknesses are derived from field integral measurements and the response functions of the magnets via a matrix inversion.

15.2 Dynamic multipoles

Undulators feature complicated three-dimensional magnetic fields which cause e-beam focusing (so-called edge focusing) and higher order dynamic effects which can not be described in terms of two-dimensional multipoles (the straight line integrals are zero). They do not obey two-dimensional Maxwell's equations and, sometimes, they are called pseudo-multipoles. The terms as derived in Ref. [121] have the form

$$\theta_{x/y} = -\frac{1}{(B\rho)^2} \int \left\{ \int B_x dz' \cdot \int \frac{\partial B_x}{\partial x/y} dz' + \int B_y dz' \cdot \int \frac{\partial B_y}{\partial x/y} dz' \right\} dz$$

which for an undulator structure reduces to

$$\theta_{x/y} = -\frac{L}{2(B\rho)^2} \frac{\lambda_0^2}{(2\pi)^2} \left\{ B_x^0 \cdot \frac{\partial B_x^0}{\partial x/y} + B_y^0 \cdot \frac{\partial B_y^0}{\partial x/y} \right\}.$$

The pseudo-multipoles scale quadratically with the period length, the maximum field, and the inverse of the energy. Choosing too-small a pole width of a high field wiggler may result in beam dynamic effects which reduce the injection efficiency [122]. APPLE II undulators show strong horizontal field gradients, and hence high pseudo-multipoles. In the elliptical mode the terms can be compensated passively with L-shaped iron shims [123, 124]. In the inclined mode active compensation schemes have to be adopted. The low electron energy of BESSY II (1.7 GeV) requires an active compensation of the pseudo-multiples of the UE112 APPLE II undulator (112 mm period length). Twenty-eight flat wires are glued onto the vacuum chamber. They are powered individually and arbitrary transverse field integral distributions can be produced with a high accuracy [125, 126]. It is worth mentioning that the wires produce two-dimensional multipoles which, by principle, can not compensate the pseudo-multipoles. The wires or L-shims minimize the effects only in the midplane while adding non-linear terms out of plane. Here, they are less harmful if the vertical betatron function is smaller than the horizontal one.

16 Next-generation light sources

Today, many storage ring based third-generation light sources are in operation all over the world serving thousands of users every year. The fields of research include protein crystallography (aiming for an understanding of the function of large biological relevant molecules from its structure); investigation of magnetic materials employing magnetic circular dichroism (MCD) in the soft X-ray regime; element-specific microscopy on the nanometre scale; time-dependent spectroscopy down to the picosecond regime with low alpha optics; non-destructive 3D investigation of matter with respect to the chemical composition, morphology, internal stress and strain; high-pressure study of matter; investigation of electron correlation phenomena such as high-temperature superconductivity; archeometry; and many more topics.

In storage rings time-dependent studies can be extended down to the 100–200 fs using the femtosecond slicing technique [127]. A high-power infrared fs laser interacts with the electron bunch within an undulator (modulator) which is resonant to the laser. The energy of the particles in the interaction region is modulated. In a dispersive section the off-energy electrons are transversely displaced or deflected with respect to the unperturbed beam which permits a spatial separation of the two photon beams as produced in the next undulator (radiator). The polarization of the fs pulses is defined by the radiator. Femtosecond slicing facilities have been built at the ALS [128, 129], at BESSY II [130–132], and at the SLS [133]. At the ALS and the SLS the fs photon beam is linearly polarized whereas at BESSY II the polarization is variable (APPLE II).

The next-generation light sources are free electron lasers (FEL). They provide a high peak brightness, short pulses in the tens of fs regime, and longitudinal and transverse coherent light beams. These properties open new research areas. Currently, two soft X-ray FELs are in operation. FLASH at DESY provides photons up to 200 eV [134] and the SPRING-8 Compact SASE Source (SCSS) test accelerator [135] delivers photons up to 50 eV. FLASH uses fixed-gap undulators (the energy has to be tuned with the electron energy) whereas the SPRING-8 facility operates variable-gap in-vacuum undulators. In spring 2009 the first X-ray FEL, the Linac Coherent Light Source (LCLS), went into operation lasing at 0.15 nm [136]. The fixed-gap device has a transverse canting of the poles for a fine-tuning of the energy. Further X-ray FELs are under construction or are planned at DESY [137], at the SPRING-8 site [138], and at PSI [139]. These facilities will have several undulator beamlines with

typical lengths of 100 m each. The total length of the first three European X-FEL undulator lines adds up to 555 m. The weight of the magnetic material of the three devices is 60 tonnes.

Another challenging linac-based light source concept is the energy recovery linac. A high-quality low-emittance beam as generated in an rf gun is injected into a circular machine and serves many users in parallel. Before the electron bunch is damped, which would be accompanied with growth of emittance and bunch length, it is recovered and replaced by a fresh bunch. ERLs provide a high averaged photon flux, a low emittance with a round beam, and short pulses in the 200 fs regime. Apart from the Jefferson Lab ERL [140], the existing and planned soft X-ray ERLs have a prototype character [141–146]. They are operated for the development and investigation of relevant technologies which are needed in X-ray ERLs. X-ray ERLs are planned at Cornell University [147], at the APS [148], and at KEK JAEA [149–150].

The new light sources require new undulator concepts. The period lengths of the undulators determine the electron energy and shorter period lengths permit shorter linacs, which reduces the total costs. Following this argument, in-vacuum undulators are planned at the SPRING-8 X-FEL and the PSI X-FEL. Circularly polarizing devices covering the regime up to 3 keV are under consideration. Studies on crossed undulators [151–153] predict a power which is one order of magnitude lower as compared to an APPLE II. The degree of polarization is only 80% whereas it is close to 100% for an APPLE II. Most challenging are variable polarizing devices operated under ultra high vacuum conditions. Temnykh built a 30 cm prototype of an in-vacuum Delta-magnet undulator [154]. The fixed-gap, variable polarization device has a four-fold symmetry without any access from the side. This is affordable in linac-based machines which do not require side access for injection. For similar gap-to-period length ratios the peak fields are comparable to the APPLE III design (Fig. 31) [155].

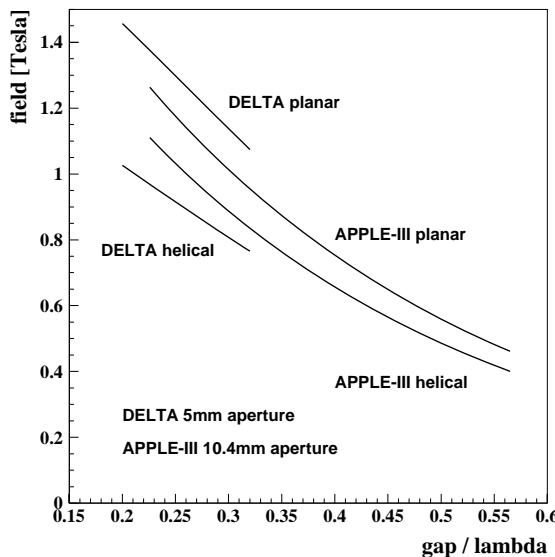


Fig. 31: Field comparison of a Delta undulator and an APPLE III

New acceleration concepts are used in so-called table-top FELs where electrons are accelerated up to the GeV regime within a length of a few centimetres [156]. A high-power laser ionizes the hydrogen atoms in a region of about 30 μm generating extremely high field gradients which accelerate the electrons. These devices are optimized for small dimensions, and short-period permanent magnet undulators as well as permanent magnet quadrupoles are essential. The small vertical apertures require magnets with high radiation stability. It is expected that in-vacuum undulators based on Pr-Fe-B material operated at low temperatures will play a key role in this field.

References

- [1] Product leaflet of Vacuumschmelze 2007, <http://www.Vacuumschmelze.com>
- [2] Y. Luo, 20th Workshop on Rare Earth Permanent Magnets, Crete, Greece, 2008.
- [3] Y. Kaneko, 18th Workshop on HPMA, Annecy, 2004.
- [4] Y. Iwashita et al., Proc. of the PAC, Portland, OR, USA (2003) 2198–2200.
- [5] J. Osborn, Phys. Rev. 67 (1945) 351–357.
- [6] A. Aharoni, J. Appl. Phys. 83 (1998) 3432–343.
- [7] K. Bertsche et al., Proc of PAC, Dallas, TX, USA (1995) 1381–1383.
- [8] P. F. Tavares et al., LNLS-2, Preliminary Conceptual Design Report, Campinas, Brazil, April 2009.
- [9] W. Rodewald et al., Hagener Symposium für Pulvermetallurgie, Band 18 (2002) 225–245.
- [10] Hitachi Metals Ltd., private communication, 2009.
- [11] P. Colomp et al., Machine Technical Note 1-1996/ID, 1996.
- [12] M. Petra et al., Nucl. Instrum. Methods Phys. Res., A 507 (2003) 422–425.
- [13] S. Sasaki et al., Proc. of the PAC, Knoxville, TN, USA, 2005, 4126–4128.
- [14] H. Delsim-Hashemi et al., Proc. of the PAC, Vancouver, BC, Canada (2009).
- [15] R. Qiu et al., Nucl. Instrum. Methods Phys. Res., A 594 (2008) 111.
- [16] J. F. Herbst, Rev. Mod. Phys. 63 (1991) 819.
- [17] T. Tanaka et al., Rev. Sci. Instrum. 71 (2000) 3010–3015.
- [18] J. Chavanne, contribution to IMME 14, CERN, Geneva, Switzerland (2009).
- [19] J. F. Herbst et al., Annu. Rev. Mater. Sci. 16 (1986) 467.
- [20] W. Rodewald et al., IEEE Trans. Magn. 36 (2000) 3279–3281.
- [21] M. Zakotnik et al., J. Iron Steel Res., International, 13 (2006) 289–295.
- [22] M. Zakotnik et al., J. Alloys Comp. 450 (2008) L1–L3.
- [23] J. Cahn, J. App. Phys. 34 (1963) 3581–3586.
- [24] A. Mager, Ann. Phys (Leipzig), 446 (1952) 15–16.
- [25] K. Üstüner et al., IEEE Trans. Magn. 42 (2006) 2897.
- [26] W. Rodewald et al., IEEE, Trans. Magn. 39 (2003) 2932–2934.
- [27] M. Katter, IEEE Trans. Magn. 41 (2005) 3853–3855.
- [28] W. Kuch et al., Phys. Rev. B 65 (2002) 0064406.
- [29] S. Eisebitt et al., Phys. Rev. B 68 (2003) 104419. S. Eisebitt et al., Nature 432 (2004) 885–888.
- [30] F. Pfeiffer et al., Phys. Rev. Lett. 96 (2006) 215505. C. Grünzweig et al., Phys. Rev. Lett. 101 (2008) 025504.
- [31] OPERA 2D and 3D: <http://www.cobham.com/about-cobham/avionics-and-surveillance/about-us/technical-services.aspx>
- [32] MAXWELL: <http://www.ansoft.com/products/em/maxwell/>
- [33] O. Chubar et al., J. Synchr. Rad. 5 (1998) 481–484.

- [34] P. Elleaume et al., Proc. of the PAC, Vancouver, BC, Canada (1997) 3509–3511.
- [35] RADIA: <http://www.esrf.eu/Accelerators/Groups/InsertionDevices/Software>
- [36] J. Chavanne et al., Proc. of the EPAC, Vienna, Austria (2000) 2316–2318.
- [37] K. Halbach, Nucl. Instrum. Methods, 169 (1980) 1–10.
- [38] K. Halbach Nucl. Instrum. Methods, 64 (1968) 278.
- [39] K. Halbach, Nucl. Instrum. Methods, 74 (1969) 147.
- [40] K. Halbach, Understanding modern magnets through conformal mapping, Proc. Symposium on Conductivity and Magnetism: the Legacy of Felix Bloch, Stanford, CA, USA, 1989, Ed. W. A. Little (World Scientific, Singapore, 1990), pp. 1201–1222.
- [41] T. Eichner et al. Phys. Rev. ST Accel. Beams, 10 (2007) 082401.
- [42] S. Becker et al. Phys. Rev. ST Accel. Beams, 12 (2009) 102801.
- [43] Y. Iwashita, Proc. of the PAC, Portland, OR, USA (2003) 2198–2200.
- [44] R. Gluckstein et al., Nucl. Instrum. Methods, 187 (1981) 119–126.
- [45] T. Sugimoto et al., Proc. of the EPAC, Genoa, Italy (2008) 583–585.
- [46] Y. Iwashita et al., Proc of the PAC, Vancouver, BC, Canada (2009).
- [47] Y. Iwashita et al., Proc. of the EPAC, Edinburgh, Scotland (2006) 2550–2552.
- [48] PETRA III Technical Design Report, DESY 2004-035. <http://petra3.desy.de/>
- [49] NSLS II Conceptual Design Report: <http://www.bnl.gov/nsls2/project/CDR/>
- [50] K. Halbach, Nucl. Instrum. Methods, 187 (1981) 109–117.
- [51] K. Halbach, J. Phys. Colloques, 44 (1983) C1-211–C1-216.
- [52] P. Elleaume et al., Nucl. Instrum. Methods Phys. Res., A 455 (2000) 503–523.
- [53] R. Walker, Nucl. Instrum. Methods Phys. Res., A 335 (1993) 328–337.
- [54] T. Tanaka et al., Proc of the FEL Conference, Stanford, CA, USA (2005) 370–377.
- [55] T. Hara et al., Phys. Rev. ST Accel. Beams, 7 (2004) 050702.
- [56] T. Tanaka et al., J. Synchr. Rad. 14 (2007) 416–420.
- [57] C. Kitegi et al., Proc. of the EPAC, Edinburgh, Scotland (2006) 3559–3561.
- [58] J. Chavanne et al., Proc. of the EPAC, Genoa, Italy (2008) 2243–2245.
- [59] J. Chavanne et al., Synchr. Rad. News 22 (2009) 34–37.
- [60] J. Chavanne et al., Proc. of the SRI Conference, Melbourne, Australia (2009) to be published by the AIP.
- [61] T. Tanaka et al., Phys. Rev. ST Accel. Beams, 12 (2009) 120702.
- [62] J. Hu et al., Phys. Stat. Sol. b 188 (1995) 807.
- [63] L. M. Garcia et al., Phys. Rev. Lett. 85 (2000) 429 and references therein.
- [64] K. Üstüner et al., 20th Conference on Rare Earth Permanent Magnets, Crete (2008).
- [65] J. Bahrdt et al., Proc. of the SRI Conference, Melbourne, Australia (2009), to be published by the AIP.

- [66] T. Tanabe et al., Proc. of the SRI Conference, Melbourne, Australia (2009), to be published by the AIP.
- [67] T. Tanaka et al., *New J. Phys.* 8 (1006) 287–303.
- [68] J. Goulon, et al., *Nucl. Instrum. Methods Phys. Res., A* 254 (1987) 192–201.
- [69] J. Pflueger et al., *Nucl. Instrum. Methods Phys. Res., A* 289 (1990) 300–306.
- [70] J. Pflueger, *Rev. Sci. Instrum.* 63 (1992) 295–300.
- [71] P. Elleaume et al., *IEEE Trans. Magn.* 28 (1992) 601.
- [72] J. Chavanne et al., *Rev. Sci. Instrum.* 66 (1995) 1868–1871.
- [73] J. Chavanne et al., *Nucl. Instrum. Methods Phys. Res., A* 421 (1999) 352–360.
- [74] S. Yamamoto et al., *Jpn. J. Appl. Phys.* 26 (1987) L1613–L1615.
- [75] S. Yamamoto et al., *Rev. Sci. Instrum.* 60 (1989) 1834–1837.
- [76] S. Yamamoto et al., *Phys. Rev. Lett.* 62 (1989) 2672–2675.
- [77] X. Marechal et al., *Rev. Sci. Instrum.* 66 (1995) 1937–1039.
- [78] X. Marechal et al., *J. Synchr. Rad.* 5 (1998) 431–433.
- [79] S. Sasaki, *Nucl. Instrum. Methods Phys. Res., A* 347 (1994) 83–86.
- [80] J. Bahrdt et al., Proc of the PAC, Vancouver, BC, Canada (2009), to be published.
- [81] J. Bahrdt et al., Proc. of the SRI Conference, Melbourne, Australia (2009), to be published by the AIP.
- [82] P. Elleaume, *Nucl. Instrum. Methods Phys. Res., A* 291 (1990) 371–377.
- [83] B. Diviacco et al., *Nucl. Instrum. Methods Phys. Res., A* 292 (1990) 517–529.
- [84] T. Hara et al., *J. Synchr. Rad.* 5 (1998) 426–427.
- [85] S. Sasaki et al., *Rev. Sci. Instrum.* 66 (1995) 1953–1955.
- [86] S. Hashimoto et al., *Nucl. Instrum. Methods Phys. Res., A* 361 (1995) 611–622.
- [87] S. Sasaki, Technical Report ELETTRA, Trieste, Italy, ST/M–TN–98/24 (1998).
- [88] M. Kawai et al., Proc. of the EPAC, Sitges (Barcelona), Spain (1996) 2549–2551.
- [89] B. Diviacco et al., Proc. of the EPAC, Stockholm, Sweden (1998) 2216–2218.
- [90] S. Sasaki et al., Proc. of the EPAC, Stockholm, Sweden (1998) 2237–2239.
- [91] J. Chavanne et al., Proc. of the EPAC, Stockholm, Sweden (1998) 2213–2215.
- [92] J. Chavanne et al., Proc. of the PAC, New York, NY, USA (1999) 2662–2664.
- [93] B. Diviacco et al., Proc. of the EPAC, Vienna, Austria (2000) 2322–2324.
- [94] J. Bahrdt et al., *Nucl. Instrum. Methods Phys. Res., A* 467–468 (2001) 130–133.
- [95] T. Schmidt et al., Proc. of the EPAC, Paris, France (2002) 2631–2633.
- [96] H. Kitamura, *Rev. Sci. Instrum.* 66 (1995) 2007–2010.
- [97] T. Tanaka and H. Kitamura, *Nucl. Instrum. Methods Phys. Res., A* 364 (1995) 368–373.
- [98] T. Tanaka and H. Kitamura, *J. Synchr. Rad.* 3 (1996) 47.
- [99] T. Tanaka and H. Kitamura, *J. Electron Spectrosc. Rel. Phenom.* 80 (1996) 441–444.

- [100] T. Tanaka et al., *J. Synchr. Rad.* 5 (1998) 459–461.
- [101] T. Tanaka et al., *Rev. Sci. Instrum.* 70 (1999) 4153.
- [102] T. Tanaka et al., *J. Appl. Phys.* 88 (2000) 2101.
- [103] B. Diviacco et al., *Proc. of the PAC*, Chicago, IL, USA (2001) 2468–2470.
- [104] B. Diviacco et al., *Proc. of the EPAC*, Paris, France (2002) 2610–2612.
- [105] D. Zangrando et al., *Proc. of the PAC*, Portland, OR, USA (2003) 1050–1052.
- [106] T. Tanaka et al., *Nucl. Instrum. Methods Phys. Res., A* 449 (2000) 629–637.
- [107] T. Tanaka et al., *Rev. Sci. Instrum.* 73 (2002) 1724–1727.
- [108] K. Shirasawa et al., *Phys. Rev. ST Accel. Beams*, 7 (2004) 020702.
- [109] R. Tatchyn et al., *IEEE Trans. Magn.* 26 (1990) 3102–3123.
- [110] J. W. Lewellen et al., *Nucl. Instrum. Methods Phys. Res., A* 358 (1995) 24–76.
- [111] G. K. Shenoy, *J. Synchr. Rad.* 10 (2003) 205–213.
- [112] R. Klein et al., *J. Synchr. Rad.* 5 (1998) 451–452.
- [113] A. Bernhard et al., *Proc. of the EPAC*, Genoa, Italy (2008) 2231–2233.
- [114] D.C. Quimby et al., *Nucl. Instrum. Methods Phys. Res., A* 285 (1989) 281.
- [115] S.C. Gottschalk et al., *Nucl. Instrum. Methods Phys. Res., A* 296 (1990) 579.
- [116] B. Diviacco et al., *Nucl. Instrum. Methods Phys. Res., A* 368 (1996) 522.
- [117] P. Elleaume et al., *Synchr. Rad. News* 8 (1995) 18.
- [118] J. Pflüger et al., *Nucl. Instrum. Methods Phys. Res., A* 429 (1999) 368.
- [119] J. Bahrdt et al., *Nucl. Instrum. Methods Phys. Res., A* 516 (2004) 575–585.
- [120] S. Marks et al., *Proc. of the PAC*, New York, NY, USA (1999) 162–164.
- [121] P. Elleaume, *Proc. of the EPAC*, Berlin, Germany (1992) 661–663.
- [122] J. Safranek et al., *Proc. of the EPAC*, Vienna, Austria (2000) 295–297.
- [123] J. Chavanne et al., *Proc. of the EPAC*, Vienna, Austria (2000) 2346–2348.
- [124] J. Bahrdt et al., *Proc. of the SRI Conference*, Daegu, South Korea, AIP Conference Proceedings 879 (2006) 315–318.
- [125] J. Bahrdt et al., *Proc. of the EPAC*, Genoa, Italy (2008) 2222–2224.
- [126] J. Bahrdt et al., *Proc. of the PAC*, Vancouver, BC, Canada (2009).
- [127] A. A. Zholents et al., *Phys. Rev. Lett.* 76 (1996) 912–915.
- [128] R. W. Schoenlein et al., *Science*, 287 (2000) 2237–2240.
- [129] R. W. Schoenlein, *Appl. Phys. B* 71 (2000) 1–10.
- [130] K. Holldack et al., *Phys. Rev. ST Accel. Beams*, 8 (2005) 040704.
- [131] K. Holldack et al., *Phys. Rev. Lett.* 96 (2006) 054801.
- [132] S. Khan et al., *Phys. Rev. Lett.* 97 (2006) 074801.
- [133] P. Beaud et al., *Phys. Rev. Lett.* 99 (2007) 174801.
- [134] <http://flash.desy.de/>

- [135] T. Tanaka et al., Proc. of the FEL Conference, Gyeongju, Korea (2008) 537–542. H. Tanaka et al., Proc. of the FEL Conference, Liverpool, UK (2009) 758–765.
- [136] <http://lcls.slac.stanford.edu/>
- [137] <http://www.xfel.eu/de/>
- [138] <http://www.riken.jp/XFEL/>
- [139] <http://fel.web.psi.ch>
- [140] G. R. Neil et al., Nucl. Instrum. Methods Phys. Res., A 557 (2006) 9–15.
- [141] L. Merminga, Proc. of the PAC, Albuquerque, NM, USA (2007) 22–26.
- [142] E. Minehara et al., Nucl. Instrum. Methods Phys. Res., A 557 (2006) 16–22.
- [143] V. P. Bolotin et al., Nucl. Instrum. Methods Phys. Res., A 557, (2006) 23–27.
- [144] S. L. Smith, Proc. of the PAC, Albuquerque, NM, USA (2007) 1106–1108.
- [145] V. Litvinenko et al., Proc. of the PAC, Albuquerque, NM, USA (2007) 1347–1349.
- [146] M. Abo-Bakr, Proc. of the SRF Conference, Berlin, Germany (2009) 223–227.
- [147] D. Bilderback et al., Synchr. Rad. News, 19, No. 6 (2006) 30–35. G.H. Hoffstaetter et al., Proc. of the PAC, Albuquerque, NM, USA (2007) 107–109.
- [148] M. Borland et al., Proc. of the PAC, Albuquerque, NM, USA (2007) 1121–1123. N. Sereno et al., Proc. of the PAC, Albuquerque, NM, USA (2007) 1145–1147.
- [149] T. Kasuga et al., Proc. of the PAC, Albuquerque, NM, USA (2007) 1016–1018.
- [150] S. Sakanaka et al., Proc. of the EPAC, Genoa, Italy (2008) 205–207.
- [151] Y. Ding et al., Phys. Rev. ST Accel. Beams, 11 (2008) 030702.
- [152] Y. Li et al., Proc of the EPAC, Genoa, Italy (2008) 2282–2284.
- [153] Y. Li et al., Nucl. Instrum. Methods Phys. Res., A (2009) accepted.
- [154] A. Temnykh, Phys. Rev. ST Accel. Beams, 11 (2008) 120702.
- [155] J. Bahrdt et al., Proc. of FEL Conference, Trieste, Italy (2004) 610–613.
- [156] F. Grüner et al., Appl. Phys., B 86 (2007) 431–435.

Bibliography

R. J. Parker, Advances in Permanent Magnetism (Wiley, New York, 1990) ISBN 0-471-82293-0.

A. Hubert and R. Schäfer, Magnetic Domains: The Analysis of Magnetic Microstructures (Springer, Berlin, 1998).

H. Onuki and P. Elleaume, editors, Undulators, Wigglers and their Applications (Taylor and Francis, London, 2003) ISBN 0-415-28040-0.

J. A. Clarke, The Science and Technology of Undulators and Wigglers (Oxford University Press, New York, 2004) ISBN 0-19-850855-7.

R. Walker, Proceedings of CERN Accelerator School, Fifth Advanced Accelerator Physics Course, Rhodes, Greece, 1993, CERN 95-06 v2, editor S. Turner.

R. Walker, Proceedings of CERN Accelerator School, Synchrotron Radiation and Free Electron Lasers, Grenoble, France, 1996, CERN 98-04, editor S. Turner.

P. Elleaume, Proceedings of CERN Accelerator School, Synchrotron Radiation and Free Electron Lasers, Brunnen, Switzerland, 2003, CERN-2005-012, editor D. Brandt.

J. Bahrdt, Proceedings of CERN Accelerator School, Intermediate Accelerator Physics, DESY Zeuthen, Germany, 2003, CERN-2006-002, editor D. Brandt.

Specifications, quality control, manufacturing, and testing of accelerator magnets

D. Einfeld
CELLS-ALBA, Barcelona, Spain

Abstract

The performance of the magnets plays an important role in the functioning of an accelerator. Most of the magnets are designed at the accelerator laboratory and built by industry. The link between the laboratory and the manufacturer is the contract containing the Technical Specifications of the magnets. For an overview of the contents of the Technical Specifications, the specifications for the magnets of ALBA (bending, quadrupole, and sextupole) are described in this paper. The basic rules of magnet design are reviewed in Appendix A.

1 Introduction

The so-called Technical Specifications have to be prepared for the Call for Tender process of the magnets (see Appendix B). This will be the dominant document of the contract with the manufacturer. The main topics covered by the Technical Specifications document include:

- Specifications
- Quality control
- Manufacturing
- Acceptance test

Accelerator magnets include:

- Normal-conducting magnets (bending, quadrupole, sextupole, corrector).
- Superconducting magnets.
- Magnets using permanent magnet material.
- Fast-pulsed magnets.
- Specialized magnets such as septa and kickers used for injection and extraction.

The scope of this paper will be limited to conventional room-temperature, iron-dominated accelerator magnets, mainly bending, quadrupole, sextupole and corrector (see Fig. 1). The main components of the different types of iron-dominated accelerator magnets are (see Fig. 2):

- Iron yoke
- Pole profile
- Coils
- Manifolds
- Sensors or thermocouples
- Supports

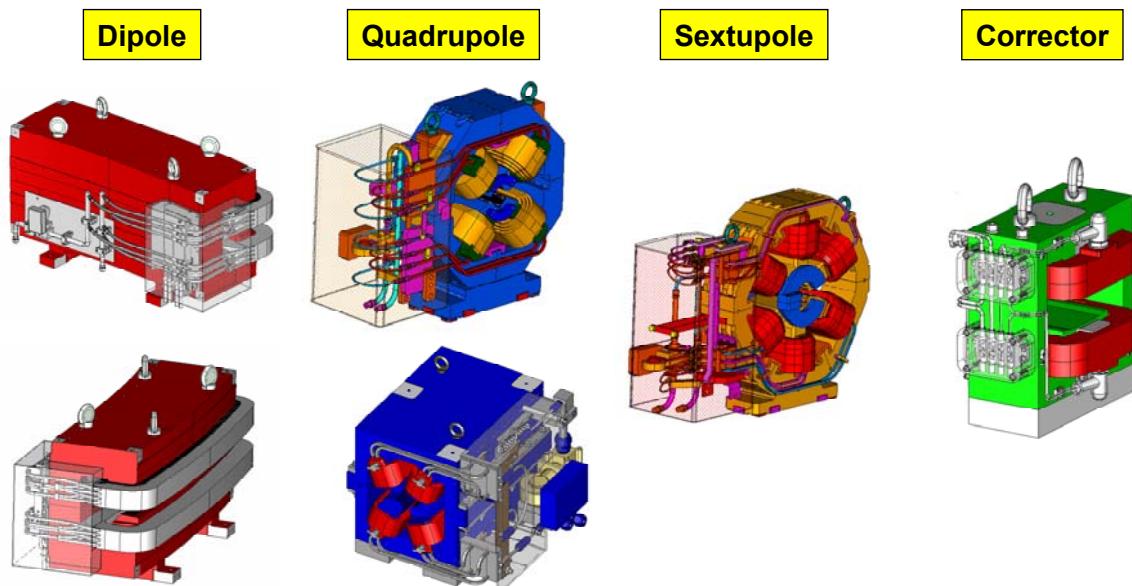


Fig. 1: Examples of different types of iron-dominated accelerator magnets

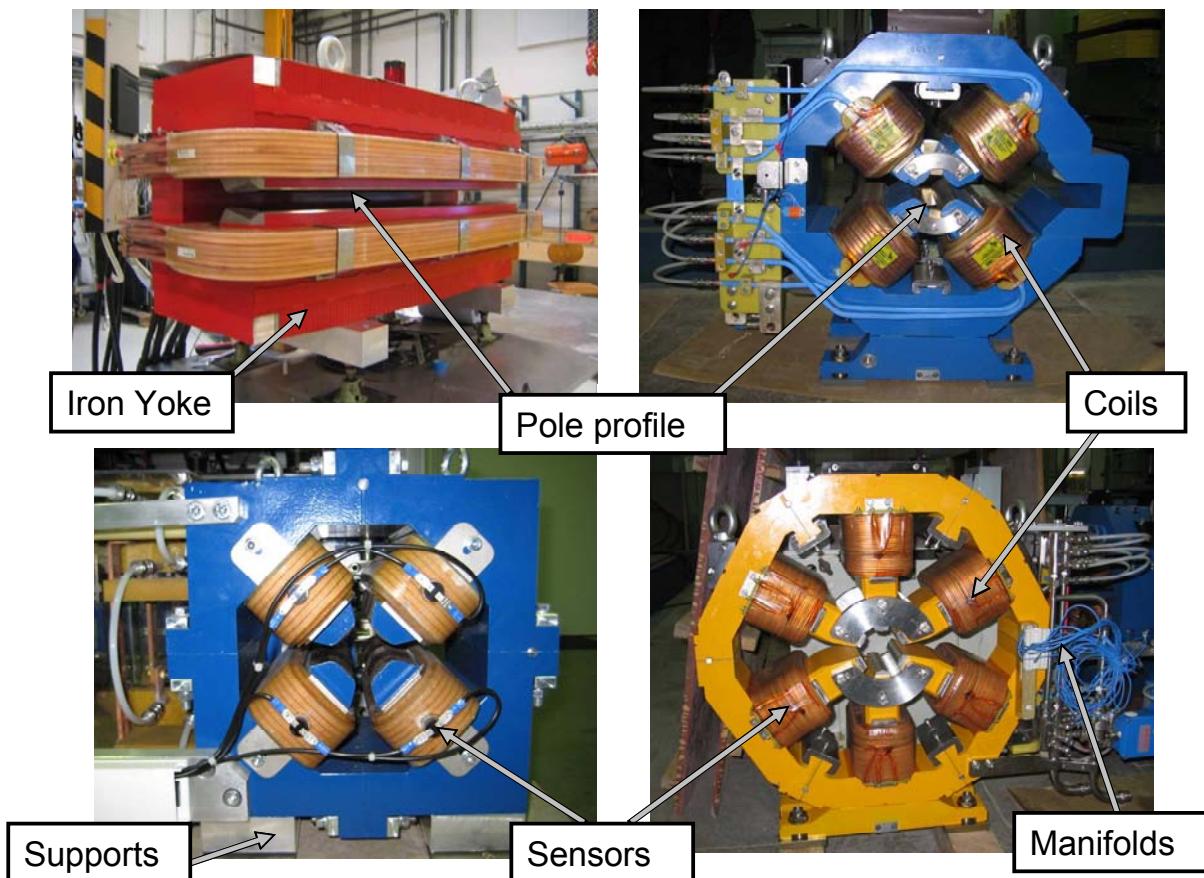


Fig. 2: Main components for different types of iron-dominated accelerator magnets

2 Contents of the Technical Specifications

A typical Technical Specifications document will have a structure similar to the following one:

2.1.) INTRODUCTION

2.2.) SCOPE OF CONTRACT

2.3.) BENDING MAGNET DETAILS

 2.3.1.) Description of the bending magnets

 2.3.2.) Parameter list

 2.3.3.) Specification drawings

 2.3.4.) Trim coils

 2.3.5.) Magnet support feet

 2.3.6.) Survey monument

 2.3.7.) Lifting brackets

2.4.) PHASING OF THE CONTRACT

2.5.) SCHEDULE

2.6.) TENDERING AND CONTRACT MANAGEMENT

 2.6.1.) Tendering

 2.6.2.) Information required with the tender

 2.6.3.) Contract management

 2.6.4.) Quality assurance

 2.6.5.) Documentation

 2.6.6.) Numbering

 2.6.7.) Guarantee

2.7.) MAGNETIC STEEL

 2.7.1.) Steel characteristics for the magnet

 2.7.2.) Laminations

 2.7.3.) Testing of steel

 2.7.4.) Steel supplier

2.8.) LAMINATIONS AND YOKE

 2.8.1.) Laminations stamping

 2.8.2.) Lamination stamping tests

 2.8.3.) Laminations shuffling

 2.8.4.) Yoke

 2.8.5.) Bending magnet ends

 2.8.6.) Mechanical yoke testing

 2.8.7.) Protection and painting

2.9.) COILS

- 2.9.1.) Coils manufacturing
- 2.9.2.) Conductor
- 2.9.3.) Conductor supplier
- 2.9.4.) Conductor tests before winding
- 2.9.5.) Pancake winding
- 2.9.6.) Pancake insulation and impregnation
- 2.9.7.) Terminations of the coil
- 2.9.8.) Coils testing

2.10.) MECHANICAL AND ELECTRICAL TESTS ON COMPLETE MAGNETS

- 2.10.1.) Mechanical and electrical tests on complete pre-series magnet
- 2.10.2.) Mechanical and electrical tests on complete production magnets
- 2.10.3.) Acceptance tests after delivery (Site acceptance test)

2.11.) PACKING AND TRANSPORTATION

- 2.11.1.) Packing
- 2.11.2.) Transportation

Now we will proceed to examine in detail as an example the Technical Specifications document in the case of the bending magnets for the Storage Ring of ALBA. Some aspects regarding the specifications for the quadrupoles and the sextupoles will be described as well.

2.1 INTRODUCTION

- 2.1.1.) The Consortium for the Construction, Equipment and Exploitation of the Synchrotron Light Laboratory (CELLS) is responsible for the construction of a new synchrotron radiation facility, named ALBA. The facility will comprise a 3 GeV electron storage ring, injected from a 100 MeV Linac through a full energy booster synchrotron, and an initial complement of five beam lines.
- 2.1.2.) The Storage Ring is composed of 32 bending magnets, 112 quadrupole magnets and 120 sextupoles magnets, plus a given number of spare magnets.
- 2.1.3.) The present specification contains the technical specifications for the bending magnets. All dimensions and tolerances of the magnets are defined in the specification and the attached drawings. The small electron beam size of ALBA and its stability results in tight tolerances on the magnets.
- 2.1.4.) The contracts for the bending magnets will include the responsibility for the electrical, mechanical and thermal design of the magnets, their construction, followed by their mechanical and electrical testing; CELLS will retain responsibility for the magnetic field generated by the magnet poles and for the magnetic measurements. This specification therefore defines the required pole profile, together with the electrical and mechanical operational performance. The final design arrangements to achieve the specified features in a reproducible way in a correctly engineered magnet will be the responsibility of the manufacturer.
- 2.1.5.) The contracts for the quadrupoles will include the responsibility for the electrical, mechanical and thermal design of the magnets, their construction, followed by their mechanical

and electrical testing, and the measurement of their magnetic performance; CELLS will retain responsibility for the magnetic field generated by the magnet poles. This specification therefore defines the required pole profile, together with the electrical and mechanical operational performance. The final design arrangements to achieve the specified features in a reproducible way in a correctly engineered magnet will be the responsibility of the manufacturer.

2.1.6.) It is desired to have full magnetic measuring of the production magnets carried out at the manufacturer under his responsibility. The manufacturer will provide the necessary magnetic field measuring equipment and suitable technical operators. After testing, it may be necessary to store magnets on the manufacturer's premises. Both these requirements are more fully described in this specification.

2.2 SCOPE OF CONTRACT

- 2.2.1.) This specification covers the engineering design, manufacture, assembly, testing and delivery of bending magnets for the Storage Ring of the ALBA facility.
- 2.2.2.) The specification also covers the supply of all materials and the construction of all tools, jigs and fixtures required to complete the contract.
- 2.2.3.) The magnets, as manufactured, will consist of a magnetic yoke, coils, and all other mechanical brackets and fittings required for their full assembly. They will also be fitted with support feet, water manifolds, electrical termination blocks, coil interconnections and insulated protective covers.
- 2.2.4.) There are not any items which have to be provided by CELLS.

2.3 MAGNET DETAILS

2.3.1 *Description of the bending magnets*

- 2.3.1.1.) ALBA requires a total of 32+1 bending magnets.
- 2.3.1.2.) The bending magnets are curved parallel-ended C-type with a central gap of 36 mm.
- 2.3.1.3.) The magnets generate a magnetic field with a gradient ($G = 5.656 \text{ T/m}$) and operate at a maximum induction of 1.42 T at the central point.
- 2.3.1.4.) The cores will be laminated and the laminations will be stacked together along a curved line with uniform radius thus forming a curved magnet with parallel ends or you have sector magnets.
- 2.3.1.5.) The bending magnets are required with an iron yoke length of 1340 mm.
- 2.3.1.6.) The location of the electron photon beam vacuum chamber in the gap of the magnets leads to the definition of a 'vacuum chamber stay clear area'. This is defined in the appropriate drawings. It will be part of the manufacturer's responsibility, during the mechanical and electrical design of the magnet, to keep this space clear for the vacuum components and to make allowance for this when designing the coil.
- 2.3.1.7.) The magnets will be excited by coils mounted on the poles. These coils will be wound from solid conductor with a central water cooling hole. The bending magnets will be powered by one power supply, all magnets are in series.

2.3.2 Parameter list and drawings

2.3.2.1 Bending magnets

Table 1 shows the list of specified parameters for the bending magnets of ALBA Storage Ring and Booster accelerators. Figures 3–7 show some of the design drawings for Storage Ring bending magnets.

Table 1: Parameter list of ALBA Storage Ring and Booster bending magnets

Storage Ring Bending Magent			Booster Bending Magents		
				5 Degr.	10 Degr.
Magnetic properties			Magnetic properties		
Beam Energy (E)	GeV	3	Beam Energy (E)	GeV	3
Central Field (Bo)	T	1.42	Central Field (Bo)	T	0.8733
Field gradient (Go)	T/m	5.656	Field gradient (Go)	T/m	2.292
Sextupole component (B'')	T/(m^2)	0	Sextupole component (B'')	T/(m^2)	18
Effective length (Lo)	m	1.384	Effective length (Lo)	m	1
Mechanical properties			Mechanical properties		
Bending radius (Ro)	m	7.047	Bending radius (Ro)	m	11.4592
Bending angle (phi)	degrees	11.25	Bending angle (phi)	degrees	5
Central Gap (h)	mm	36	Central Gap (h)	mm	22.6
Length of Fe-yoke L(Fe)	mm	1340	Length of Fe-yoke L(Fe)	mm	0.972
Coil and conductor			Coil and conductor		
Number of coils		2	Number of coils		2
Number of pancakes per coil		4	Number of pancakes per coil		1
Number of turns per pancake		10	Number of turns per pancake		12
Conductor size	mm^2	16.3*10.8	Conductor size	mm^2	12*12
Cooling channel diameter (D)	mm	6.6	Cooling channel diameter (D)	mm	5
Number of ampere turns per coil	A-turns	20340	Number of ampere turns per coil	A-turns	7906
Current (I)	A	509	Current (I)	A	659
Current density (j)	A/mm^2	3.59	Current density (j)	A/mm^2	6.08
Resistance at 23 degrees	mΩ	34.5	Resistance at 23 degrees	mΩ	9.2
Inductivity	mH	40	Inductance	mH	1.3
Voltage drop	V	17.6	Voltage drop	V	6.1
Power	kW	8.93	Power	kW	2
Cooling			Cooling		
Maximim DT	Celsius	8.6	Maximim DT	Celsius	11
Nominal input temperature	Celsius	23	Nominal input temperature	Celsius	23
Number of cooling circuits per coil		2	Number of cooling circuits per coil		2
Maximum pressure drop per magnet	bar	7	Maximum pressure drop per magne	bar	7

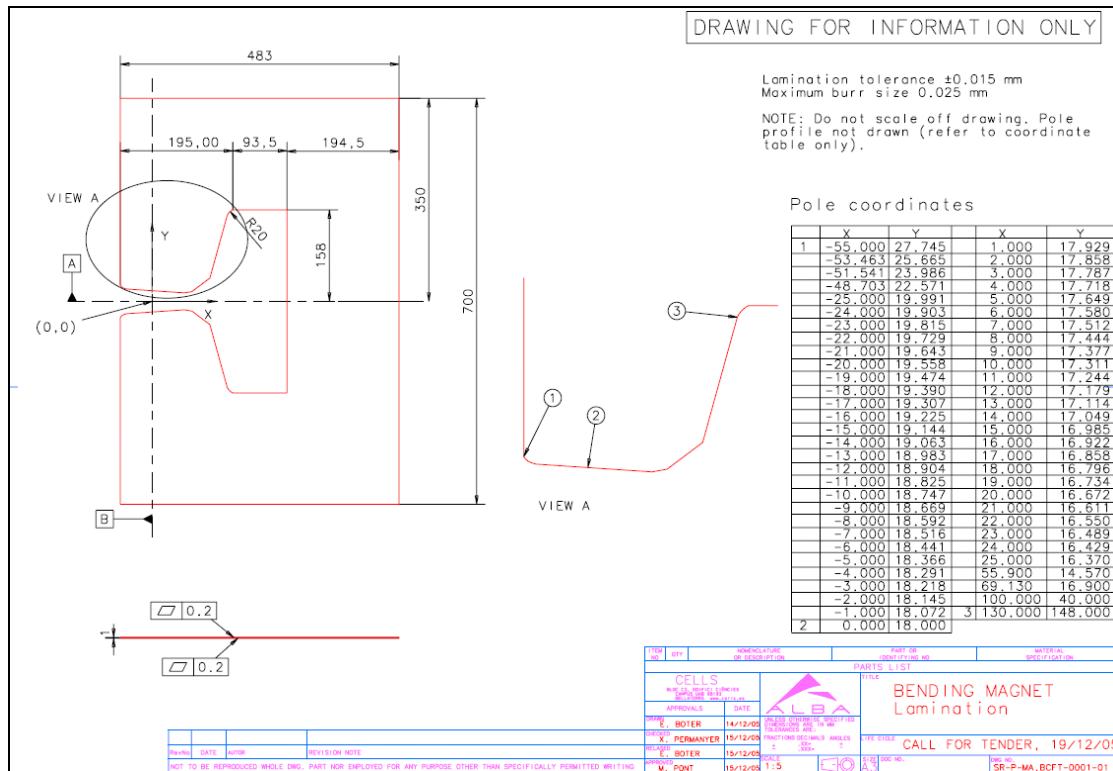


Fig. 3: Drawing showing the dimensions and the pole profile of the laminations for ALBA Storage Ring bending magnets. The required production tolerance is $\pm 15 \mu\text{m}$.

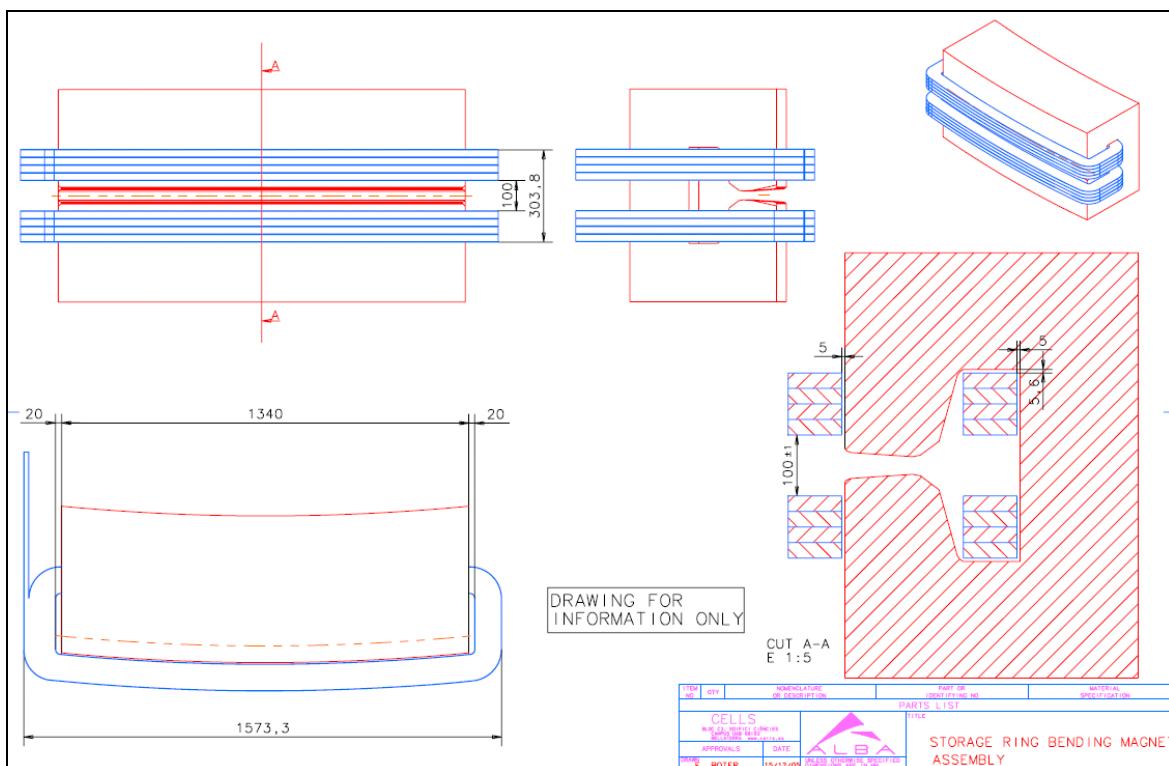


Fig. 4: Drawing showing the dimensions of the yoke and the coils for ALBA Storage Ring bending magnets

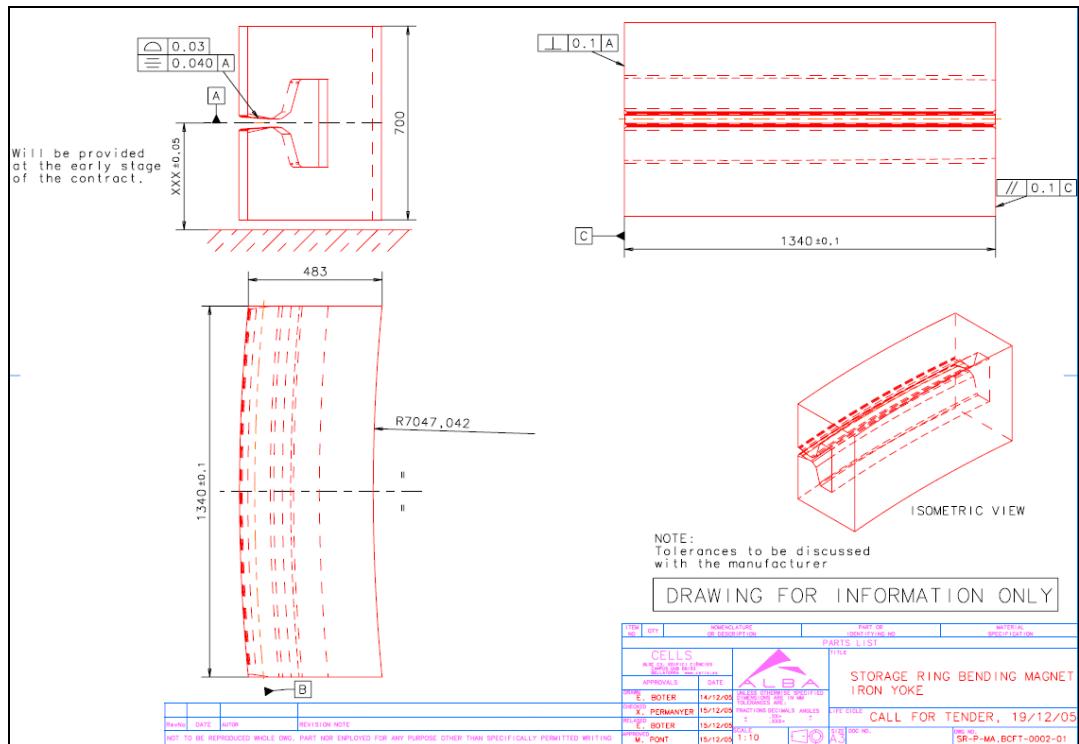


Fig. 5: Drawing showing the dimensions of the yoke for ALBA Storage Ring bending magnets

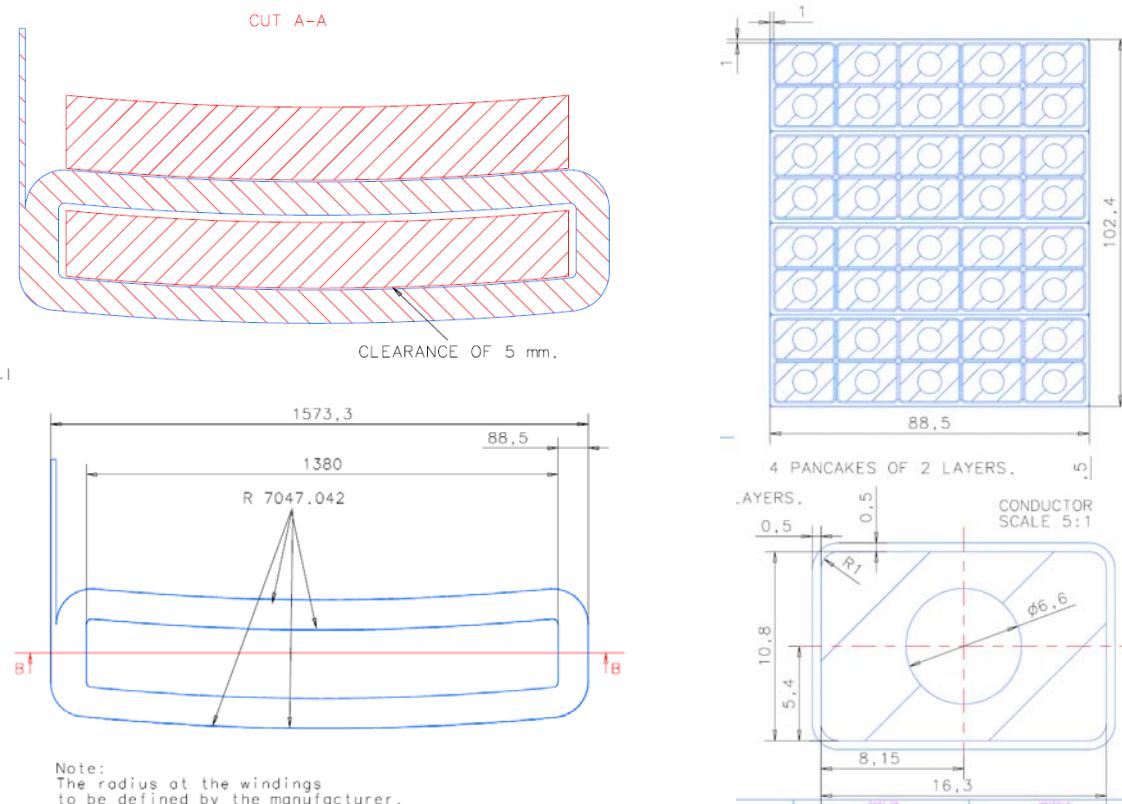


Fig. 6: Drawing showing the dimensions and characteristics of the coils for ALBA Storage Ring bending magnets

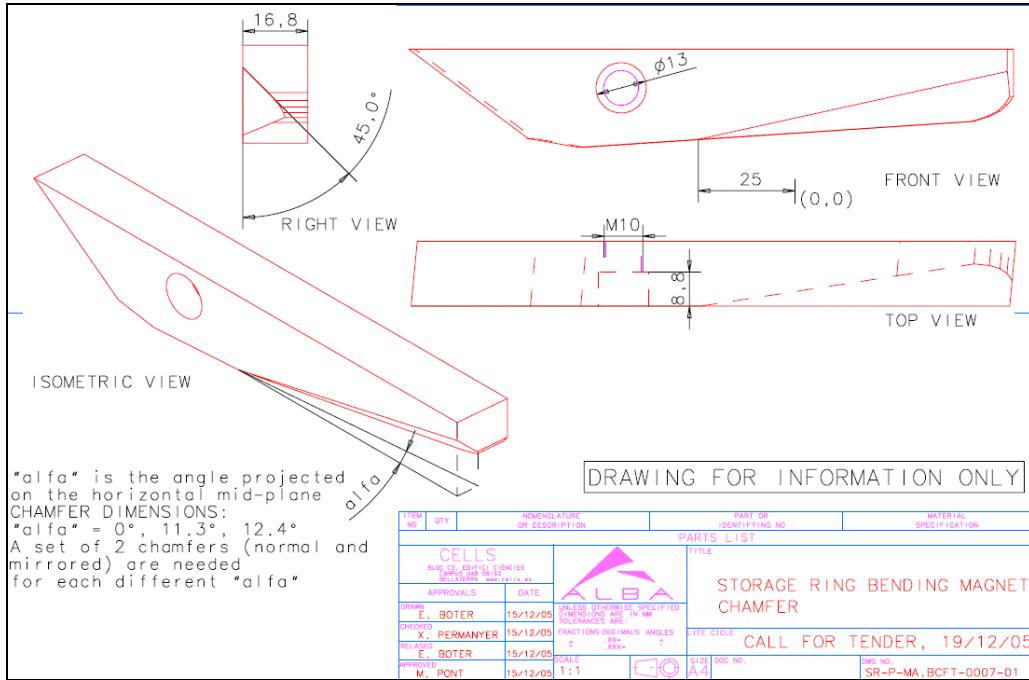


Fig. 7: Drawing showing the dimensions of end chamfers for ALBA Storage Ring bending magnet prototype. The end chamfer determines the overall length and the edge focusing of the magnet.

2.3.2.2 Quadrupole Magnets

Table 2 shows the list of specified parameters for the quadrupole magnets of ALBA Storage Ring and Booster accelerators. Figures 8–12 show some of the design drawings for Storage Ring quadrupole magnets.

Table 2: Parameter list of ALBA Storage Ring and Booster quadrupole magnets

Storage Ring Quadrupole Magents					Booster Quadrupole Magents		
	Q200	Q260	Q280	Q500	QS180	QS340	QC340
Magnetic properties							
Beam Energy (E)	GeV	3	3	3	3	3	3
Field gradient (G ₀)	T/m	19.8	21	21.4	21.9	17.45	17.45
Sextupole component (B'')	T/(m ²)	0	0	0	0	0	5
Effective length (L ₀)	m	0.23	0.29	0.31	0.53	0.36	0.36
Mechanical properties							
Aperture radius	mm	30.5	30.5	30.5	30.5	18	18
Length of Fe-yoke L(Fe)	m	0.2	0.26	0.28	0.5	0.18	0.34
Maximum length of magnet	m	0.298	0.358	0.378	0.598	0.28	0.44
Coil and conductor							
Number of coils		4	4	4	4	4	4
Number of turns per coil		46	46	46	46	17	17
Conductor size	mm ²	8*8	8*8	8*8	8*8	5*5	5*5
Cooling channel diameter (D)	mm	5	5	5	5	3	3
Number of ampere turns per coil	A-turns	7327	7771	7919	8104	2250	2250
Current (I)	A	160	169	172.1	176.2	132.4	132.4
Current density (j)	A/mm ²	3.59	3.81	3.88	3.97	3.78	4.02
Resistance at 23 degrees	mΩ	51.9	60.8	63.8	96	34.6	59
Inductivity	mH	24.6	32.4	34.6	59.2	3	6
Voltage drop	V	8.26	10.3	11	16.9	4.6	7.8
Power	kW	1.32	1.73	1.89	2.98	606	1034
Cooling							
Maximum DT	Celsius	8	8	8	8	8	8
Nominal input temperature	Celsius	23	23	23	23	23	23
Number of cooling circuits per coil		2	2	2	4	1	1
Maximum pressure drop per magnet	bar	7	7	7	7	7	7

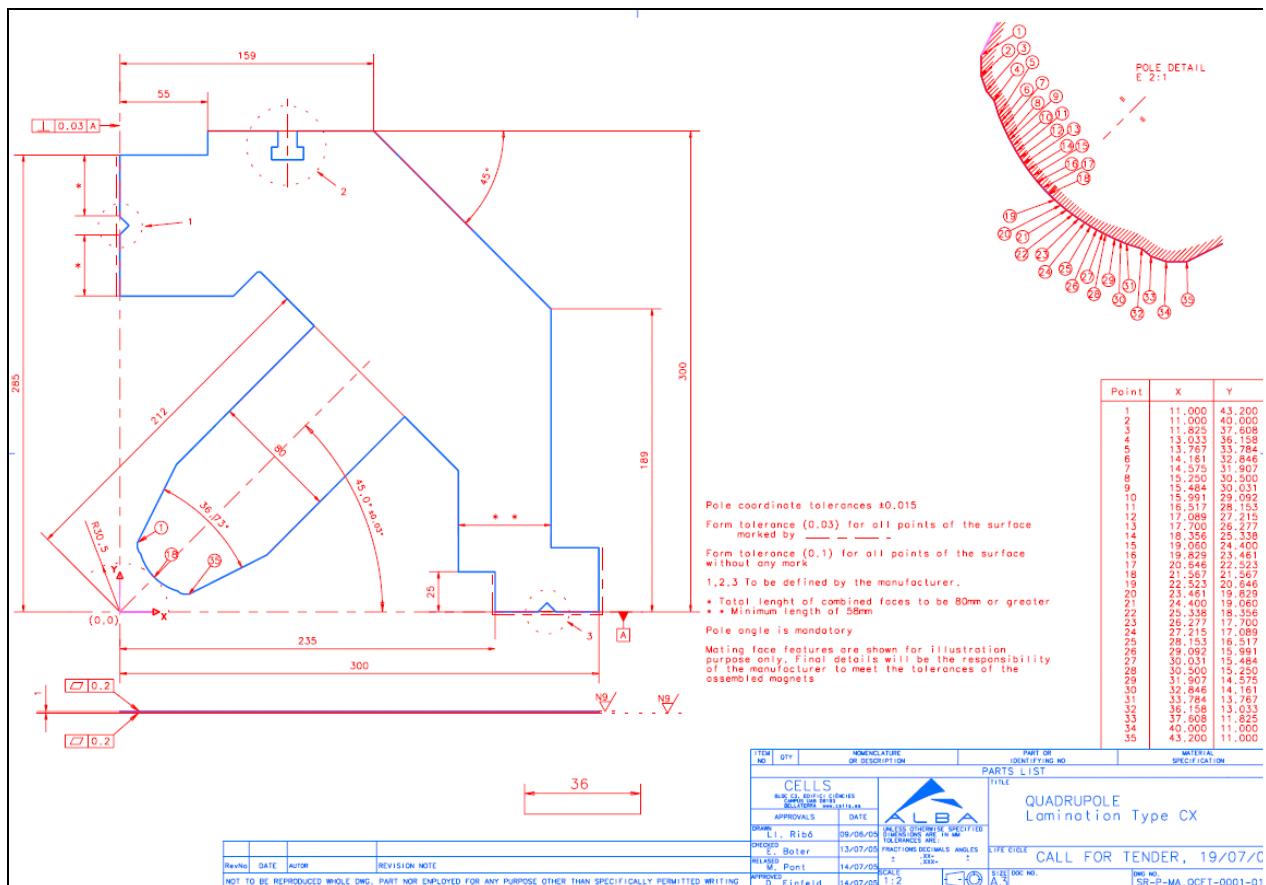


Fig. 8: Drawing showing the dimensions and the pole profile of the laminations for ALBA Storage Ring quadrupole magnets

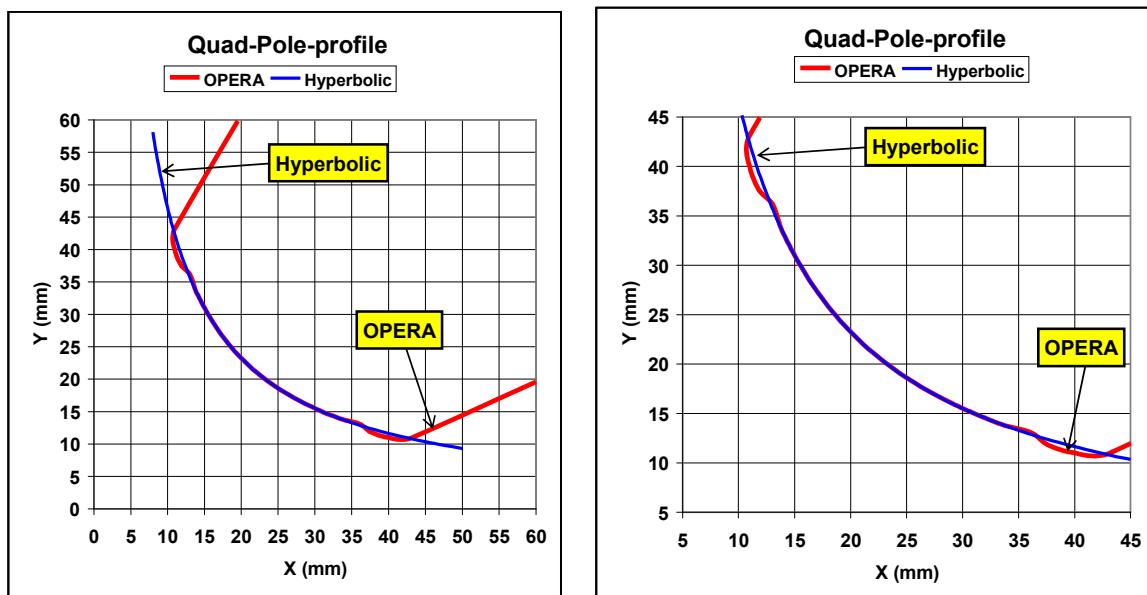


Fig. 9: Comparison between ideal (hyperbolic) pole profile and pole profile determined by means of OPERA magnetic field simulations corresponding to ALBA Storage Rings quadrupoles

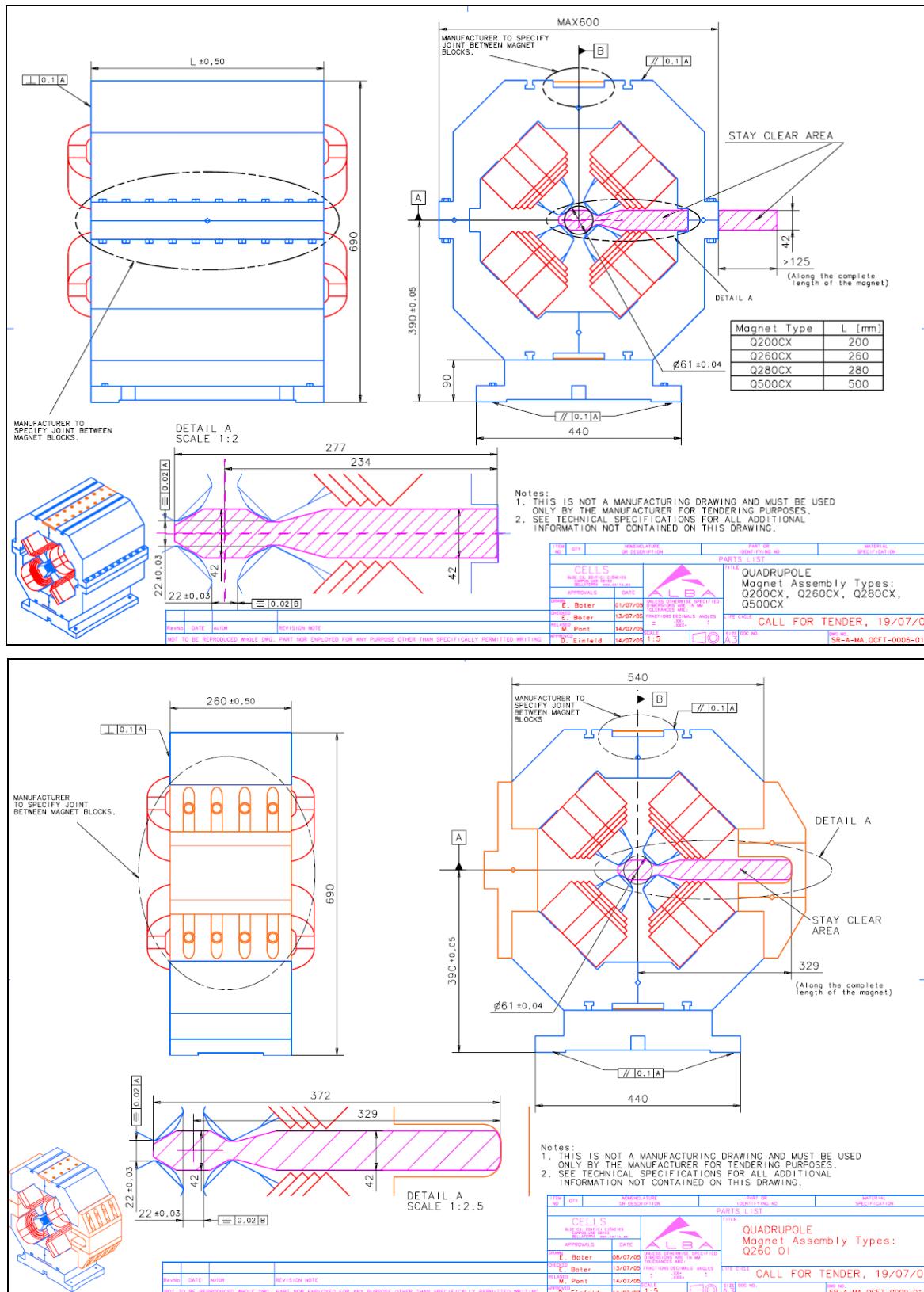


Fig. 10: Dimensions and stay clear area for two types of ALBA Storage Ring quadrupole magnets

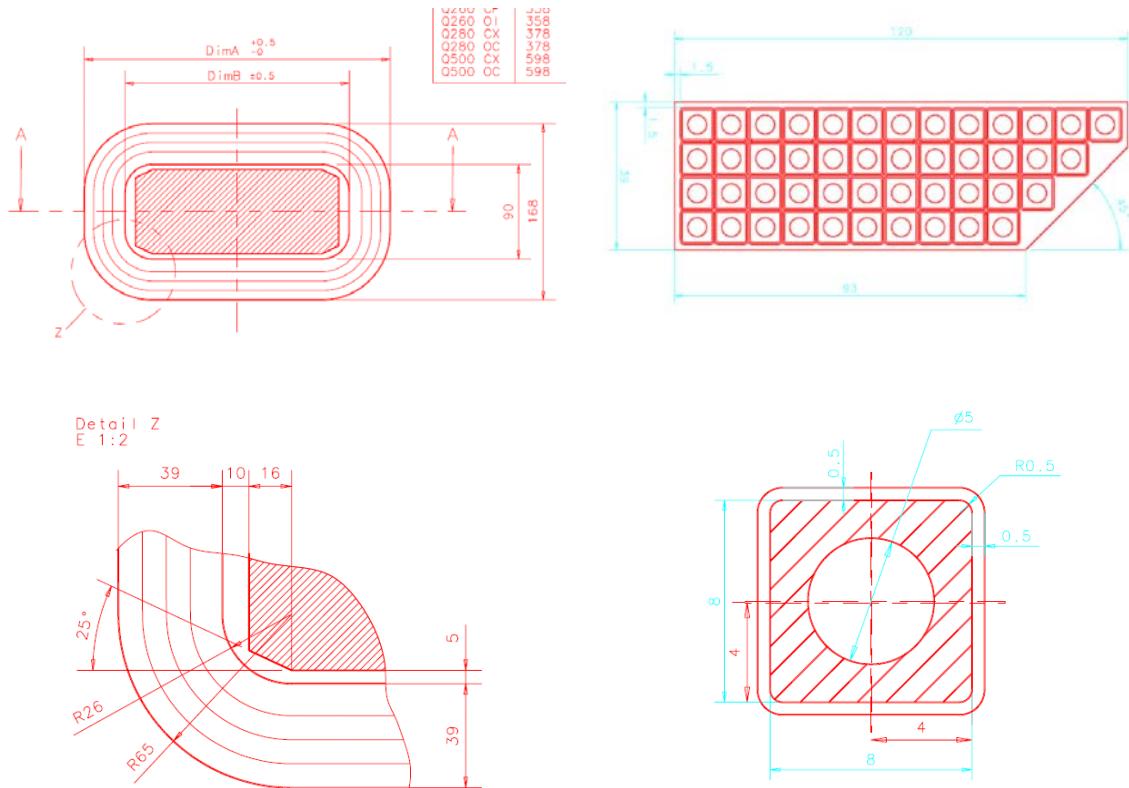


Fig. 11: Drawing showing the dimensions and characteristics of the coils for ALBA Storage Ring quadrupole magnets

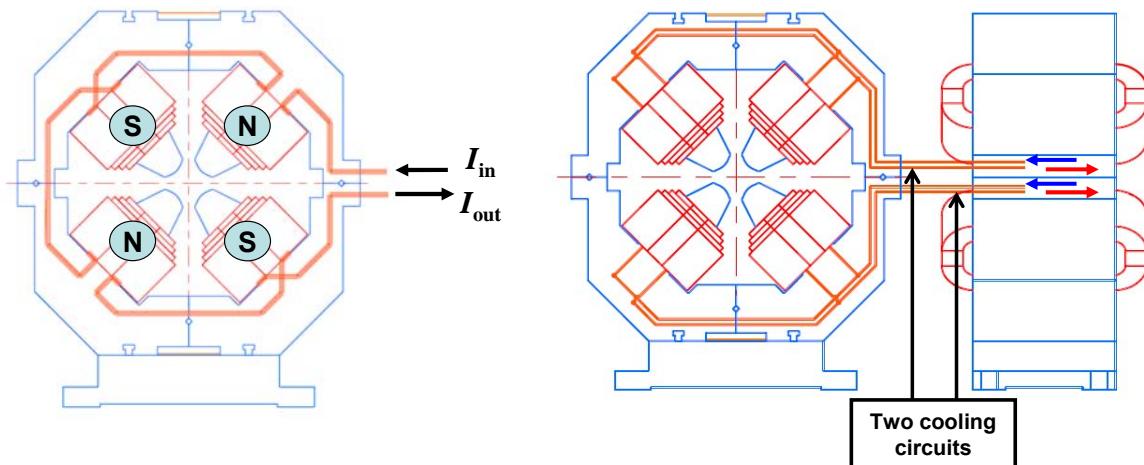


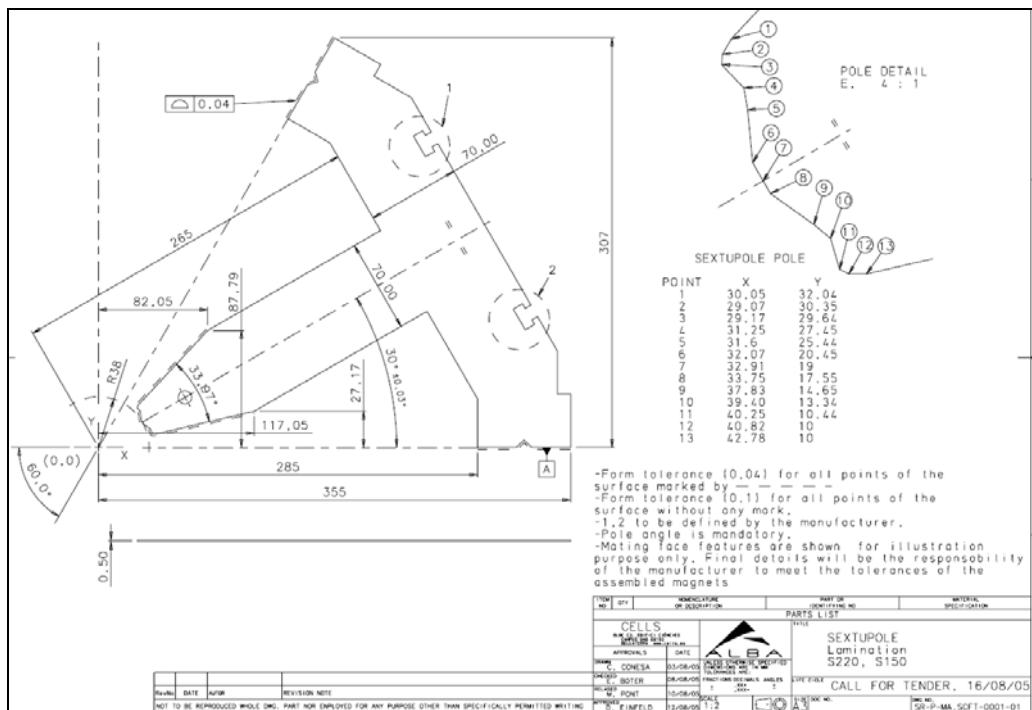
Fig. 12: Electrical (left) and water cooling (right) connections arrangement for ALBA Storage Ring quadrupoles

2.3.2.3 Sextupole magnets

Table 3 shows the list of specified parameters for the sextupole magnets of ALBA Storage Ring and Booster accelerators. Figures 13–15 show some of the design drawings for Storage Ring sextupole magnets.

Table 3: Parameter list of ALBA Storage Ring and Booster sextupole magnets

Storage Ring Sextupole Magents				Booster Sextupole Magents			
	S-150	S-220	S-200				
Magnetic properties				Magnetic properties			
Beam Energy (E)	GeV	3	3	Beam Energy (E)	GeV	3	
Sextupole component (B'')	T/(m^2)	700	700	Sextupole component (B'')	T/(m^2)	400	
Magnetic field at pole tip	T	0.51	0.51	Magnetic field at pole tip	T	0.065	
Effective length (Lo)	m	0.175	0.245	Effective length (Lo)	m	0.2	
Mechanical properties				Mechanical properties			
Aperture radius	mm	38	38	Aperture radius	mm	18	
Length of Fe-yoke L(Fe)	m	0.15	0.22	Length of Fe-yoke L(Fe)	m	0.2	
Maximum length of magnet	m	0.252	0.322	Maximum length of magnet	m	0.3	
Coil and conductor				Coil and conductor			
Number of coils		6	6	Number of coils		6	
Number of turns per coil		28	28	Number of turns per coil		50	
Conductor size	mm^2	7*7	7*7	Conductor size	mm^2	2.8*1	
Cooling channel diameter (D)	mm	3.5	3.5	Cooling channel diameter (D)	mm		
Number of ampere turns per coil	A-turns	5094	5094	Number of ampere turns per coil	A-turns	310	
Current (I)	A	182	182	Current (I)	A	6.2	
Current density (j)	A/mm^2	4.62	4.62	Current density (j)	A/mm^2	2.21	
Resistance at 23 degrees	mΩ	54.2	60.1	Resistance at 23 degrees	mΩ	886	
Inductivity	mH	13.1	18.3	Inductivity	mH	34	
Voltage drop	V	9.87	11	Voltage drop (resistive)	V	5.5	
Power	kW	1.8	2	Power	W	33	
Cooling				Cooling			
Maximum DT				Maximum DT			
Nominal input temperature	Celsius	9	9	Nominal input temperature	Celsius		
Number of cooling circuits per coil	Celsius	23	23	Number of cooling circuits per coil	Celsius		
Maximum pressure drop per magnet	bar	3	3	Maximum pressure drop per magnet	bar		
	5.6	7.46					

**Fig. 13:** Drawing showing the dimensions and the pole profile of the laminations for ALBA Storage Ring sextupole magnets

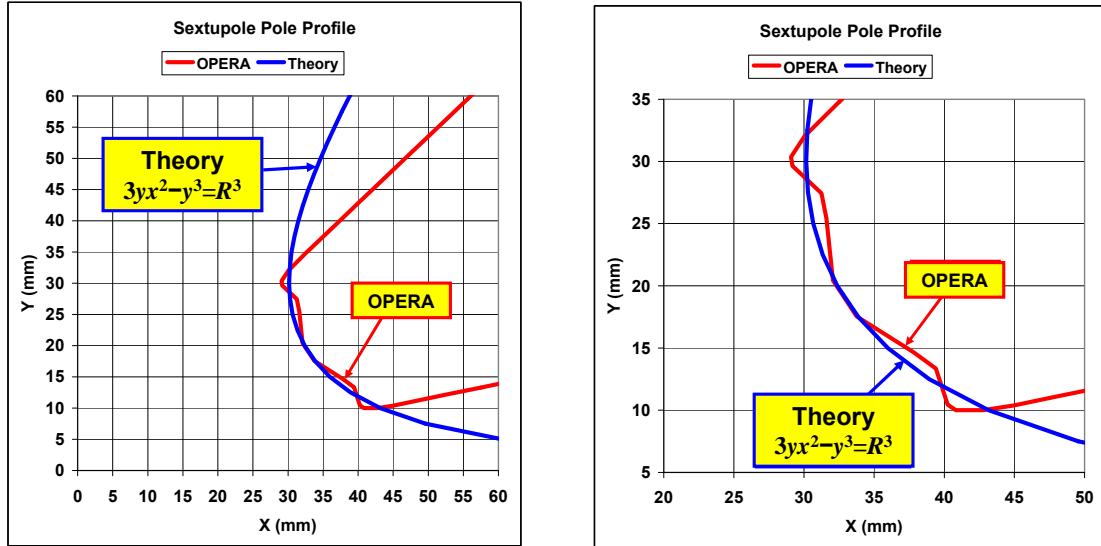


Fig. 14: Comparison between ideal pole profile and pole profile determined by means of OPERA magnetic field simulations corresponding to ALBA Storage Ring sextupoles

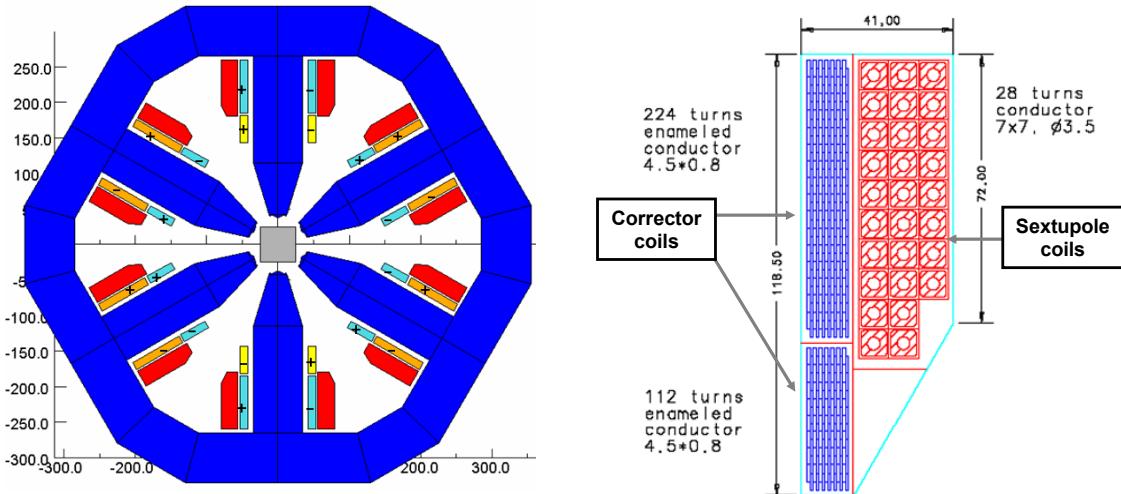


Fig. 15: Drawing showing the characteristics of the coils for ALBA Storage Ring sextupole magnets, which include additional windings for introduction of corrections (horizontal/vertical steering and skew quadrupole correction)

2.3.3 Specification drawings

2.3.3.1.) The drawings are specified in section 2.3.2. To produce a comprehensive set of drawings for the specification, it has been necessary to make certain assumptions concerning parts of the magnet design that will be the responsibility of the manufacturer. Tenderers shall therefore be aware that certain features shown on the drawings are tentative, and will be subject to adjustment by the manufacturer during the design phase. This reservation applies particularly to: (a) coil shape and cross section; (b) shape of magnet lamination, except for the pole profile.

2.3.4 Trim coils

2.3.4.1.) The bending magnets shall be equipped with trim coils capable of a 5 per mil variation in the magnetic field (see Fig. 16). These coils are not shown in the drawings.

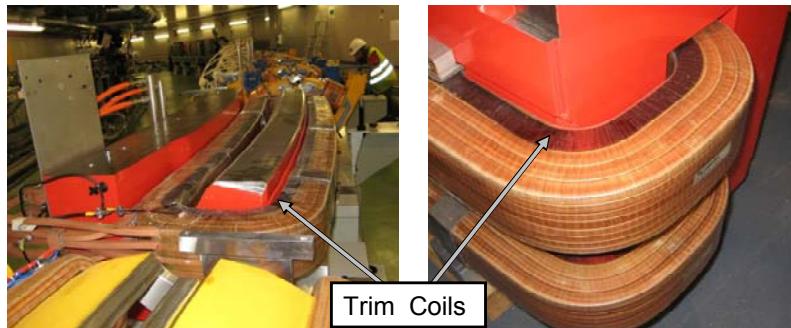


Fig. 16: Images of the trim coils of ALBA Storage Ring bending magnets

2.3.5 Magnet support feet

2.3.5.1.) The bending magnets shall be supplied with supporting feet (see Fig. 17). Details of these feet, as well as of their location will be agreed with the manufacturer at an early stage of the contract.

2.3.6 Survey monument

2.3.6.1.) The bending magnets shall be supplied with target mounting features on their upper face which are required to mount up to 3 survey monuments (see Fig. 17). Details of these monuments will be provided by CELLS at an early stage of the contract.

2.3.6.2.) The bending magnets shall be supplied with reference surfaces in the front and the side of the magnet (see Fig. 17). These surfaces shall be adequate for aligning the magnet with the magnetic bench by using, for example, a dial indicator. Details of these surfaces will be provided by CELLS at an early stage of the contract.

2.3.7 Lifting brackets

2.3.7.1.) The bending magnets shall be supplied with at least four lifting brackets adequate to support the complete magnet (see Fig. 17).

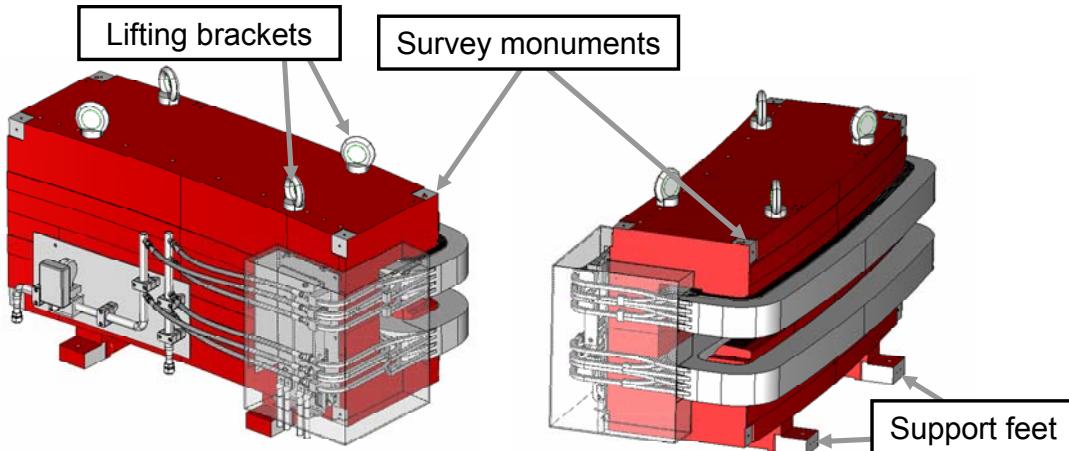


Fig. 17: Location of magnet support feet, survey monuments and lifting brackets for ALBA Storage Ring bending magnets

2.4 PHASING OF THE CONTRACT

2.4.1.) The contract will have two phases.

2.4.2.) The first phase includes:

- All mechanical and electrical design.
- The provision of all tools, jigs and fixtures.
- The procurement of all materials for one (1) pre-series magnet.
- The procurement of all materials that are not available “ex-stock” for the series production magnets.
- The manufacture, mechanical and electrical testing of one (1) pre-series magnet.
- The manufacture of 3×2 different sets of end chamfers as shown on the appropriate drawing.
- The mechanical measurement and testing of the pre-series magnet, as specified in section 2.10.1.

2.4.3.) CELLS will perform the magnetic measurements of the pre-series magnet with the provided end chamfer pieces in order to define the final end chamfer to be used on the production series magnets and asses the viability of the magnet design.

2.4.4.) The second phase includes:

- Any necessary modification of the tooling, which results from the mechanical and magnetic testing of the pre-series.
- The provision of all additional materials required for the production of magnets.
- Manufacture, electrical and mechanical testing and delivery of thirty-two (32) bending magnets.
- Manufacture, testing and delivery four (4) sets of spare coils.

2.4.5.) CELLS will not authorise the second phase until tests and measurements on the pre-series magnet have been successfully completed by the manufacturer and approved by CELLS.

2.4.6.) All designs, tools and materials obtained during phase one shall remain the property of CELLS, and shall be surrendered to CELLS at any time during the contract, within one month of the receipt of written notification.

2.4.7.) Tenderers are required to provide separate quotations for the two phases, which together will fully cover the ‘scope of contract’ defined above.

2.4.8.) The quotation for both phases of the contracts will be fixed price.

2.5 SCHEDULE

2.5.1.) The following program is required for the timescales of the design, construction and delivery:

Contract award	Week 0
Engineering design ready (including tooling design)	Week 16
Pre-series magnet completed	Week 32
Completion of pre-series tests at the factory	Week 34
Magnetic measurements at CELLS and approval of pre-series	Week 42
Completion of delivery of production bendings	Week 66

2.6 TENDERING AND CONTRACT MANAGEMENT

2.6.1 *Tendering*

2.6.1.1.) All interested contractors are strongly encouraged to contact CELLS and discuss details of the specification to ensure that the bidder understands completely the requirements and implications of the specification before making an offer. Enquiries of a technical nature shall be directed to X.YYYY, CELLS, tel: xx-xx-xxxxyy, e-mail: x.yyy@cells.es. Enquiries of a contractual nature shall be directed to Mr. z.wwwwww, CELLS, Tel: xx-xx-xxxwww, e-mail: z.wwwwww@cells.es.

2.6.1.2.) CELLS shall adjudicate the bids by considering the technical, and value for money aspects of the formal bid. See the folder of administrative clauses.

2.6.2 *Information required with the tender*

2.6.2.1.) The bidder shall provide with the tender documents sufficient information to allow an informed choice of contractor. These shall include:

- A confirmation of acceptance of every clause of the present specification or a detailed explanation of any departure from the conditions defined in this specification.
- A breakdown of the price into main categories.
- Details of the quality assurance scheme that the contractor operates.
- A draft time schedule showing the principal design, ordering, manufacturing, and testing of the bending magnets.
- Indications of proposed work packages to be undertaken by any sub-contractors with the identity of the proposed subcontractor.
- A list of previous projects, similar or comparable in size and scope, to enable CELLS to assess the contractors viability and ability to accomplish the contract.

2.6.2.2.) Specific information on:

- Engineering Design
 - Where major features of the magnet design have been left to the judgement of tenderers, an indication of the solutions or parameters shall be given.
 - Proposed method to hold the yoke together.
 - Proposed method to keep the strict mechanical tolerances on the gap region as set in the corresponding drawings.

- Magnetic Steel
 - Proposed source of supply of the steel that will be used in the magnet yokes
 - Technical information required for the proposed steel shall include: Thickness / Grade or type designation / Proposed nature of insulated coating / Quoted permeability / Quoted coercivity.
 - The tenderer shall indicate the test and measurement methods that are proposed for quality control of the mechanical, electrical and magnetic properties of the magnet steel.
- Lamination Stamping and Yoke Assembly
 - A brief description of the tooling proposed for stamping laminations, and whether a one or two stage process is intended shall be given. The size of the stamping press to be used shall be indicated.
 - An estimate of the burr height and extent of shear edge taper expected on the lamination, together with the estimated frequency of tool regrinding to meet the specification, shall be given.
 - Details of the methods proposed and the equipment available for the accurate measurement of the lamination profile.
 - Details of the storage procedure and the method of the subsequent shuffling operation.
 - A description of proposed lamination stacking fixture, including stacking technique. The proposed method for holding the laminations together, and the techniques recommended for maintaining the required dimensional tolerances on the yoke shall be explained.
- Lamination Stamping and Yoke Assembly
 - Details of the measurement procedures proposed for mechanical checking of the yoke after stacking and bonding/welding, and indication of how the dimensional tolerances given by CELLS can be checked.
 - Details of the proposed techniques for any required machining.
 - A preliminary proposal of the instrument that is to be used for the continuous measurement of the gap, taking into account that the gap is not constant in the transversal direction.
- Coil Production and Testing
 - Indication of the expected source of supply for copper conductor, and details of the proposed conductor dimensions.
 - Details of the proposed coil winding operation, including information on the glass cloth to be used. The method of production and all materials to be used (including required packing pieces) must be described. The number of coil winding formers to be used must be stated.
 - The tenderer shall indicate the proposed source of supply of the epoxy resin chemical system, together with details of its expected mechanical, thermal and radiation properties.
 - Details of the proposed impregnation technique and curing operation for the resin system must be given, including information on all jigs, tools and moulds which will be required. The number of moulds to be used must be stated.

- A brief description shall be given of the equipment, which is proposed for the various tests listed in the coil test schedule.
- Details of the proposed over-temperature switches shall be given. The proposed method of mounting the temperature sensor on the coil shall also be explained.
- Details of the proposed water flow switch shall be given. The proposed method of mounting the sensor on the magnet shall also be explained.
- Electrical Connections
 - The tenderer shall indicate which type of connections are proposed for the power cables as well as the proposed layout of terminal boards and protective covers.
 - The tenderer shall indicate the proposed method of coil electrical interconnection and the way that radiation damage is to be avoided.
 - The tenderer shall give details of the proposed connections of the interlock switches that are proposed.
- Water Connections
 - The tenderer shall give details of the proposed water distribution, including drawings of the manifolds and indicating which materials will be used.
 - The water flow rate shall be given.
- The proposed layout, avoiding organic materials in the bending magnet vertical median plane shall be described.

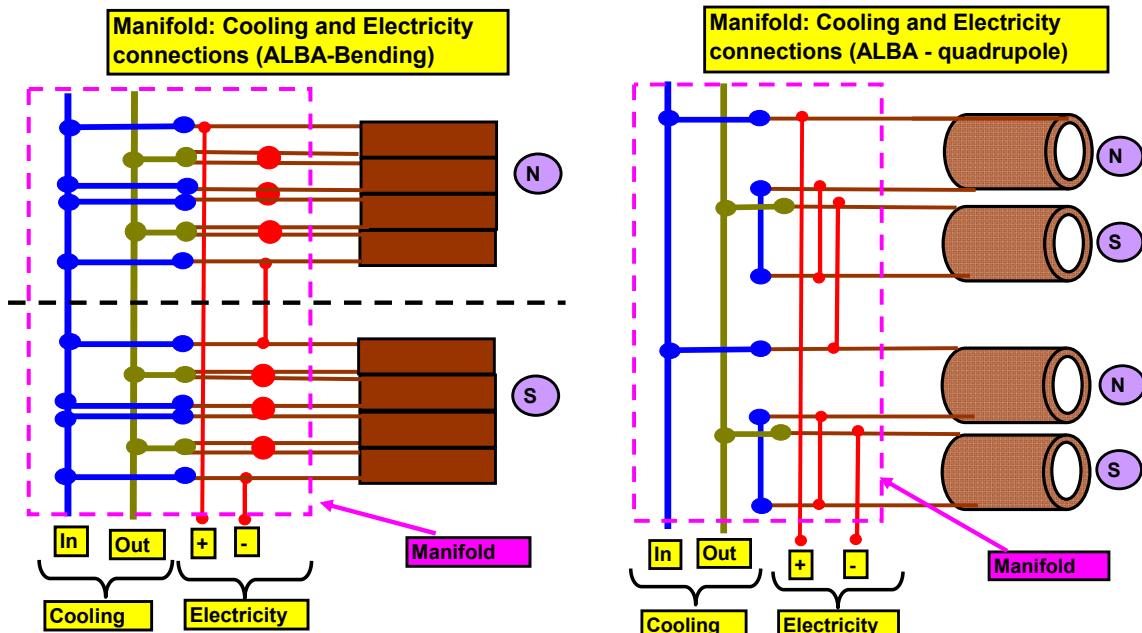


Fig. 18: Scheme of water and electrical connections for ALBA bending and quadrupole magnets

2.6.3 Contract management

2.6.3.1.) At the start of the contract the contractor shall assign contact persons for Technical and Administrative matters who will be responsible for all reporting to, and contact with CELLS.

- 2.6.3.2.) Within 2 weeks of the commencement of the contract the contractor must issue a detailed program covering the design, manufacturing, and testing phases in sufficient detail to allow regular progress monitoring.
- 2.6.3.3.) Within 2 weeks of the commencement of the contract a program of technical and progress meetings will be agreed between the contractor and CELLS.
- 2.6.3.4.) Thereafter, and throughout the contract, the Technical Contact shall supply a written report to CELLS every month detailing progress with respect to the program.
- 2.6.3.5.) The manufacturer will be responsible for the final design, the production methods and the correct manufacture of all magnets, irrespective of whether they have been chosen by the contractor or suggested by CELLS.
- 2.6.3.6.) CELLS's approval of the design and components does not release the contractor from his responsibility to correct errors, oversights and omissions to ensure conformance to the specifications in this document.
- 2.6.3.7.) In the event of the contractor having misinterpreted any of the specifications provided by CELLS, CELLS expects that the misinterpretation will be corrected at no extra cost.
- 2.6.3.8.) The contractor must declare any sub-contractor that will be used in the execution of this contract and inform CELLS of any change of sub-contractor. The change must be accepted by CELLS in writing.
- 2.6.3.9.) Nominated members of CELLS staff, or their appointed representatives must be guaranteed reasonable access to the premises of the contractor, for the purpose of progress meetings, inspection visits etc., even without notice.
- 2.6.3.10.) Nominated members of CELLS staff, or their appointed representatives must be guaranteed reasonable access to the premises of any subcontractor, for the purpose of progress meetings, inspection visits etc., with the main contractor present, even at short notice.
- 2.6.3.11.) CELLS will have the right to observe the tests and suggest additional ones. The contractor shall give at least 10 working days notice of any test date to allow the necessary travel arrangements to be made.
- 2.6.3.12.) CELLS reserves the right to require additional or more extensive tests to be conducted in the event of marginal design or performance.
- 2.6.3.13.) The contract will be completed when the magnets have been delivered to the CELLS indicated site and satisfactorily completed the acceptance tests and comply fully with this specification document.

2.6.4 *Quality assurance*

- 2.6.4.1.) The contractor shall maintain and apply a quality assurance program compliant with ISO-9001 for the design, manufacture and testing of all systems and equipment provided by them.
- 2.6.4.2.) All equipment shall be manufactured in accordance with the best existing techniques and recognised good engineering practices available at the time of construction. All systems shall be designed and constructed with an expected operational lifetime longer than 10 years. The magnets shall be designed and constructed for continuous use.

2.6.5 *Documentation*

- 2.6.5.1.) The contractor shall provide 2 (two) sets of paper copies of the following documentation as soon as it becomes available and in accordance with the schedule presented in section 5:

- Schedule
- Construction drawings of coils and magnetic circuit
- Travellers of control executed at each step of the manufacturing
- A manufacturing booklet which will be composed of:
 - For the coils: Mechanical, hydraulic and electric tests
 - For the magnetic yoke: Dimensional checks
 - For the magnet assembled: Mechanical, hydraulic and electric tests
 - The list of the non compliances processed

2.6.5.2.) All the technical documents delivered by the contractor shall be in English.

2.6.5.3.) The contractor shall provide 2 (two) full sets of paper copies of construction drawings.

2.6.5.4.) In addition, the contractor shall provide 2 (two) full sets of electronic copies on physical media of construction drawings. These shall be preferably in IDEAS format, although other formats e.g. DXF, or IGES are acceptable.

2.6.6 Numbering

2.6.6.1.) Each individual coil, yoke and each completed magnet will be identified and numbered. The position of the identification number shall be agreed with CELLS.

2.6.6.2.) A stainless steel or aluminium plate shall be fixed on each magnet. The following information shall be on the plate:

- Magnet name & Serial number
- Year of manufacture & Maximum current
- Gross weight of the unit & Cooling requirements (if required)
- Coolant flow rate (l/min) & Pressure drop (MPa)
- Maximum Temperature Rise (K) & Contractor's name

2.6.7 Guarantee

The contractor shall guarantee the magnets against failure due to either faulty components or faulty manufacture for a period of 24 months after the magnets have been accepted by CELLS. It is warranted that no modifications will be undertaken without the written permission of the contractor.

2.7 TECHNICAL SPECIFICATIONS: MAGNETIC STEEL

2.7.1 Steel characteristics for the magnet

2.7.1.1.) It is envisaged that this specification will be met by cold rolled, fully annealed, non-oriented, laminated steel.

2.7.1.2.) Table 4 below gives the minimum values of induction under d.c. excitation acceptable at the stated values of field parallel to the rolling direction; values of relative permeability are shown for convenience. Tests are assumed to be made on strips, so that properties parallel and perpendicular to the rolling direction can be separately assessed. Induction measured perpendicular to the rolling direction shall be not less than 20% lower than the values given in Table 4.

Table 4: Required induction and permeability for the magnet steel

Magnetic field [A/m]	Minimum induction parallel to rolling direction [T]	Relative d.c. permeability
116	0.50	3430
208	1.00	3826
300	1.30	3448
597	1.50	1999
1343	1.60	948
3236	1.70	418
6855	1.81	210
12490	1.91	122

2.7.1.3.) The coercivity is defined as the field required to produce zero induction in a mixed sample of the steel after repeated cyclical excursions to high induction with a field of $\geq 10\,000.0$ A/m. The maximum allowed coercivity in a single sample is 80 A/m, and the maximum allowed variation from the mean is $\pm 15\%$.

2.7.1.4.) The steel magnetic properties shall be guaranteed on bulk samples after stamping without any further heat treatment or annealing.

2.7.1.5.) The steel is required to be coated, at least, on one side with an inorganic insulating coating with a maximum thickness of 5 μm .

2.7.2 *Laminations*

2.7.2.1.) Nominal thickness of the lamination is 1 mm.

2.7.3 *Testing of steel*

2.7.3.1.) Steel is normally produced in 'batches'. All batches of steel produced by the steel suppliers shall have test samples taken from the beginning and end of the batch, together with a further sample from the middle of the batch. In exceptional circumstances, where tests on these samples indicate that a large variation of magnetic properties, greater than 15% peak-to-peak, is present within a single batch, CELLS shall be entitled to call for further samples to be taken at one quarter and three quarters through the batch.

2.7.3.2.) All three samples of each batch shall have the following measurements carried out with a strip sample technique for properties parallel and normal to the rolling direction

- i) Permeability at all values of induction specified in section 2.7.1.2 (see Table 4).
- ii) Coercivity as defined in section 2.7.1.3.

2.7.3.3.) The surface insulation shall be checked on three samples taken at locations indicated above, using the standard insulation measuring technique of the steel manufacturer.

2.7.3.4.) The thickness of the steel shall be checked on samples, taken at locations indicated above.

2.7.3.5.) The tenderers shall provide information at tender on the proposed methods for magnetic, electrical and physical measurements of the magnet steel.

2.7.4 Steel supplier

2.7.4.1.) Steel of the described quality is available from a number of suppliers and the manufacturer has full liberty to choose any source of suitable material. CELLS believes that e.g. steel type 1200-100A coated on both sides with Stablit 70 will comply with this specification. CELLS has identified the supplier for this steel:

ThyssenKrupp Electrical Steel GmbH (EBG), Altendorfer Str. 120, D-45143 Essen, Germany

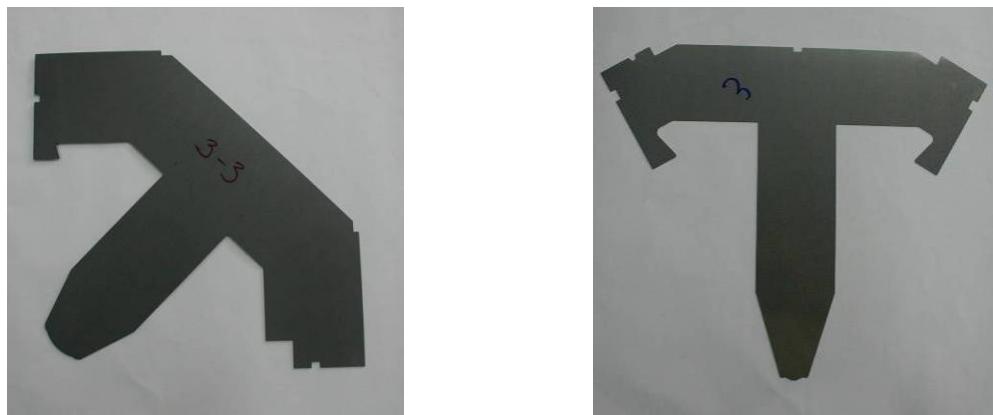


Fig. 19: Examples of laminations for ALBA Storage Ring quadrupoles (*left*) and sextupoles (*right*)



Fig. 20: *Left:* Stacking fixture of laminations for ALBA Storage Ring quadrupole. *Right:* Mechanical test of finished yoke for ALBA Storage Ring sextupole magnet.

2.8 LAMINATIONS AND YOKES

2.8.1 Laminations stamping

2.8.1.1.) The manufacturer is responsible for achieving the required dimensional tolerances for all laminations to meet the mechanical specifications.

2.8.2 Lamination stamping tests

2.8.2.1.) Prior to start production stamping and after each re-grinding of the stamping tools, a certain number of lamination will be punched and three of them shall be measured. In addition, after every 5000 laminations during the stamping operation, a lamination will be taken at random from the production line, and measured. The suitability of the lamination will be judged

with respect to the lamination drawing that will have been approved by CELLS. The measurements shall be submitted to CELLS for approval.

- 2.8.2.2.) CELLS reserves the right to be present during all such measurements.
- 2.8.2.3.) The tested laminations, duly marked with an identifying label, shall then be made available to CELLS for independent verification of dimensional tolerances.
- 2.8.2.4.) CELLS is entitled to reject any lamination which is not within tolerances, as specified in the appropriate drawing.
- 2.8.2.5.) The maximum burr at any location on the lamination must not exceed 0.025 mm; no subsequent deburring operation will be permitted.
- 2.8.2.6.) Re-grinding of the tooling must be carried out at such intervals as are necessary to maintain the required dimensional tolerances of all laminations.
- 2.8.2.7.) After stamping, the laminations produced from each batch of steel shall be stored separately to allow the shuffling operation as described below to be undertaken. The batch of origin of all laminations shall continue to be identifiable.
- 2.8.2.8.) In order to successfully carry out the shuffling operation, all steel delivered for use in Phase 2 of the contracts must be stamped and stored before commencing any further handling and processing operations on the laminations. The steel shall be protected against corrosion.

2.8.3 *Laminations shuffling*

- 2.8.3.1.) This section applies only to laminations produced for use in the second phase.
- 2.8.3.2.) The required magnetic identity between the magnets requires appropriate shuffling of the laminations in such a way that each magnet contains steel representative of each steel production batch, in roughly equal proportions. Manufacturers shall therefore plan a strategy to achieve this, taking into account the number of steel batches required for the magnet production, and the expected variation of magnetic properties within each batch.
- 2.8.3.3.) Final details of the shuffling process shall be agreed with CELLS; these will depend on the results obtained during the magnetic testing of the steel and the periodic measurement of the laminations defined in section 2.8.2.1.

2.8.4 *Yoke*

- 2.8.4.1.) Tenderers shall indicate whether it is their intention to stack and glue laminations to produce yokes in a final state or whether it is their intention to stack and weld the laminations to produce a yoke.
- 2.8.4.2.) Tenderers shall indicate whether it is their intention to perform further machining on one or more faces of the block after yoke assembly. Where such machining is proposed, tenderers shall give a clear explanation of how they intend to ensure that the specified dimensions required on the yoke will be achieved. In any case machining of any part of the yoke is restricted to:
 - a) Machining of the pole profile
 - b) Machining of the reference planes for the alignment features, as set out in section 2.3.6.
 - c) Machining of the reference for the support feet.

The pole has one machined end chamfer to correct the end field in order to provide the correct harmonic analysis. After magnetic measurements of the pre-series magnet the final shape of the end chamfer will be confirmed.

- 2.8.4.3.) During the stacking of the initial pre-series, a record shall be kept of the weight of laminations used for the yoke. After acceptance of the pre-series, this weight will be defined as the standard weight of laminations in a given yoke.
- 2.8.4.4.) During stacking of the production magnet yokes, the number of laminations to be included in the yoke will be determined by weight, and shall be within plus or minus half the weight of a single lamination of the standard weight of laminations in a yoke block.
- 2.8.4.5.) The packing factor has to be at least 98 % and has to be within ± 0.5 % for all of them.

2.8.5 *Bending magnet ends*

- 2.8.5.1.) To control both the effective length and the integrated field quality of the bending magnet, the ends of the poles of each magnet will be chamfered, as specified on the appropriate drawing.
- 2.8.5.2.) To decide the length and the angle of the end chamfer a set of different chamfers, as specified on the appropriate drawing, will be produced during the first phase and a set of magnetic measurements will be taken for the pre-series magnet equipped with all different chamfers. After evaluation of the magnetic measurements done by CELLS the length and angle of the end chamfer to be machined on the production magnets will be communicated to the manufacturer before the second phase starts.
- 2.8.5.3.) Irrespective of the technique used for producing the end modification, the laminations that are to be included in the end region of the poles will be taken from the same stock of shuffled laminations as the unmodified pole region.

2.8.6 *Mechanical yoke testing*

- 2.8.6.1.) After the laminations have been assembled and machined into a yoke the main geometrical dimensions will be carefully checked according to the tolerances of the relevant drawings. Features that shall be controlled are:
- a) Length measured at three different locations along the yoke
 - b) Flatness of the whole assembly
 - c) Squareness of the sides and of the end faces
 - d) Longitudinal shape of the yoke will be checked with a jig, the construction of which is part of this contract.

2.8.6.2.) After assembling of the complete yoke the functional tolerances among which the distance between the reference surfaces on the top and bottom of the bending magnets and the median plane will be checked with the magnet both powered and unpowered. The dimensional controls are specified on the appropriate drawings.

2.8.7 *Protection and painting*

- 2.8.7.1.) After assembly and control, the yokes will be protected against rust by painting. Two-component epoxy paint shall be used, which the manufacturer shall ensure is hard and mechanically resistant. The unpainted areas, as indicated in the appropriate drawings shall be protected by a light oil or other rust preventative measures. The colour will be RAL 3026.

2.9 TECHNICAL SPECIFICATIONS: COILS

2.9.1 *Coils manufacturing*

- 2.9.1.1.) Bending magnets have two (2) coils. Each coil is made of four (4) individual pancakes. Each pancake is composed of 10 turns; therefore, each coil has 40 turns. The coils shall be manufactured using a solid copper conductor with a central hole for the passage of cooling water, according to the CELLS design.
- 2.9.1.2.) The coil is designed to have a maximum temperature rise in the cooling water of 10°C with a differential pressure of 7 bar. Under these conditions the maximum coil temperature will be $\leq 35^{\circ}\text{C}$. It is essential that the insulation system withstands repeated thermal cycling without mechanical or electrical failure.
- 2.9.1.3.) The coils will operate in a radiation environment, and must therefore be built using glass and epoxy resin insulation system. No materials other than those specified in this section of the specification will be permitted. The materials that will be used for insulation will be subject to written authorization from CELLS.
- 2.9.1.4.) Inter-turn resistance shall be provided by wrapping the copper conductor with a borosilicate glass cloth, half-lapped to produce a minimum insulation thickness of 0.5 mm turn-to-turn. After completion of winding, an outer ground insulation shall be provided by further layers of half-lapped glass cloth of minimum thickness 1.0 mm.
- 2.9.1.5.) On completion of the coil winding, electrical and water terminations shall be attached to the leads.
- 2.9.1.6.) Each pancake shall have one (normally closed) over-temperature switch, set to open an electrical circuit at $60^{\circ}\pm 5^{\circ}\text{C}$. These switches shall be fitted each with two external leads or connections, and shall either be fitted before the impregnation process, or glued to the surface of the coil after impregnation. Whichever technique is used, the manufacturer will ensure that good thermal and mechanical contact is obtained using materials that meet the requirements of this specification.
- 2.9.1.7.) All voids arising within the pancakes shall be packed with glass roving in order to avoid the occurrence of resin-rich areas and delamination. The material used shall conform to the glass requirements described below.
- 2.9.1.8.) The insulation in the vicinity of the pancakes leads and terminations will require special attention, in order to provide adequate strength and to avoid the presence of excessive resin in that area. Glass roving or pre-formed glass epoxy inserts must be utilised in these areas, and all materials used must conform to the requirements described below.
- 2.9.1.9.) The manufacturer shall estimate the degree of conductor keystoning that will occur in certain areas of the coil. Resulting voids consequently introduced into the coil must therefore be filled, using the methods indicated above in 2.9.1.8.
- 2.9.1.10.) No joints within the pancakes will be permitted.

2.9.2 *Conductor*

- 2.9.2.1.) The copper shall be Cu-OF Oxygen free (ISO designation) annealed after cold work (dead soft fully annealed temper).
- 2.9.2.2.) The copper shall be free of cracks, porosity and voids. It shall not have any tendency for hydrogen embrittlement. Very good characteristics for brazing are required as well as a ductility which permits the winding of the conductor into magnet coils with tight bends.

2.9.2.3.) The composition shall be at minimum 99.99 % Cu (+Ag).The oxygen content shall be kept below 10 ppm.

2.9.2.4.) The electrical resistivity shall be less than $17.1 \times 10^{-9} \Omega \cdot \text{m}$ at 20°C.

2.9.2.5.) The uniformity of the conductor shall be such that the resistance of all coils constructed from it shall be equal to within $\pm 1\%$.

2.9.3 *Conductor supplier*

2.9.3.1.) Copper conductor of the described quality is available from a number of suppliers and the manufacturer has full liberty to choose any source of suitable material. CELLS has identified a supplier for a conductor that will comply with this specification:

Luvata, Kuparitie, P.O.Box 60, FIN-28101 Pori, Finland, www.luvata.com

2.9.4 *Conductor tests before winding*

2.9.4.1.) Test certificates shall be available relating to tests undertaken by the copper manufacturer, to include dimensions, resistivity and Brinell hardness.

2.9.4.2.) The cooling channel must permit the free passage of a 5.3 mm diameter ball.

2.9.4.3.) Before construction commences, the conductor shall be hydraulically tested at a pressure of 100 bar for five minutes. Conductors revealing any evidence of leakage shall be rejected.

2.9.4.4.) Before the winding of the coil, the conductor shall be cleaned and sandblasted.

2.9.5 *Pancake winding*

2.9.5.1.) Scrupulous care shall be exercised at all stages of the coil construction in the handling of all the components, which shall be undertaken in a clean environment. All working surfaces shall be cleaned immediately prior to be used, and protective gloves shall be worn by all the staff involved. Quality control of this design will have to guarantee the lack of any conductive occlusion between wires (cutting, dust...). Moreover, excessive hammering (hardening) of the conductor which could destroy the fibreglass tape, shall be avoided.

2.9.5.2.) After completion of pancake winding, the pancakes shall be tested with a gas (helium or halogen) at a pressure of 15 bar for thirty minutes. The soundness of the pancake shall be checked at the end of the thirty minutes by passing the probe of a leak detector over the full outer surface of the pancake. This detector will be a mass spectrographic device tuned to the test gas, or similar system. Impregnation shall not be undertaken on any pancake exhibiting evidence of leakage.

2.9.6 *Pancake insulation and impregnation*

2.9.6.1.) The pancakes shall be vacuum impregnated. The use of a mould is considered to be essential, and such a mould must apply direct contact pressure to as great a surface area of the pancake as can be achieved.

2.9.6.2.) Impregnation and curing shall be preceded by oven drying of a pancake and degassing of the resin, pancake and mould. Use of an open mould for the impregnation operation is preferred.

2.9.6.3.) The thickness of unreinforced resin on the surface of a finished pancake must not exceed 0.5 mm.

2.9.6.4.) After completion, the resin on the pancakes must be fully transparent, with no colouriser or additive that would limit observation of the copper turns used within the resin system. No paint or other external coating will be allowed.

2.9.6.5.) No pancake shall be repaired after its initial impregnation without the written approval of CELLS.

2.9.7 *Terminations of the coil*

- 2.9.7.1.) The magnets shall be designed with all mechanical services (i.e. cooling water manifolds) and all electrical connections (including power terminals) on the inside of the ring at the downstream end of the bending magnet.
- 2.9.7.2.) Magnets shall be supplied with an inlet manifold and outlet manifold mounted on the mechanical services panel. The manifold shall be manufactured from metric stainless steel tube, grade 316, suitable for connection to the supply and return water system via a single Swagelok compression fitting onto each manifold. The manifold pipe will be mounted vertically on the mechanical services panel and the connection point will be at the bottom end of the tube.
- 2.9.7.3.) Magnets shall be equipped with a water flow controller Eletta type at the outlet manifold. All part of this controller in contact with water will be in brass with Canigen coating. The Eletta switch will include a witness window for visual flow indication.
- 2.9.7.4.) The water connections on the coils shall be manufactured from phosphor bronze, and shall be attached to the coil by a silver brazing technique to give a system which is unaffected by demineralised water.
- 2.9.7.5.) The electrical connections for power terminals shall be mounted on the electrical and mechanical services panel which shall be rigidly supported on the magnet. These terminals shall be suitable to receive the incoming supply cable connectors, and will be designed to withstand a maximum force of 50 kg exerted by the incoming cables on the terminals. The connection between the coil terminals and the services panel shall be the responsibility of the manufacturer.
- 2.9.7.6.) The manufacturer shall, during initial design, avoid locating any organic based material in the median plane of the magnet centre where it would be subject to long-term radiation damage from the beam. Coil water connections to the manifolds shall be well above and below the beam-line. Where water conduits cross the beam-line horizontal plane, pipe work shall be metallic.
- 2.9.7.7.) The coil terminals, the connection posts and all metallic parts connected to them will be protected against accidental contact by an insulating, transparent cover, which can only be removed by the use of tools; tenderers' proposals for this cover shall be described in the offer.
- 2.9.7.8.) The water connection between the coil terminals and the manifold shall be of non-conducting tube, having suitable mechanical properties, and suitable for use in a high radiation environment. Tenderers shall indicate in their quotation, the type of material that they propose for these tubes.
- 2.9.7.9.) The cabling of the over-temperature interlock switches shall be part of the contract. The two terminals of each switch shall be mounted on the electrical services board as specified.
- 2.9.7.10.) A single terminal connection post able to receive a 10 mm² cable shall be provided for earthing the yoke. The manufacturer shall ensure that there is adequate electrical connection between the yoke, manifolds and other components so that all the exposed metallic parts of the magnet are safely earthed by this terminal post.

2.9.8 *Coils testing*

- 2.9.8.1.) The brazed termination shall be tested at a pressure of 60 bars for 10 minutes.

- 2.9.8.2.) During the total immersion of each coil in water the conductor shall be pressurised at 30 bar with water and sealed. The pressure shall be recorded, and any drop of pressure larger than 2 % during the 24 hours period shall result in rejection of the coil.
- 2.9.8.3.) The water flow for each of the water channels in a coil shall be separately measured with a pressure differential across the channel of 7 bar. The flow rate shall not be less than the flow rate, as calculated by the manufacturer and communicated to CELLS during the design phase, on which the coil thermal calculations are based. The flow in any coil shall also be in the range of $\pm 10\%$ of the mean for all coil flow measurements.
- 2.9.8.4.) The electrical resistance of all coils shall be measured with a DC bridge. The values shall be corrected to 23°C , and must be within $\pm 1\%$ of the mean value for all coils.
- 2.9.8.5.) Each coil shall be immersed in tap water at ambient temperature, but with the terminals exposed above the water level. Any other part of the coil body not then completely immersed shall be covered with wet cloths the ends of which are in contact with the water. The following test sequence shall then be carried out:
- Record insulation resistance between coil terminals and water bath, using minimum voltage of kV. Insulation resistance shall be above $50 \text{ M}\Omega$.
 - Apply direct voltage of 5 kV between coil terminals and water bath for one minute, and record the leakage current.
 - Repeat measurement as in (a).
- Any coil exhibiting evidence of breakdown or significant changes of insulation resistance during these tests shall be rejected.
- 2.9.8.6.) After completion of the tests in 2.9.8.5 each coil shall be energised until the coil temperature increases to 60°C , as measured by the change in electrical resistance. During this period water shall be sealed within the conductor by means of a valve. On attaining the required temperature the current shall be interrupted and water at room temperature allowed to flow through the coil until the conductor again assumes the ambient temperature, as measured by the conductor resistance. The valve shall then again be closed and the foregoing cycle repeated fifty (50) times. The manufacturer may wish to undertake this procedure on a number of coils simultaneously. This test shall be performed in each one of the coils for the pre-series bending magnet and for one in each 5 coils for the series production.
- 2.9.8.7.) On completion of the thermal cycling the insulation tests described in shall be repeated, and significant changes of insulation resistance or breakdown characteristics shall again be sufficient reason for rejection of a coil. Any coil exhibiting evidence of cracking or delamination shall also be rejected.
- 2.9.8.8.) Immediately after the test described in 2.9.8.7 the coil shall be tested by using it as the secondary winding of a transformer. A maximum voltage of 2 kV RMS shall be induced across the coil terminations for a period of one minute, and the corresponding primary current recorded. Any indication of short-circuiting between turns shall result in rejection of the coil.

2.10 MECHANICAL AND ELECTRICAL TESTS ON COMPLETE MAGNETS

2.10.1 *Mechanical and electrical tests on complete pre-series magnet*

- 2.10.1.1.) After the magnets have been assembled with coils and cooling hoses, the complete assembly will be measured to ensure that it complies with the dimensional tolerances as specified.

- 2.10.1.2.) All the dimensions will be checked according to the appropriate magnet assembly drawing.
- 2.10.1.3.) The measuring techniques will be specifically designed to check each and all of the dimensions and dimensional tolerances defined in the appropriate assembly drawing and in this specification. These will be subject to CELLS approval. Manufacturers are requested to give details of their proposed assembly measurement techniques in their tenders.
- 2.10.1.4.) As an essential part of these tests and the subsequent production checks, the manufacturer shall develop and manufacture an instrument that is capable of making precision measurements of the gap region. This instrumentation is required to have the following features:
- a) A measurement sensitivity and reproducibility equal to or better than $\pm 10 \mu\text{m}$;
 - b) A monitorable electrical output, allowing gap dimensions to be continuously measured as the gauge traverses through the magnet;
 - c) Accuracy and sensitivity unaffected by magnetic fields, i.e. to be capable of performing the required measurements with the bending magnet powered.
- 2.10.1.5.) A direct voltage of 5 kV shall be applied between the terminals of each coil and its magnet yoke for one minute. Any coil showing evidence of breakdown, indicated by a leakage resistance of less than $50 \text{ M}\Omega$, shall be rejected.
- 2.10.1.6.) A maximum d.c. operating current test with a coil excitation of 530 A shall be carried out for a period of at least two hours, with cooling water circuits set to provide a differential pressure not greater than 7 bar. During this test, the water inlet and outlet temperature shall be monitored, and the temperature of coil surfaces and all coil interconnections and terminals checked with contact thermometers. Results shall be judged with respect to the appropriate magnet thermal specifications. Any coil showing evidence of overheating, local hot spots or other faults during this period shall be rejected. This test can be performed at the commencement of the magnet measuring sequence, with the magnet connected to a power supply and cooling water.
- 2.10.1.7.) The manufacturer shall demonstrate the operating efficiency of the over-temperature switches by raising the temperature of each coil to the value (60°C) at which the switches are guaranteed to operate. The technique that is to be used for the necessary overheating shall be agreed with CELLS. Because of the danger of damage to the coil, CELLS strongly prefers a method involving the external heating of the circulating water. Details of the proposed method shall be included in the tender.
- 2.10.1.8.) The manufacturer shall demonstrate the operating efficiency of the water flow switch by restricting slowly the supply of water with the magnet unpowered.

2.10.2 *Mechanical and electrical tests on complete production magnets*

- 2.10.2.1.) After the magnets have been assembled with coils, cooling hoses and mounting brackets fitted, the complete assembly will be measured to ensure that it complies with the dimensional tolerances as specified.
- 2.10.2.2.) All the dimensions will be checked according to the appropriate magnet assembly drawing.
- 2.10.2.3.) A direct voltage of 5 kV shall be applied between the terminals of each coil and its magnet yoke for one minute. Any coil showing evidence of breakdown, indicated by a leakage resistance of less than $50 \text{ M}\Omega$, shall be rejected.

2.10.2.4.) A maximum d.c. operating current test with a coil excitation of 530 A shall be carried out for a period of at least two hours, with cooling water circuits set to provide a differential pressure not greater than 7 bar. During this test, the water inlet and outlet temperature shall be monitored, and the temperature of coil surfaces and all coil interconnections and terminals checked with contact thermometers. Results shall be judged with respect to the appropriate magnet thermal specifications. Any coil showing evidence of overheating, local hot spots or other faults during this period shall be rejected. This test can be performed at the commencement of the magnet measuring sequence, with the magnet connected to a power supply and cooling water.

2.10.3 Acceptance tests after delivery

2.10.3.1.) After delivery, the bending magnets shall be visually inspected for mechanical damage suffered in transit. Any such damage shall be reported to the manufacturer. Possible repair shall be subject to agreement with CELLS. Where the damage has resulted in alteration to the magnet iron geometry or to the soundness or shape of coil conductor, insulation or terminals, the magnet shall be rejected.

2.10.3.4.) Electrical tests shall be carried out by staff of CELLS after delivery. The manufacturer has the right to be represented during these tests but shall notify CELLS in writing if this right is to be exercised.

2.10.3.5.) A direct voltage of 5 kV will be applied between the terminals of each coil and its magnet yoke for one minute. Any coil showing evidence of breakdown, indicated by a leakage resistance of less than 50 MΩ, shall be rejected.

2.10.3.6.) The magnet will be energised at the maximum current of 530 A for a period of at least two hours. Any coil showing evidence of breakdown, local hot spots or other faults during this period shall be rejected.

2.11 PACKING AND TRANSPORTATION

2.11.1 Packing

2.11.1.1.) The contractor will submit to CELLS a solution for the packing. This packing will have to use the classical handling tools. The magnet packed shall be protected against the elements, the projections and the breaks during transportation and storage.

2.11.1.2.) The packing of each magnet shall be dust proof, water proof and will have to protect the steel parts against the oxidation. Moreover, the pieces shall be protected against the strain, impacts and rubbing which can damage their surfaces.

2.11.1.3.) A particular protection shall be required for brittle parts (reference surfaces, electrical connections and coils).

2.11.1.4.) The coils shall be rinsed, dried and sealed before expedition in order to avoid any frost.

2.11.2 Transportation

2.11.2.1.) The contractor will include in his tender the transport from the factory:

- a) To the site where CELLS will control and measure the magnets. This place will not be necessarily in the CELLS site.
- b) Or to the storage area that CELLS will indicate. This storage area will be at CELLS site or nearby.

2.11.2.2.) The contractor keeps the responsibility of the bending magnets until the delivery to one of the sites listed above. CELLS will supply the local handling tools.

2.11.2.3.) The transfer of risks shall take place when the load is laid down on the ground.

Acknowledgement

The author wants to acknowledge the help and contribution of different colleagues: Montse Pont (CELLS/ALBA), Davide Tommasini (CERN), Neil Marks (Crockcroft Institute), and G. Moritz (GSI). The author is indebted to Jordi Marcos (ALBA/CELLS) for his excellent assistance in preparing the manuscript.

Bibliography

Jack T. Tanabe, *Iron Dominated Electromagnets* (World Scientific, Singapore, 2005) ISBN 981-256-327-X.

Neil Marks, Magnets (warm): Conventional magnets for accelerators, CAS – Introduction to Accelerator Physics, Zakopane, Poland, 1–13 October 2006.

D. Einfeld, Magnets (warm), CAS – General Accelerator Physics Course: Introductory Level, Frascati, Italy, 2–14 November 2008.

Appendix A: Design criteria for different types of magnets

In this appendix we recall the main points concerning the design of iron-dominated magnets (bending, quadrupole, and sextupole).

A.1 Design criteria of a dipole magnet

A dipole or bending magnet has two poles generating a constant field which steers the particle beam. The purpose of the complete set of bending magnets in a circular accelerator is to bend the beam by exactly 360 degrees. Some examples of bending magnets are shown in Fig. A1.

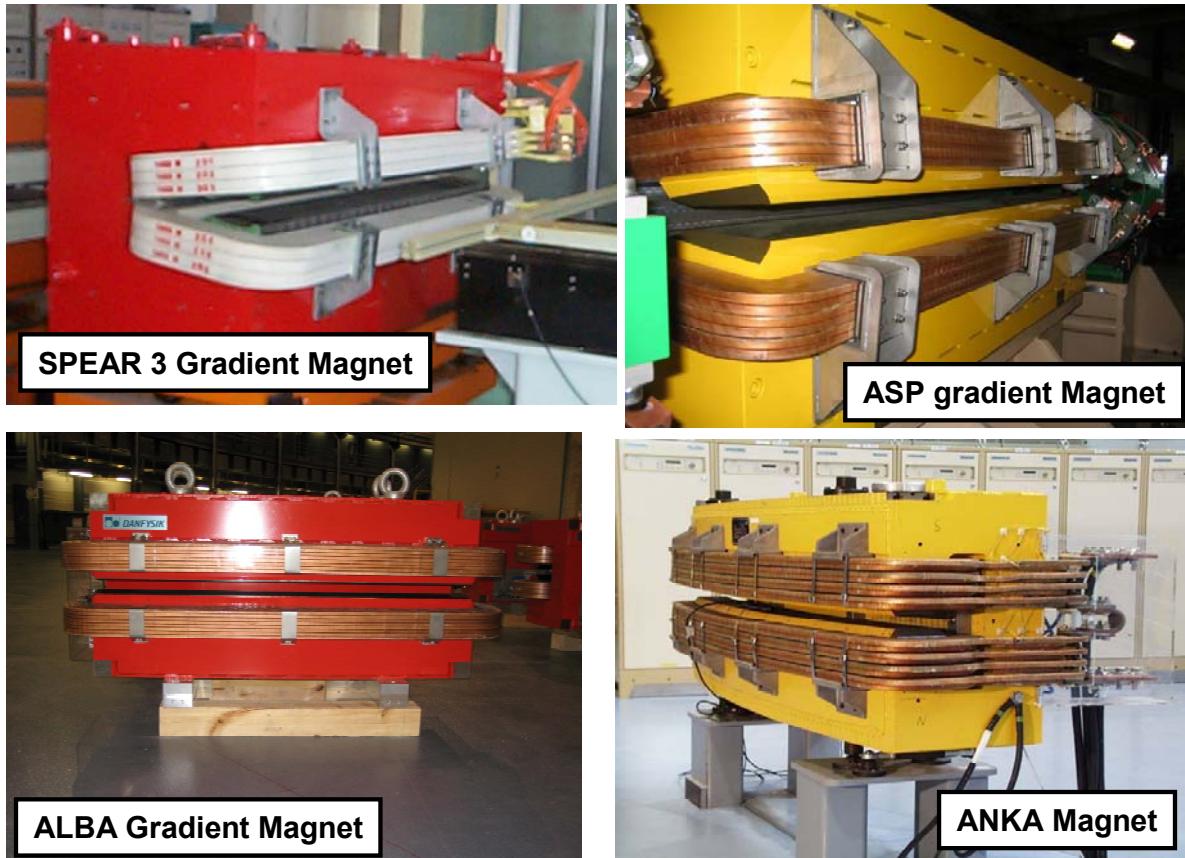


Fig. A1: Examples of dipole magnets for different synchrotron light source facilities

The gap ($g = 2 h$) in a normal bending magnet is constant:

$$h(x) = \text{const.} = \frac{gap}{2} \quad (\text{A1})$$

At the ends of the pole there are shims introduced to optimize the ‘good field region area’ (see Fig. A2).

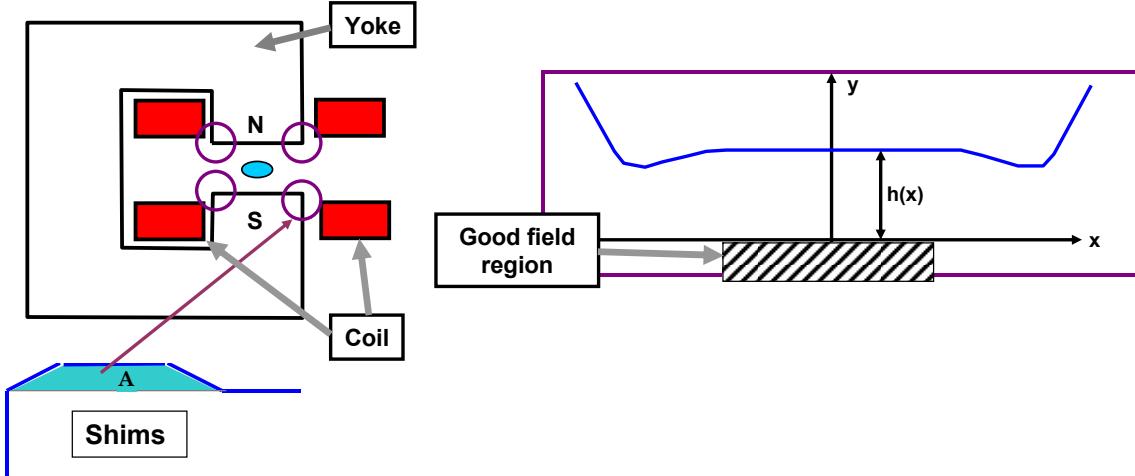


Fig. A2: Left: Schematic cross section of a bending magnet and detail of the end of the pole with a shim. Right: Pole profile with an indication of the good field region of the magnet.

The size of a dipole is determined by the dimensions of the ‘good field region’ required by accelerator physicists. For example, in a typical bending magnet for a synchrotron light source a good field region of roughly ± 15 to ± 20 mm is required. In heavy-ion accelerators or other machines the dimensions of the good field region can be completely different (usually much larger).

To optimize the pole profile the so-called shims are used at the end of the poles (as already shown in Fig. A2). The contour of the shims has to be determined with a 2D (Poisson) or 3D (TOSCA) magnetic field simulation code.

The main characteristic of a bending magnet is to deflect the particle beam by an angle φ . The deflection angle is proportional to the integrated field of the magnet, and the change $\Delta\varphi$ of the deflection angle is given by the change of the magnetic field (ΔB) over the length of the magnet, Δs . For the correction of $\Delta\varphi$ so-called correctors are introduced into the lattice.

$$\varphi = \frac{\int B ds}{B_0 \rho_0} \quad (A2)$$

$$\Delta\varphi = \frac{\Delta B \Delta s}{B_0 \rho_0}$$

The bending angle of the Storage Ring bending magnets of ALBA is 11.25 degrees (or 196.35 mrad). The maximum allowed change of $\Delta\varphi$ is determined by the acceptance of the machine. For instance, in the case of ALBA $\Delta\varphi$ should be smaller than 0.3 mrad, which means that $\Delta\varphi/\varphi \leq 0.0015$ or 1.5×10^{-3} . A typical requirement is that $\Delta\varphi/\varphi \leq 0.001$ or 1×10^{-3} . That is to say, with a constant field B_0 everywhere in the bending magnet the tolerance of the length has to be better than $\pm 0.5 \times 10^{-3}$. For a 1 metre-long bending magnet the resulting tolerance is ± 0.5 mm, which can be achieved.

The change of the deflection angle $\Delta\varphi$ is proportional to the change of the flux density ΔB and to the change of the gap Δg .

$$\frac{\Delta B}{B_0} = \frac{\Delta\varphi}{\varphi} = \frac{\Delta g}{g} \quad (A3)$$

With a gap of roughly 40 mm and the requirement of $\Delta\varphi/\varphi \leq 1 \times 10^{-3}$, the change of the gap has to be smaller than 40 μm . As a consequence the gap height has to be very accurate, in the range of $\pm 20 \mu\text{m}$.

A.1.1 Pole profiles in a bending magnet

In some cases there is a gradient (G) in the bending magnet so as to have a so-called combined function bending magnet. To reach this a slope has to be built into the gap of the bending magnet. The change of the slope is roughly given by the product of the gradient and the gap divided by the field.

$$\frac{\Delta y}{\Delta x} = \frac{G(g/2)}{B_0} = \frac{(g/2)}{X_0}. \quad (\text{A4})$$

For the Storage Ring of ALBA ($G = 5.6 \text{ T/m}$, $g = 40 \text{ mm}$, $B_0 = 1.42 \text{ T}$) the resulting slope is roughly 78.9 mrad or, for a distance of $x = 25 \text{ mm}$ away from the centre, a value of 2.5 mm. With the required tolerance of 1×10^{-3} , the change of the gap should have an accuracy of roughly 2.5 μm . This is pretty tight and a fairly impossible tolerance to be fulfilled.

More precisely, if a gradient has to be introduced into the bending magnet the gap needs a hyperbolic shape (Fig. A3). The change of the half-gap height with the horizontal position is given by

$$h(x) = \frac{h(0)}{1 - \frac{xG}{B_0}} = \frac{h(0)}{1 - \frac{x}{X_0}} \quad (\text{hyperbolic approach}) \quad (\text{A5})$$

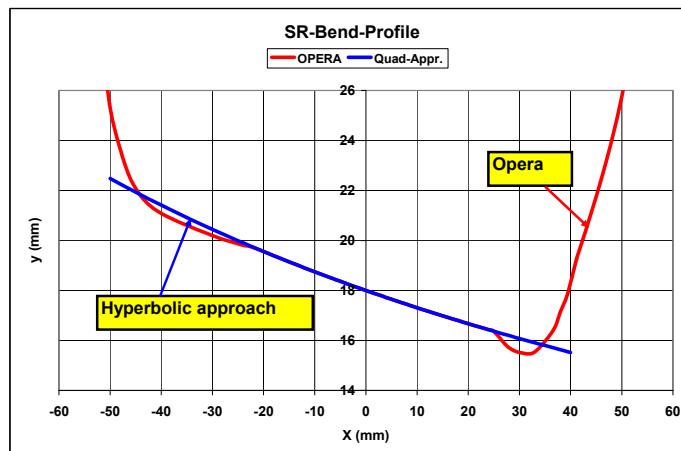


Fig. A3: Pole profile of the gradient bending magnet for the Storage Ring of ALBA determined using a hyperbolic approach (blue line) or using OPERA3D field simulation code (red line)

If a sextupole component B'' has to be introduced into the bending magnet the gap needs a quadratic shape. In this case the change of the half-gap is given by

$$h(x) = h(0) \left(1 + \frac{(x/X_0)^2}{1 - (x/X_0)^2} \right) \quad \text{with} \quad X_0 = \sqrt{\frac{2B_0}{B''}}. \quad (\text{A6})$$

For example, for a sextupole component of $B'' = 40 \text{ T/m}$, the half-gap has to be increased (at a transverse distance of 30 mm) by 0.257 mm. The pole profile can be machined with an accuracy of roughly $\pm 20 \mu\text{m}$, which means that the sextupole component in a dipole does not have a high accuracy (5 to 15%).

A.2 Design criteria of a quadrupole magnet

A quadrupole magnet has four poles. The field varies linearly with the distance from the magnet centre. It focuses the beam within one plane while defocusing the beam within the orthogonal plane.

The field of the quadrupole has to be proportional to the distance from the centre (along either x or y). The excitation in general is given by

$$B_0 = \frac{\mu_0 N I}{g} \quad \text{or} \quad B(x) = \mu_0 N I x . \quad (\text{A7})$$

This means that the pole profile $g(x)$ of a quadrupole has to be a hyperbola, as shown in Fig. A4.

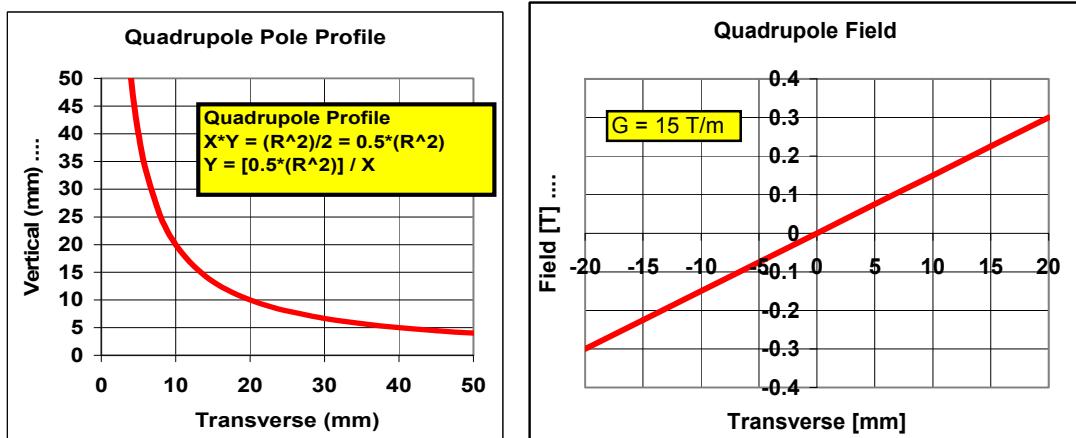


Fig. A4: Pole profile (left) and dependence of the magnetic field along the horizontal direction (right) for a quadrupole magnet

Some examples of quadrupole magnets are shown in Fig. A5.

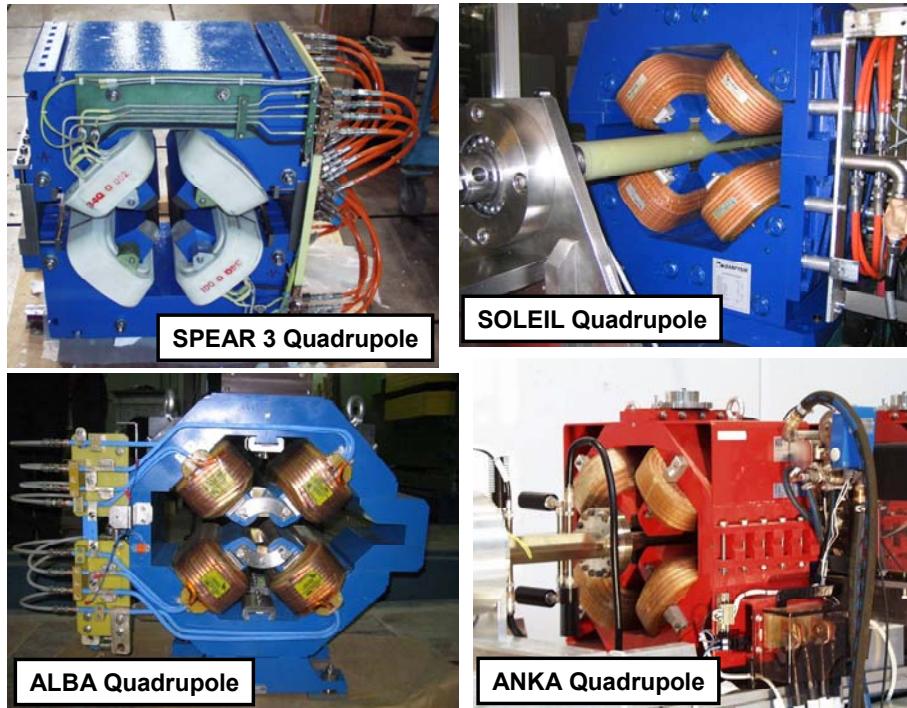


Fig. A5: Examples of quadrupole magnets for different synchrotron light source facilities

By cutting the pole profile in order to have space for the introduction of the coils, the field distribution will be disturbed and higher order systematic multipoles will be introduced. In the case of fully symmetric magnets, the allowed systematic multipoles depending on the number of poles of the magnet are listed in Table A1.

Table A1: Allowed systematic multipoles depending on the fundamental geometry of the magnet, for fully symmetric magnets

Fundamental geometry	Systematic multipoles
Dipole, $n=1$	$n = 3, 5, 7\dots$ (6-pole, 10-pole, 14-pole...)
Quadrupole, $n=2$	$n = 6, 10, 14\dots$ (12-pole, 20-pole...)
Sextupole, $n=3$	$n = 9, 15, 21\dots$ (18-pole, 30-pole...)
Octupole, $n=4$	$n = 12, 20, 28\dots$ (24-pole, 40-pole...)

Assembly errors introduce higher multipoles as well. In this case they are called random multipoles. In the case of a quadrupole magnet some of the possible assembly errors are illustrated in Fig. A6.

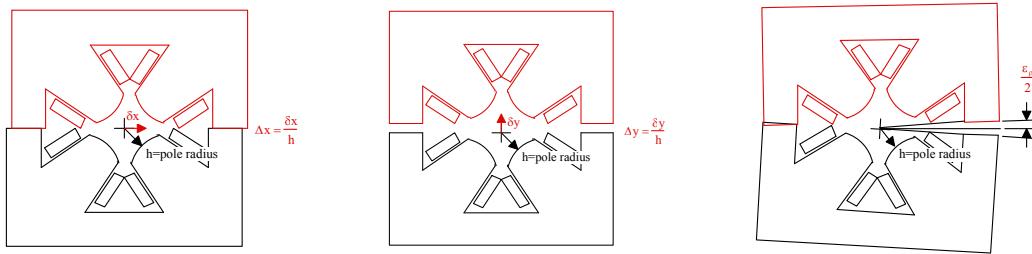


Fig. A6: Some of the possible assembly errors for a quadrupole magnet

Figure A7 shows two possible ways of segmenting the laminations of a quadrupole. In approach (a), each segment can be assembled with errors with three kinematic motions, x , y and e (rotation). Thus, combining the possible errors of the three segments with respect to the datum segment, the core assembly can be assembled with errors with $3 \times 3 \times 3 = 27$ degrees of freedom. In comparison, approach (b) has the advantage that the two core halves can be assembled kinematically with only three degrees of freedom for assembly errors. Thus assembly errors are more easily measured and controlled.

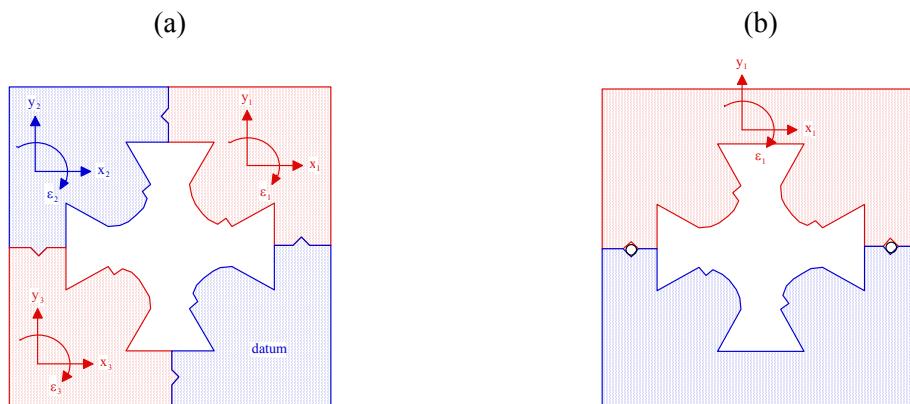


Fig. A7: Two possible ways of segmenting the laminations of a quadrupole magnet

A.2.1 Higher multipoles in quadrupoles

The geometrical parameters involved in the design of a quadrupole magnet and determining its content of higher multipoles are illustrated in Fig. A8. The effect of introducing different modifications of these parameters on the content of higher multipoles is listed in Table A2. It can be concluded that to avoid higher multipoles the manufacturing of the magnets has to be very accurate ($A = B$ and $a = b = c = d$).

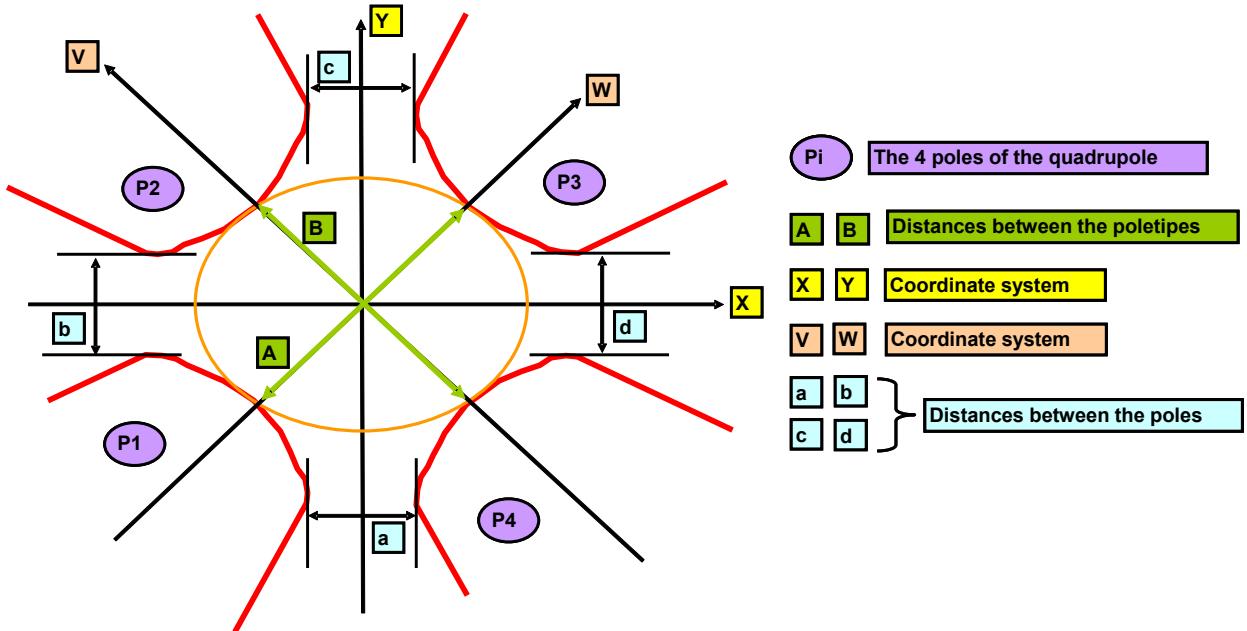


Fig. A8: Geometrical parameters involved in the design of a quadrupole

Table A2: Introduced higher order multipoles for different modifications of the geometrical parameters of a quadrupole magnet as defined in Fig. A8

	Geometrical change	Allowed multipoles
0)	Fully symmetric quadrupole; $A=B$ and $a=b=c=d$ and poles are truncated	$n=6$
1)	P1 in neg. W	$n=3, n=4, n=5, n=6, n=10$
2)	P4 in neg. V	$n=3, n=4, n=5, n=6, n=10$
3)	P3 in neg. W	$n=3, n=4, n=5, n=6, n=10$
4)	P2 in pos. Y	$n=3, n=4, n=5$
5)	P2 in neg. Y	$n=3, n=4, n=5$
6)	P2 in neg. V and P4 in pos. V	$n=4, n=6, n=8, n=10$
7)	P2 in pos. X	$n=3, n=4, n=5$
8)	P2 in neg. X	$n=3, n=4, n=5$
9)	A and B are increased by the same amount	$n=6, n=10$
10)	A and B are decreased by the same amount	$n=6, n=10$
11)	A and B not equal	$n=4, n=6, n=8, n=10$

A.3 Design criteria of a sextupole magnet

A sextupole magnet has six poles. The field varies quadratically with the distance from the magnet centre, as shown in Fig. A9. The purpose of sextupoles is to affect the beam at the edges, much like an optical lens which corrects chromatic aberration. Sextupoles are needed for the compensation of the chromaticity of the machine or, in other words, in order to make the focusing of the machine independent of the energy of the beam in a small range. Note that the sextupole also steers the beam along the 60 and 120 degree lines.

Several examples of sextupole magnets are shown in Fig. A10.



Fig. A9: Dependence of the magnetic field along the horizontal direction for a sextupole magnet

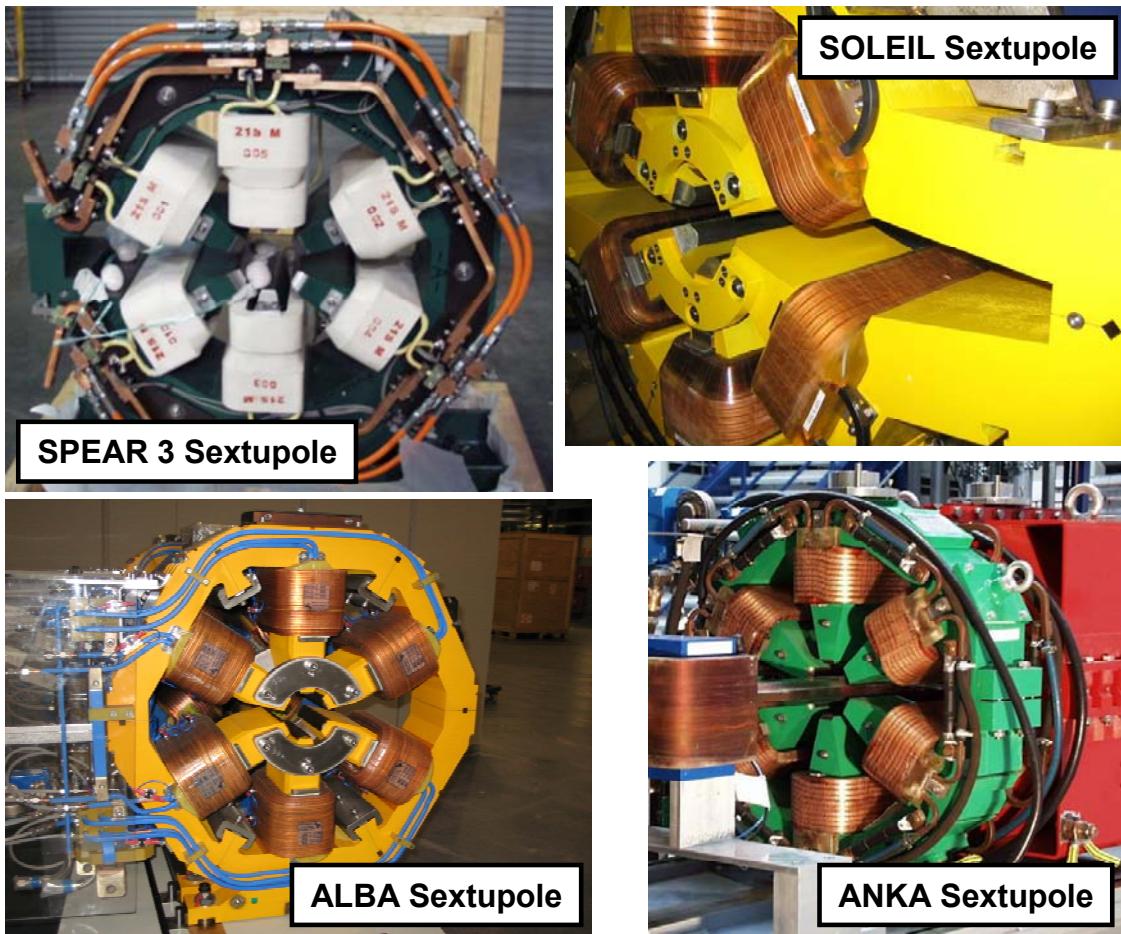


Fig. A10: Examples of sextupole magnets for several synchrotron light sources

A.4 Specification of coils

The current and the number of windings of a coil determine the excitation of a magnet through the following expression:

$$NI = \frac{B_0}{\mu_0} \left(g + \frac{\ell_{Fe}}{\mu_r} \right) \approx \frac{B_0 g}{\mu_0}. \quad (A8)$$

For ALBA Storage Ring bending magnets with a gap of 36 mm and a flux density of 1.42 T the required excitation is

$$NI = 40\,680 \text{ Amp-turns}.$$

This number can be attained either by using a larger number of turns or by increasing the current.

The standard design for magnet coils is rectangular copper (or aluminium) conductor, with a cooling water tube inside, as shown Fig. A11. Either glass cloth or epoxy resin can be used as insulation.

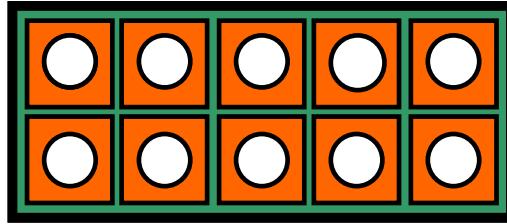


Fig. A11: Cross section sketch of the standard design for magnet coils. The shown coil corresponds to 1 pancake with 10 turns, arranged in 2 layers.

Once the required number of Amp-turns (NI) has been determined, there is still some freedom to choose the total copper area (A_{copper}) and the number of turns (N) of the coils. The choice is made based on economic criteria taking into account the current density (j) circulating through the windings of the coil, given by

$$j = \frac{NI}{A_{copper}}. \quad (A9)$$

The arrangement of the windings within the cross section of the coil has to be decided taking into account the space requirements of the machine. Usually, it is desirable to reduce the space between the components of the machine as much as possible. A small number of layers needs more space in the longitudinal directions, whilst a high number of layers needs more space for the magnets in the vertical direction. Therefore a compromise that will depend on the characteristics of the machine has to be reached.

A.4 Other design criteria

For the design of the magnets one has to take further into account:

- The requirements of the vacuum system.
- The requirements (fixing) of the diagnostics elements.
- The requirements of the front ends.
- Fixing of the magnets on the girders/supports.

Appendix B: Sequence for magnet production

A typical sequence for the production of the magnets of an accelerator includes the following steps:

- 1) The requirements of the magnets (field, length, gradient, sextupole field, good field region, etc.) are established by the Beam Dynamics Group.
- 2) A basic design of the magnets is prepared by the Magnet Group (pole profile, H- or C-type of magnet, cross section of the lamination, coil design, supports, etc.), including the required space needed for vacuum, diagnostics, cooling, etc.
- 3) Steps 1) and 2) have to be made according to the space requirements for the vacuum, diagnostics, etc., and hence some iterations are needed before finishing the basic design.
- 4) At this point one of the following decisions has to be taken:
 - a) Will the detailed design be made in-house by the magnet group/engineering group?
 - b) Will the detailed design be made by the manufacturer?
- In most cases the detailed design is made by the manufacturer. The reason for this is that the manufacturer has to prepare the production drawings anyway and therefore he should also make the detailed design.
- 5) The fulfillment of the specifications can only be checked by means of magnetic measurements. Therefore it has to be decided by whom and where the magnetic measurements will be carried out. This has to be written down in the Technical Specifications for the Call for Tender.
- 6) Going out for the Call for Tender.
- 7) Evaluation of the different offers, including some discussions with the bidders. Choice of the manufacturer.
- 8) Signature of the contract.
- 9) Preparation of the detailed design by the manufacturer.
- 10) Acceptance of the detailed design (coils, manifolds, etc.) by the customer.
- 11) Preparation of the manufacturing drawings.
- 12) Production of stamping tools.
- 13) Procurement of raw materials: steel, copper, etc.
- 14) Acceptance test of laminations.
- 15) Production of prototype or pre-series magnet.
- 16) Magnetic measurement of prototype magnet.
- 17) Determination of end chamfer, acceptance of prototype and agreement of modifications for the series production.
- 18) Series production.
- 19) Mechanical test of the yokes.
- 20) Magnetic and electrical test of the coils.
- 21) Magnetic measurements of the series magnets.
- 22) Acceptance of the magnets.

All the previous points have to be addressed in the Technical Specifications document.

Dimensional metrology and positioning operations: basics for a spatial layout analysis of measurement systems

A. Lestrade

Synchrotron SOLEIL, Saint-Aubin, France

Abstract

Dimensional metrology and positioning operations are used in many fields of particle accelerator projects. This lecture gives the basic tools to designers in the field of measure by analysing the spatial layout of measurement systems since it is central to dimensional metrology as well as positioning operations. In a second part, a case study dedicated to a synchrotron storage ring is proposed from the detection of the magnetic centre of quadrupoles to the orbit definition of the ring.

1 Introduction

The traditional approach in Dimensional Metrology (DM) consists in considering the sensors and their application fields as the central point. We propose to study the geometrical structure or ‘architecture’ of any measurement system, random errors being a consequence of the methodology.

Dimensional metrology includes the techniques and instrumentation to measure both the dimension of an object and the relative position of several objects to each other. The latter is usually called *positioning* or *alignment*.

Dimensional metrology tools are split into two main categories: the sensors that deliver a measure of physical dimensions and the mechanical tools that deliver positions (centring system).

A third component has to be taken into account: time dependence of the measures coming from sensors but also from mechanical units. It is usual to consider measurement systems (whatever the techniques or the methods) as evolving in a pure 3D space. But, if ultimate precisions have to be reached, the system cannot be studied from a steady state point of view: any structure is subject to tiny shape modification, stress or displacement (e.g., thermal dependence). In other words, metrology depends on time.

Finally, space is obviously to be considered. The three-dimensional geometry (affine and vector spaces) is central. From this point of view, we could define the topic as the spatial layout (or topology) analysis of any measurement system. As an introduction, let us consider a metrology loop similar to a tolerance stack-up of a complex mechanical assembly: such metrology loops have necessarily a three-dimensional aspect imposed by the relative position of parts to each other.

Spatial analysis can be used for fields other than dimensional metrology: the design of a magnetic bench, i.e., the choice of the technology (coil, Hall probe, etc.) needs the metrology dedicated to magnetism but the complete bench design will also include a spatial analysis of the whole set-up.

The relationship between the four components of dimensional metrology can be summarized through the schematic in Fig. 1.

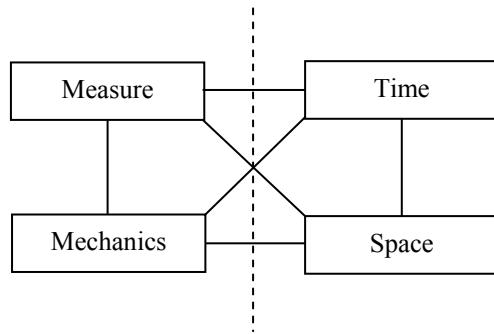


Fig. 1: The components of dimensional metrology

The examples of the lecture come deliberately from various fields and situations, the goal being to underline their common features and not their differences.

2 The sensor

Let us define the sensor in the framework of this course: the ‘sensor’ is the whole data acquisition chain from the physical detection to the output value, usually in a digital format (Fig. 2). However, any sensor includes a material (mechanical) part, even if it is minimal.

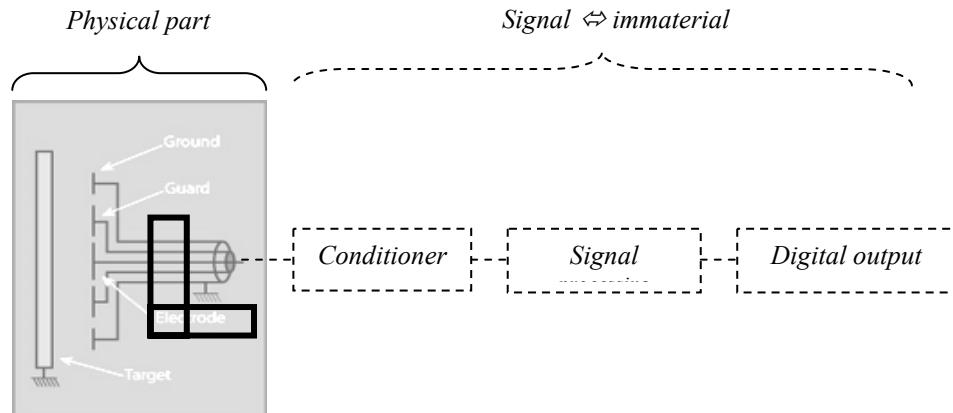


Fig. 2: Capacitive sensor for non-contact measurements with its data acquisition

2.1 Time-dependent behaviour of the sensor

The behaviour of the sensors regarding their time dependence will be linear, as a zero-order for acceleration: $y(t) = Kx(t)$, $x(t)$ and $y(t)$, being the physical detection and the digital output, respectively. The response time of the measurement device is instantaneous.

2.2 Categories of errors in the domain of the measure

The errors can be distinguished according to their origin:

- i) Random errors: in the case of the sensors, they correspond to microscopic effects coming from the devices of the measurement chain (detection, signal processing). Their magnitude is usually small, e.g., electronic noise of the micro-component of detection.

- ii) Bias errors (also called systematic errors): an offset, i.e., a bias can exist in the result of a measurement procedure. As an example, the zero value of a sensor is rarely well known. A lack of linearity can also affect the sensor. In the field of mechanics, bias errors include a misalignment between components or a shape defect of an object. The bias errors do not depend on time and their magnitude can be important.
- iii) Errors depending on external sources: they can change the environment conditions of a measuring or mechanical system. Typical examples are the temperature variations during measurement or the use of mechanical assembly, the ground settlement of a long structure such as a particle accelerator. These two examples show that the range of such external influences is extremely variable. Either it is possible to have a model of their influence and one can subtract it from the result coming from the system, or they are treated as random errors.

2.3 Statistical model

A ‘measurement’ is symbolized by a pair of figures: the measure itself and the corresponding estimation of its uncertainty due to the random errors. Here are recalled the definition of the main statistical terms: average and standard deviation of a set of measurements:

$$M = \frac{1}{n} \sum_{i=1}^n m_i . \quad (1)$$

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (m_i - M)^2} . \quad (2)$$

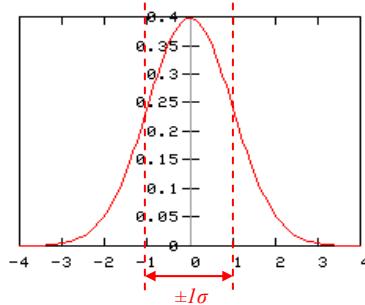


Fig. 3: Normal distribution of random errors

We shall restrict the study of random errors to these first-order statistical moments. Here σ is used as an estimator of the accuracy of a measurement (Fig. 3). In the case of poor redundancy, the metrologist has to estimate the accuracy according to his own experience or from the supplier. The tolerance is defined by $T = \pm 2\sigma$. Let us assume that the measurement system is affected by a set of random errors. Let us also assume that these errors are fully independent of each other. Then, the law of random errors combination is as follows:

$$\sigma_{tot} = \sqrt{\sum_{i=1}^n \sigma_i^2} . \quad (3)$$

The hypothesis of independence of the random errors is often very reliable in the field of DM.

3 Mechanical aspects of dimensional metrology

As mentioned in the introduction, mechanics is a full component of DM as a positioning system. Even sensors include a minimum part of mechanics: the core of an electrode, its supporting part, etc. More generally, we have to consider the necessary physical part which materializes the function of a component to be positioned: the yokes, centring, and supports of a magnet for example, are dedicated to the magnetic function necessary to control the beam orbit of an accelerator. Nevertheless, it is of major importance to keep in mind that only the function of a component has to be positioned even if it is usually immaterialized, as a magnetic axis of a magnet: we do not align its yokes, even less its support.

3.1 Categories of errors in the domain of mechanics

3.1.1 Random errors

The mechanical units may be subject to uncertainty of their dimension and to external source solicitations: an assembly in a hyperstatic situation can be bent. The spatial set-up of a mechanical assembly appears at many different levels. As examples, first, the direction of a shaft depends on the clearance of the bore hole (Fig. 4). Let assume $\sigma = 10 \mu\text{m}$ as the clearance of the bore hole, then its influence on the shaft direction is $\sigma_a = \sigma/l$ and the Z uncertainty at the point A is $\sigma_z = L \cdot \sigma_a$.

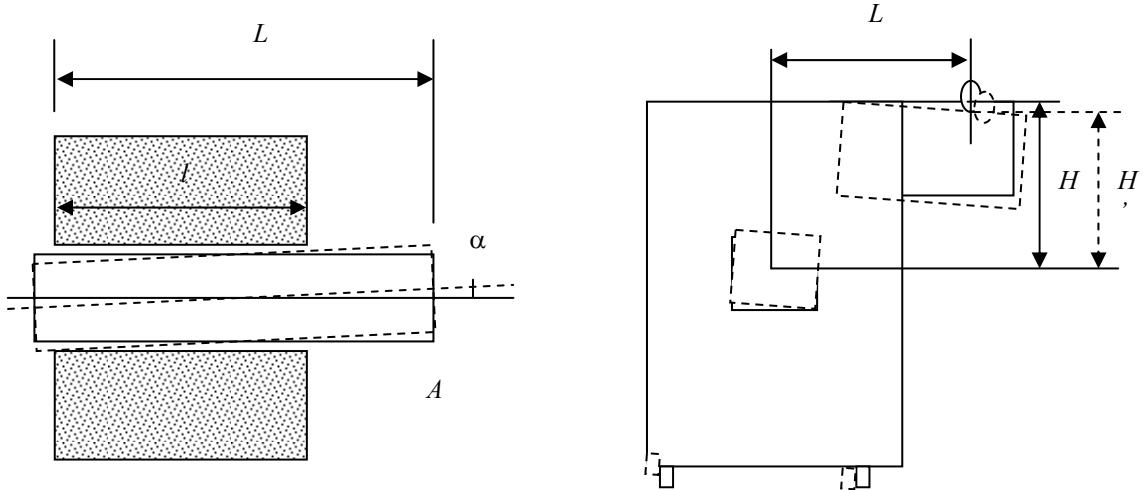


Fig. 4: Clearance of set shaft-bore

Fig. 5: Dependence on the lever arm

The height H between the two points of the mechanical assembly in Fig. 5 depends on the horizontal lever arm L . If there is a probability σ_θ of parasitic rotation, then $\sigma H \approx L \cdot \sigma_\theta$. The accuracy of machining is equivalent to random errors in the field of measurements. For that reason the tolerance stack-up of an assembly is similar to the law of error combination of measurements. However, the tolerance is preferred to the standard deviation in mechanics for practical reasons.

3.1.2 Bias errors

If a mechanical assembly is measured after having been machined, then the difference with respect to the nominal dimension called ‘offset’ can be used as a bias error regarding the whole assembly (Fig. 6).

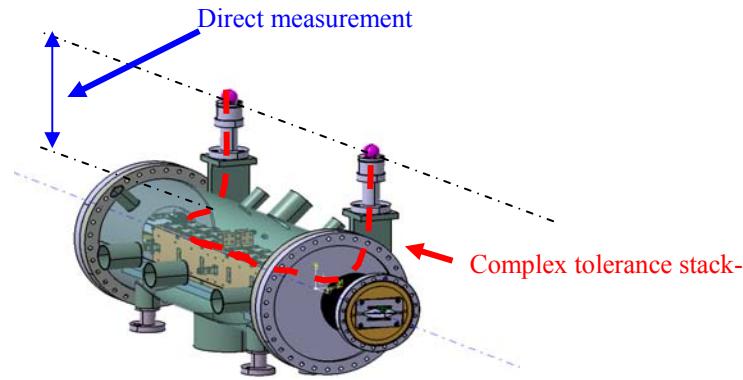


Fig. 6: Tolerance stack-up and direct measurement

3.2 Analogy between measure and mechanics

There is a formal analogy between measure and mechanics in DM. We can thus mix both sensors and mechanical units then apply the law of random errors combination and finally mix the two modes in a metrology loop. The concept of redundancy of the measures may be extended to the purely mechanical parties; there is hyperstatism in the function of positioning. Nevertheless, fully mechanical hyperstatic positioning is difficult to manage quantitatively because the mathematical modelling of ‘physical’ parts is not sufficient in comparison with that of the measures which are ‘immaterial’. It only brings a good stability to the assembly. Conversely, a network of measures between points of a structure is, in most cases, very redundant because the theory of probability random errors applies perfectly. One can notice that applying the least-squares principle ($\sum v_i^2 \min$) in the field of measures corresponds to a minimum of energy of a mechanical system at equilibrium.

A polygon which has been measured in angle and distance is equivalent to a mechanical assembly similar to a wheel. If the measurement of the angles is weak, the wheel tends to be deformed. Strengthening the network by measures of the distance of diagonals improves its sensitivity to errors. The assembly is now more rigid (Fig. 7).

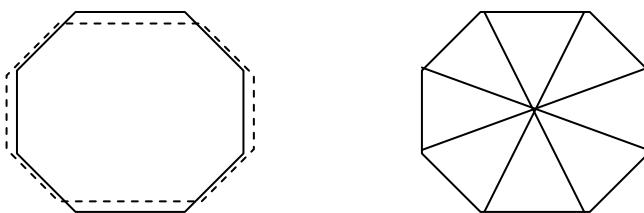


Fig. 7: Analogy between measure and mechanics

The analogy measure-mechanics has its own limit because of the fundamental causal relationship between both: mechanical units are made of parts which are somehow measured. After an ultimate and exhaustive analysis, one finds that the realization of the mechanical part is only a materialization of a measure by its machining step via a complex chain of actions. Similarly, carrying out the position of a mechanical set is done from preliminary measurements. However, in these two cases, the measure by which they are associated is necessarily more precise since there is always the causal relationship. Finally, after the manufacturing step, machining or adjustment at best, a measure is carried out to know the true situation (Fig. 8).

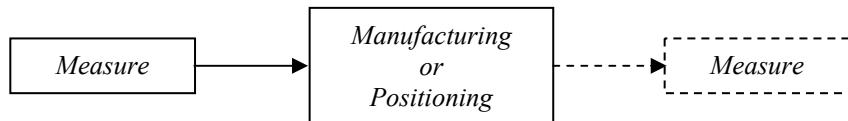


Fig. 8: Relationship between measures and mechanics

3.3 Purely mechanical aspects

The mechanical design can have a direct impact on the aspects of positioning at different stages (machining, assembly, and adjustment): the links between the sub-assemblies strongly involved in the positioning; clearance between centring systems; rolling elements not repeatable in position, etc.

- i) Hyperstatism of mechanical assemblies can have a negative influence on the positioning in terms of position repeatability. Some static mechanical mounts require a hyperstatism in the functional configuration (to limit the vibration) and an isostatism for setting operation in position. Changing from one to the other cannot necessarily ensure good repeatability of position.
- ii) The weight of the equipment itself must sometimes be taken into account for the machining of a large surface (Fig. 9): a beam support for magnets, whose upper face is intended to be perfectly flat, will be presented to the milling machine with its supporting points as laid down in operation; the sag due to its own weight is thus eliminated by the machining.

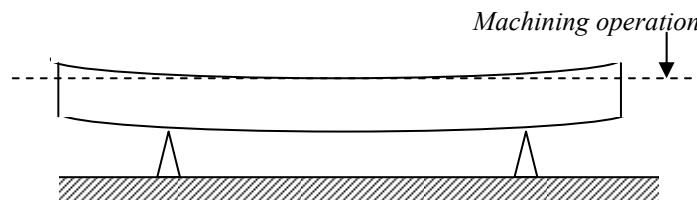


Fig. 9: Elimination of own weight of a girder

- iii) The settings of a mechanical assembly must be made in the configuration of equilibrium closest to that in operation. It is necessary to mount all the parts, especially if they are heavy or off-set.
- iv) Some parts of an accelerator must be baked out at high temperature to obtain a good quality of vacuum in the chambers. The temperature can be higher than 200°C, affecting the relaxation of the heated parts. Consequently, it is better to avoid any accurate measurement before bake-out.

4 Stability time constant

4.1 Definition

The use of a dimensional measuring instrument must take into account the stability of the set including both the instrumentation and the object to be measured or positioned. In other words, the stability analysis of the set is mandatory if ultimate accuracy is required: the stability should be at least of the same dimensional scale as the instrument precision. It is a well-known issue for metrologists, but we propose a formalization by using the concept of Stability Time Constant (STC). The STC is defined

by the acceptable duration δt during which we do not want less than a parasitic displacement quantity δd (Fig. 10):

$$STC = (\delta d, \delta t) \quad (4)$$

with $\delta d \ll \delta m$ during the duration δt , δm being the measurement accuracy.

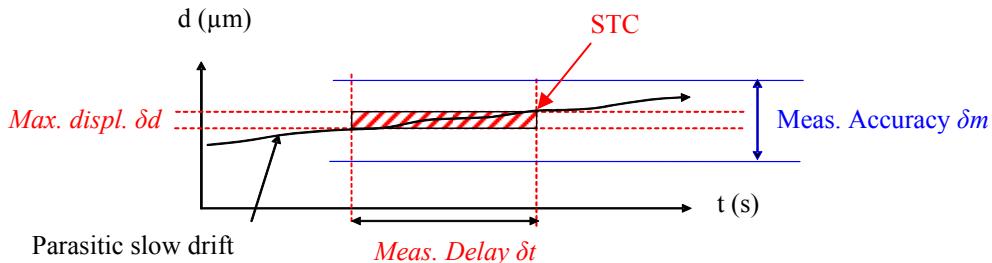


Fig. 10: Stability time constant

This concept can be applied whatever the origin of the disturbance of the system: mechanical, electronic, etc. The main interest of this concept is to keep in mind stability analysis at any step of the design of a measurement procedure. The STC can be defined for the six differential Degrees Of Freedom (DOF) between instrument and object.

4.2 Angular measurements with a theodolite

A theodolite measures angles defining the difference between two directions with, say, 3.10^{-4} deg accuracy. It includes a graduated circle of about $\phi 70$ mm as an angular encoder. The duration for measurements could be $\delta t = 30$ min when there are many directions to be measured (Fig.11).

Then $STC_\theta = (3.10^{-4} \text{ deg}; 30 \text{ min})$.

Here 3.10^{-4} deg represents only $0.2 \mu\text{m}$ displacement of the graduations on the circumference of the angular sensor around the vertical axis. It shows the high level of stability required for such angular measurements.

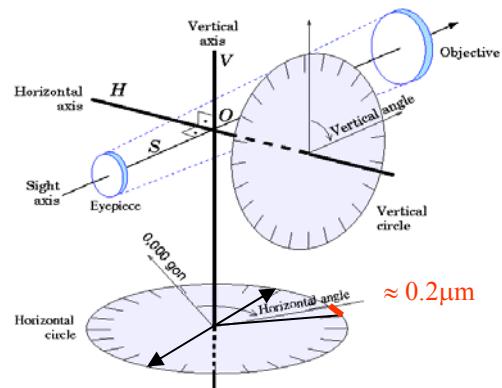


Fig. 11: STC of an angular theodolite measurement: $0.2 \mu\text{m}$ stability is necessary

4.3 Slope measurement with an inclinometer

An inclinometer measures tight slopes with respect to the horizontal plane. The measurement itself is directly linked to gravity by means of its internal sensors. Common sensors can produce measures within $10 \mu\text{rad}$ accuracy. The inner electronics needs to be stable for good results and the mechanical assembly whose physical quantity is around millimetres may vary slowly due to the thermalization of the instrument (Fig.12). Then a slow drift is always present for such instruments whose effect is equivalent to an offset on the measure origin (zero reading).

We will demonstrate in Section 7 that the half-difference of the two sensor outputs, m_1 in a position of the inclinometer and m_2 after having rotated it by 180° , removes any offset of the measurements [(Eq. 8)]: α is measured without offset in a minute (Fig. 12). If the measurement is carried out during the ramp of the thermal dependence curve of the inclinometer, there is no issue if the reversal is done immediately, unlike when the delay is important. Then $STC_{\theta z} = (10 \mu\text{rad}; 1 \text{ min})$. This is far easier to reach than the stability conditions in the example of the theodolite (Section 4.2).

In certain cases, the inclinometer can be installed for long-term measurements of a mechanical structure on which it is fixed without any capability of rotation around its vertical axis. The STC is then $(10 \mu\text{rad}; \infty)$, the ‘infinity’ symbol meaning a long duration. It is clear that these conditions are very difficult to reach for accurate measurements.

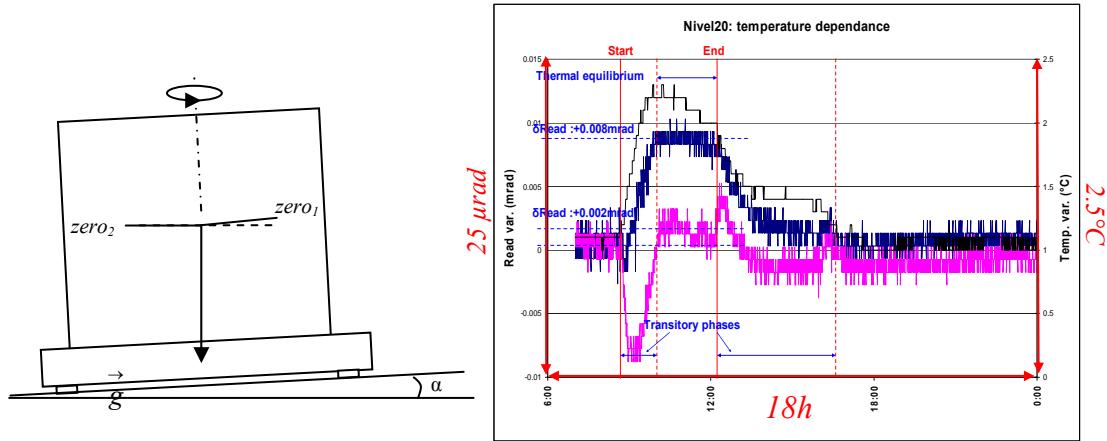


Fig. 12: Inclinometer rotation and thermal dependency of NIV20

The totality of instrumentation is concerned, whatever the origin of the instability (electronics, supporting mechanics, etc.). That is why every user has to regularly (δT) calibrate his instruments to keep its accuracy (Acc). The corresponding $STC = (Acc; \delta T)$.

4.4 The control of the stability issues

Many methods allow the STC to be reduced. Here are the main ones to be applied as soon as possible:

- short duration for the measurements,
- minimization of the metrology loop (Section 6),
- reversal of the instrument,
- regular calibration.

4.5 Choice of a referential for STC analysis

Stability analysis with STC requires a referential to define the six DOF of the system to be analysed. It is usually given by the relative position between the different parts of the unit: the distance between the object and the instrument, their relative rotation, etc. Thus in the examples of Sections 6.2 and 6.3 (translation stage), the best referential is the one defined by the direct trihedral whose abscissa axis is confused with the translation of the stage. In the example of the theodolite aiming at several targets, since the STC involves many points together, a general referential is used and described as an ‘absolute’. Absolute usually means that it is linked to the Earth but the main feature of a referential is to allow the right description of the DOF of the system.

Note that the referential for calculations, if required, can be different from the one for the STC analysis. There is a particular case which involves a true absolute measurement: gravity is used for inclinometry and altimetry measurements. The metrologist has to pay attention to what can define his referential: keep it absolute or leave it as relative by differential measurement.

5 Geometrical frame of dimensional metrology

5.1 Angle-length duality

The physical space where we live is mathematically modelled by an affine space with three dimensions. The study of the mathematical operators existing in an affine space leads one to be interested in the only class of interest for DM: the displacements, i.e., rotations and translations. There is only one physical dimensional quantity in an affine space, the length, ‘quantity **with** a dimension and with a unit’, the metre. That is not the case of the angle, ‘quantity **without** dimension and with a unit’, the radian. A basic geometrical figure, the triangle, resumes clearly these properties. Three quantities are enough to define a triangle, except for the case where the three angles are known: its shape can be defined and not its dimensions.

B. Schatz wrote about the metrology of angles [1]: “*The abstract nature of the angle unit assumes an absolute accuracy. In addition, it stays available for everybody in the laboratories as in the industry. But difficulties arise when it is necessary to materialize this standard by an instrument. It is at this level of considerations that appears a certain duality, both in the means and in the methods of control, between:*

- *the means and methods attached directly to the abstract nature of the angle unit,*
- *the means and methods attached more directly to the unit of length and indirectly the angle unit by its trigonometric lines, those angles are thus generated by variation of length. Due to this duality, the metrology of angles is deeply linked to the metrology of the lengths.”*

Methods vary according to the environment and accuracy: one can use angles to determine lengths and one can determine angles with the use of lengths (Fig. 13). As an example, alignment, i.e., the positioning of the components of a (quasi-)linear structure such as an accelerator, can be reached either with angle measurements of a theodolite placed in the continuation of structure corresponding to tangential measurement, or with distance measurements of an EDM on the edge of the structure corresponding to radial measurements.

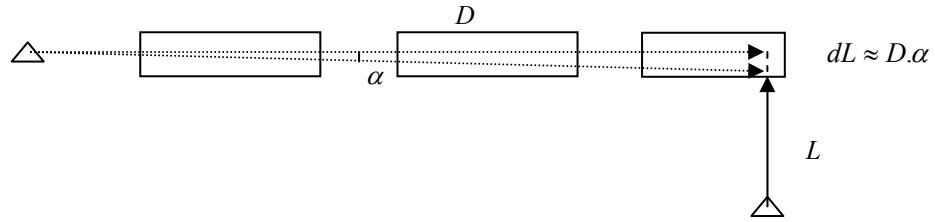


Fig. 13: Angle–length duality

This duality occurs even in the calculations after measurement. When an operation requires the two kinds of quantities, a least-squares calculation may be useful to the quality of the results. One can use the kind of quantities, angle or distance, in the equations describing the system. The following choice appears: equations homogeneous to angles or equations homogeneous to distances.

In some cases, the metrologist, as a user, may have to consider a vector space without length quantity, even if finally any use of angle requires that of a distance. When one wants to measure the direction of a vector, whose materialization is given by two points, then the centring of the instrument on a point and the aiming at the target of the other one, both intervene directly in the measure. Using a vector defined by two points means working into an affine space (Fig. 14).

Conversely, any measure involving autocollimation (guidance of a mirror plan, reciprocal collimation of instruments, etc.) means working in a vector space since this method is completely free of the concept of distance at the level of the user (Fig. 15); instruments and/or mirror can be positioned anywhere in the space to carry out this measure, provided that there is enough beam for the quality of measure.

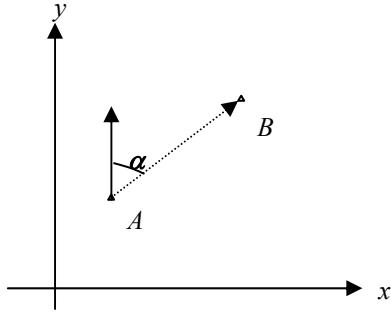


Fig. 14: Orientation and affine space

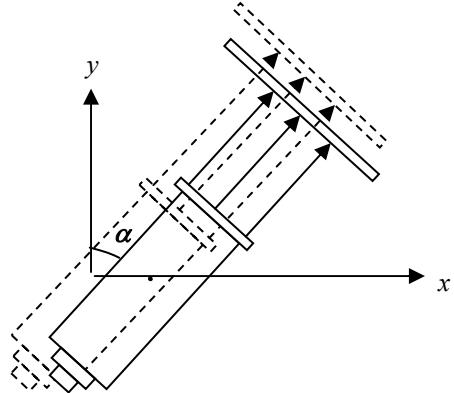


Fig. 15: Autocollimation and vector space

5.2 Sensitive direction

Many sensors have a unique direction of measurement. The typical example is the dial gauge: any transversal component of the part displacement cannot be detected by a dial gauge (Fig. 16). The sensitive direction of the sensor is given by its longitudinal axis. A second sensor, orthogonal to the first one, can be set for more complete information on the displacement; it is equivalent to (dx, dY) coordinates. In the latter case, one can calculate the displacement of the part in any other direction which can be the one of the user by the following equation:

$$R(\theta) = dX \cos(\theta) + dY \sin(\theta) . \quad (5)$$

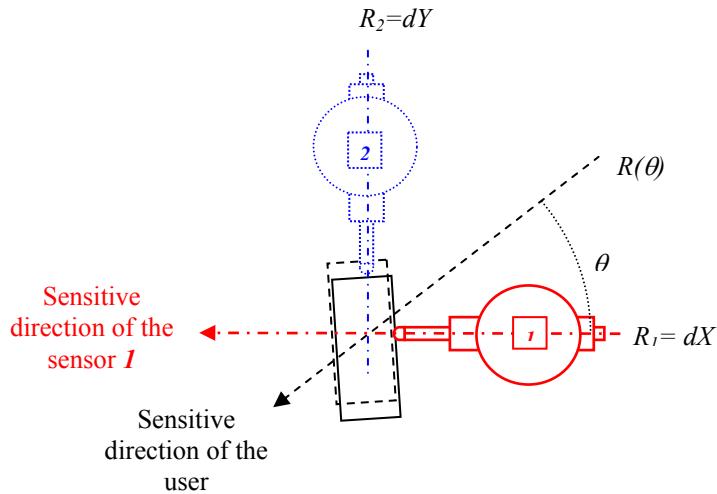


Fig. 16: Sensitive directions of the sensor and of the user

5.3 Spatial intersection

The theodolite enables 2D or 3D spatial intersection by measuring angles from two or more known observational stations to a point to be determined [(Fig 17 (a)]. That approach is very common, in DM as well as in other fields such as astronomy, physics, optics, etc. More generally, having several ‘points of view’ onto a problematic gives more consistent information.

Photogrammetry based on digital cameras is widely used at CERN for large physics detectors such as ATLAS or CMS [2].

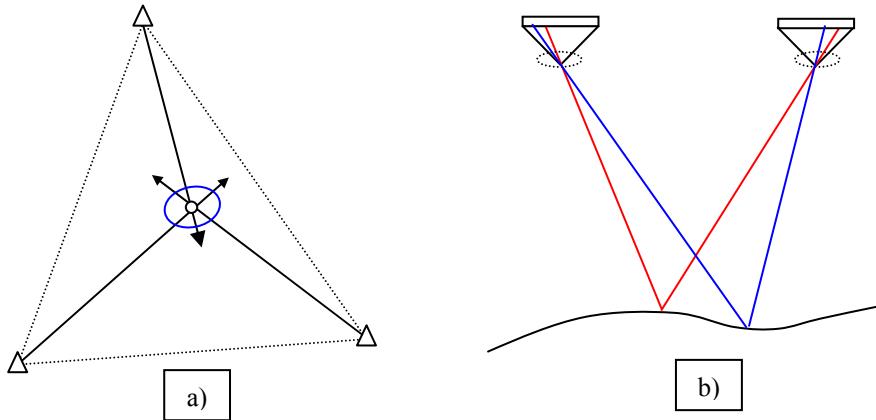


Fig. 17: (a) Angular intersection and error ellipse with theodolite measurements,
(b) schematic of photogrammetry

In the field of surveying, spatial intersections tend to have the same accuracy in all the directions of, say, the plane in 2D geometry. One defines the error ellipse of the point to be determined in the plane as the probability to find it inside with 86% confidence (Fig. 17). This probability corresponds to $T = \pm 2\sigma$ (see Section 2.3) which is the standard deviation of a set of measurements applied to 2D. A ‘good’ layout in standard surveying must give an ellipse closed to a circle whose radius is as small as possible.

But what is linked to the concept of sensitive direction is the fact that sometimes one cannot be interested in an isotropic information: aligning quasi-linear structures such as an accelerator leads to intersections with angles close to zero or to π radians (Fig. 18). The small axis of the error ellipse and therefore the sensitive direction of the angle instrumentation used are then in the radial direction of the accelerator beam axis for alignment purposes.

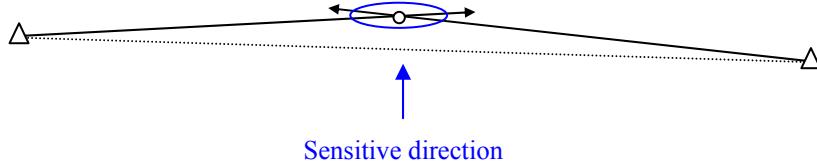


Fig. 18: Angular intersection and error ellipse of the point to be determined in the accelerator field

The use of a laser tracker with its accurate interferometer leads to the same principle of error ellipse and sensitive direction applied to distance measurements carried out in the normal direction of the beam axis.

5.4 Effective length

In the case of small angle measurement like inclinometry, the concept of ‘effective length’ is very useful. The mechanical design of such an instrument requires a stack of interfaces from the sensor to its stands in contact with the part to be measured. The user, who has to estimate the accuracy or the STC of a system, must keep in mind the fact that whatever the length of the part to be measured, the effective length of the physical detection is at most equal to the size of the sensor body. It is actually usually far smaller for inclinometers. Monitoring a structure with a fixed inclinometer at the microradian scale needs the corresponding stability for a very long duration: $STC = (\mu\text{rad}, \infty)$. That stability constraint, being applied for the whole stack of mechanical interfaces, is obviously true for the smallest one, which is usually the detector itself. Such support is about 10 mm: that is the effective length of physical detection. Reaching 1 μrad with a 10 mm lever arm is equivalent to limiting any parasitic vertical displacement of the detector bigger than 10 nm. Conversely, using HLS (non-contact displacement sensors) on a free surface of water for inclinometry (Fig. 19) can achieve easily 0.1 μrad because the effective length is the distance between the sensors, i.e., some hundred metres.

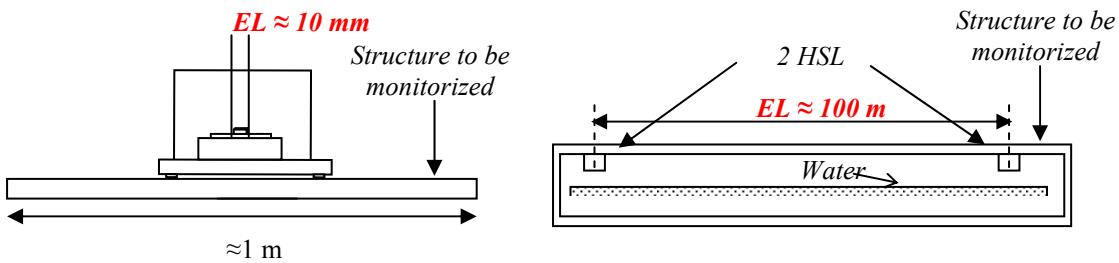


Fig. 19: Effective Length (EL) in inclinometry

Another example is the case of lamination machining for a dipole magnet (Fig. 20). The shape or size tolerance is typically ± 0.02 mm required by the designers for any point of the lamination perimeter. A usual mistake is to believe that the accuracy of the mechanical tilt (rotation around the beam axis) of the magnet is $0.02/L = 0.025$ mrad because the lateral range of the yoke is its width $L = 786$ mm. The effective length for the electron beam is actually the width of the pole $p = 128.6$ mm which is smaller. Then, the tilt accuracy due to machining is $0.02/p = 0.156$ mrad (per pole).

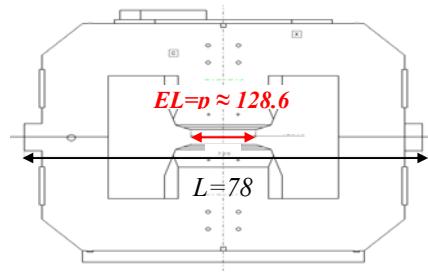


Fig. 20: Effective Length (EL) of a dipole tilt

5.5 Sine and cosine errors

The sine error usually called Abbe error, is the one committed in a length measurement whose reference axis is shifted with respect to its displacement axis or to the contact on the object to be measured (Fig. 21): any parasitic rotation θ between the instrument and user axes induces a ‘sine’ error on the result:

$$e_{\sin} = d \cdot \sin \theta \quad (6)$$

where d is an offset between the two axes called ‘Abbe length’.

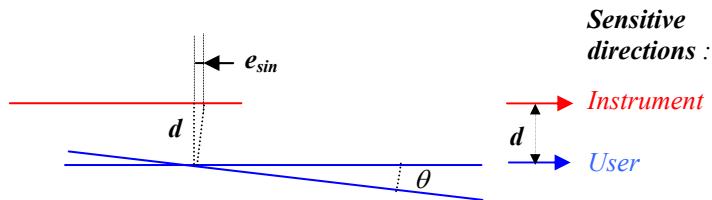
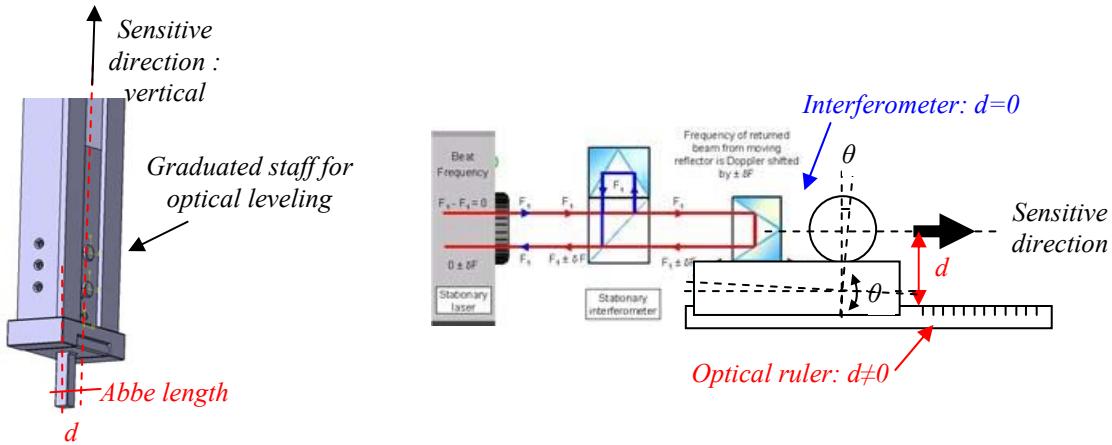


Fig. 21: Abbe or sine error

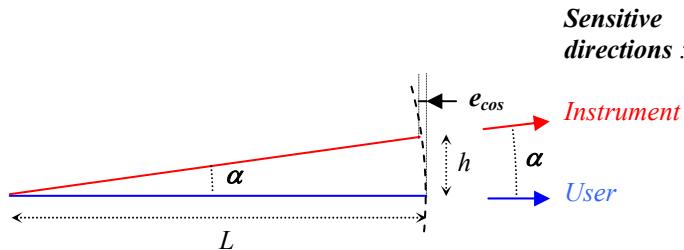
Ernst Abbe (1840–1905) wrote: ‘to realize a good measure, the measurement standard must be placed in the same line as the dimension to be checked’. In Fig. 22, the interferometer used to calibrate the unit is properly set at the same level as the point to be checked ($d = 0$) while the optical ruler shows an Abbe length d , sensitive to any parasitic rotation θ of the stage with respect to the optical ruler.

**Fig. 22:** Layouts with Abbe error

This error is often accompanied by its twin, called ‘cosine error’ and arising if the sensitive direction of the sensor is not parallel to the dimension to be checked (Fig. 23):

$$e_{\cos} = L \cdot \cos \alpha \approx \frac{h^2}{2L} \quad (7)$$

where α is an angle between the two axes.

**Fig. 23:** Cosine error

5.5.1 Calibration of a linear displacement sensor using an interferometer

Consider the previous example of the interferometer checking a translation stage without any Abbe error. Suppose now that the Abbe length is $d = 2$ mm. That length applied to a maximum rotation $\theta = 0.10$ mm/m due to the pitch of the sensor head when it moves on the optical rule gives an Abbe error of about: $e_{max} = 0.002$ mm.

On the other hand, if the direction of the interferometer is rotated by $h = 1$ mm applied to $L = 50$ mm to be measured, then the cosine error is

$$e = \frac{1^2}{2 \times 50} = 0.010 \text{ mm. In this example, the cosine error prevails.}$$

5.5.2 Combination of sine and cosine errors

If both errors are combined, the geometric formulation corresponds to a rotation in 2D (Fig. 24) and the matrix formulation is

$$M' = \begin{vmatrix} x' \\ y' \end{vmatrix} = \begin{vmatrix} x \cos \alpha - y \sin \alpha \\ x \sin \alpha + y \cos \alpha \end{vmatrix} \approx \begin{vmatrix} -y \sin \alpha \\ x \sin \alpha \end{vmatrix}_{x \approx y} \quad (8)$$

where (x') and (y') are the horizontal and vertical sensitive directions respectively.

The sine errors prevail if $x \approx y$.

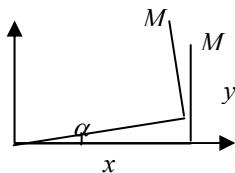


Fig. 24: Combination of sine and cosine errors

5.5.3 Fiducialization of a quadrupole magnet

Any fiducials of an accelerator component show such lever arms (d and h in Fig. 25) with respect to the magnetic axis, and therefore sine errors versus θ the rotation around the beam axis (tilt).

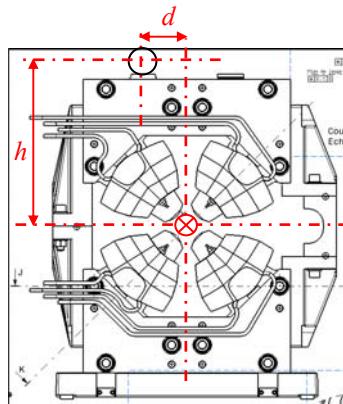


Fig. 25: Lever arms of quadrupole fiducials

5.5.4 The control of lever arms

The Abbe principle being difficult to apply, it is necessary to control the lever arms d and the parasitic rotations θ even in the three dimensions of space.

Three classes of solution exist according to the system of measurement to limit the sine error:

- i) The sensor is in the axis of measurement ($d = 0$).
- ii) Several sensors are placed symmetrically with respect to the axis of measurement; the sine function being odd, the average of the measurements delivered by the sensors corrects the error.
- iii) An angular sensor manages any parasitic rotation.

Two classes of solution exist according to the system of measurement to limit the cosine error:

- i) The sensor is in the axis of measurement ($h = 0$).

- ii) An angular sensor manages any parasitic rotation.

It is important to note that multiple sensors placed symmetrically with respect to the axis of measurement do not correct this error, the cosine function being even.

6 Metrology loop

Any system dedicated to positioning or requiring a positioning operation for its good working consists of a succession of mechanical parts and/or of sensors. This succession of elements is called ‘metrology loop’: it can be seen as the support of position information transmission [3]. Several kinds of metrological chain are used according to whether it is dedicated to the position by mechanical means or is a set of measures spatially distributed. The first one is sometimes called *mechanical, assembly or actuator loop* according to the case; the second one is sometimes called *measuring loop*. Finally, if we must analyse a system of positioning from the point of view of stability in order to integrate the effect of a parasitic slow drift, then it is necessary to take a *stability loop* into consideration. However, that vocabulary must remain flexible; any rigorous application of these names fails with the following example: a measuring loop necessarily requires mechanical parts.

The *metrology loop* is the best structure for a calculation of error budgets, bias errors, and STC analysis. By using the word ‘*structure*’ we can feel the relationship with a spatial layout.

6.1 Metrology loops dissociation

The mechanical units which need both kinematics and measurement as a set point must be designed with fully independent actuators and measuring loops. That is the only condition so that the measuring system can detect any true movement coming from actuators. The modern Co-ordinate Measuring Machines (CMM) are now designed with the condition that the two loops are fully disconnected from each other. The term ‘*dissociated metrology*’ is used in that case [3], [4]. Applying the dissociation principle brings some interesting aspects in other cases: if the precision of a unit dedicated to positioning cannot be reached by mechanical means, then it is sometimes possible to replace it by a measurement.

6.2 Translation stage with coaxial micrometer

The reading of the micrometer is used as set point information for the translation of the stage to the required position (Fig. 26). The actuator and measuring loops are coaxial, that is they correct to eliminate Abbe error, but they are not completely independent: the backlash of the displacement screw of the stage does not appear in the measuring loop that induces a limit to the positioning accuracy.

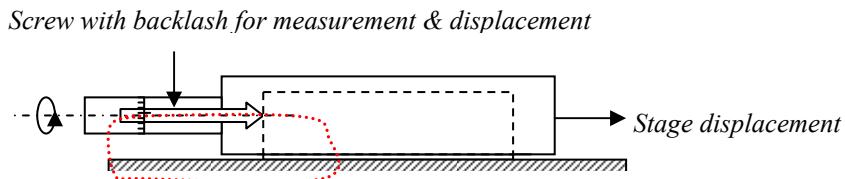
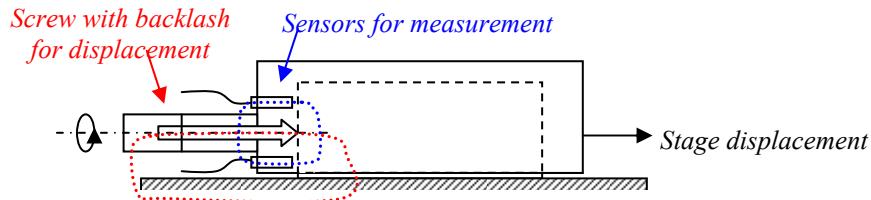


Fig. 26: Translation stage with coaxial micrometer

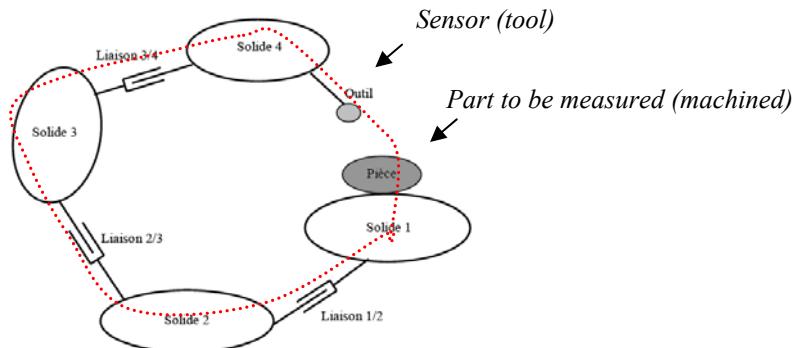
6.3 Translation stage with independent sensors

Both loops are clearly independent in Fig. 27. The sensor can detect the screw backlash delivering thus an accurate set point for the displacement. However, if the backlash induces rotations, then one sensor only is not sufficient.

**Fig. 27:** Translation stage with independent sensors

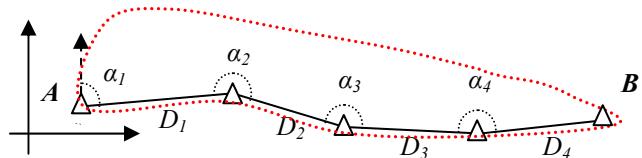
6.4 Co-ordinate measuring machine and machine tool

This kind of topology called ‘series’ is costly in terms of accuracy on the relative position between the sensor and the part to be measured, or the tool and the part to be machined (Fig. 28). All the intermediate parts of the metrology loop must be known accurately and thus for the six DOF of any mechanical subset.

**Fig. 28:** Serial layout [3]

6.5 Polygonal traverse for surveying

This topology is fairly usual in the field of surveying (Fig. 29). All the distances D_i and the angles α_i are measured from A to B . One can clearly see the analogy with the previous mechanical example. This one shows the disadvantage of such a ‘series’ topology: the errors pile up all along the traverse.

**Fig. 29:** Polygonal traverse

6.6 Hexapod

This kind of mechanical structure is called ‘parallel’ (Fig. 30). It is much less sensitive to the errors of positioning since the position of the upper stage depends on each of the six legs of the hexapod. It is

important to note that this structure is not hyperstatic: it is strictly equivalent to the six DOF of a solid in space. Decoupling actuators and measuring loops is extremely effective in such topologies [3].

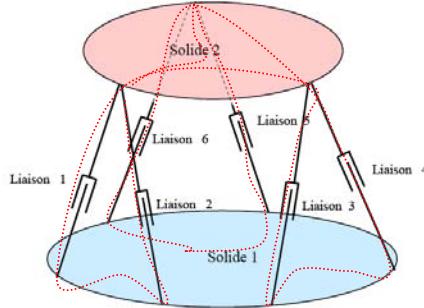


Fig. 30: Parallel layout [3]

6.7 Optical measurement

When measuring a part (say, its shape) by means of an optical instrument separate from the support of the part (Fig. 31), the metrology loop includes:

- the instrument with the target to be used to define the points on the part,
- their respective supports (tripod, target support),
- the volume of the air on the optical path of the measurement of its mechanical interface,
- the support of the part,
- the ground.

Any mechanical part and any measurement which could produce errors has to be included in the metrology loop. Both STC and random errors analysis must be analysed. The metrologist has to choose his level of detail.

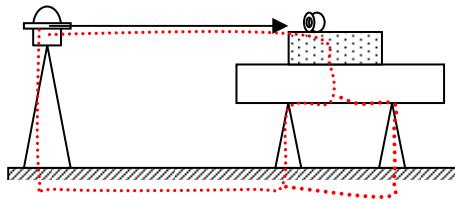


Fig. 31: Optical measurement separate from the part

6.8 Bench for magnetic measurements

A measurement bench is a facility dedicated to the calibration of series of components. In terms of positioning it locates the position of the functional part of the component with respect to a mechanical reference externally accessible for further positioning operations. The bench shown in Fig. 32 is dedicated to the detection of the magnetic axis of quadrupole magnets. Once the axis is detected by the sensor, one extends this position information to the upper yoke: this last step is called ‘fiducialization’.

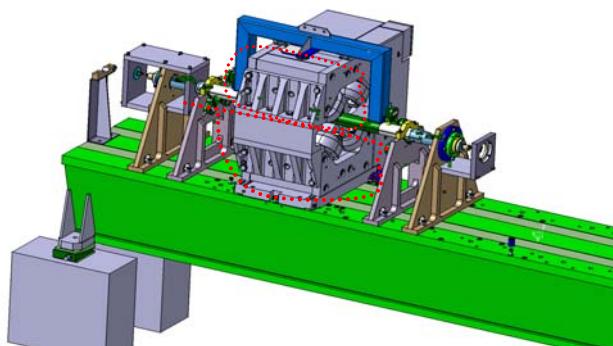


Fig. 32: Bench for magnetic measurement

6.9 Offset measurement of mechanical assembly

The relative position between the functional part and the external references (fiducials) can be assured mechanically by a drastic tolerance stack-up imposing very accurate machining and assembly (Fig. 33). It is sometimes better to release the constraints (and the cost) of mechanical machining and to measure precisely according to the chain of measurement indicated on the figure.

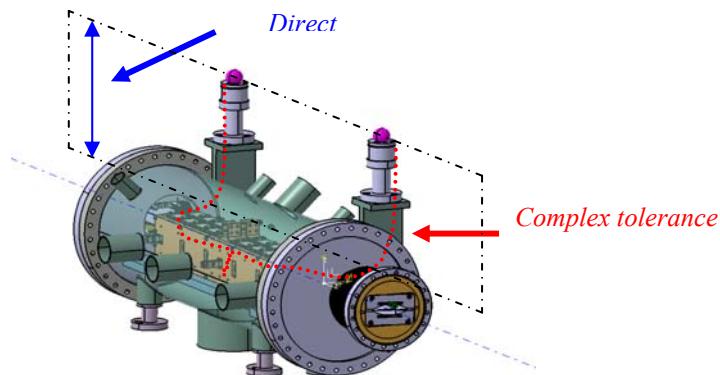


Fig. 33: Tolerance stack-up and direct measurement

6.10 Geodetic network

A network of measurements (i.e., a set of measures) to define a mechanical unit or to align them is a trivial example of a metrology loop. Since the redundancy is easier to control in the field of measurements than in mechanics, we shall use it as much as possible. The geodetic network includes all possible measurements (angles, distances, alignment, etc.) with a tacheometer between the different objects (Fig. 34).

The metrology loop is in fact more complex: the centring systems of the instruments and tools, the air where the optical beam travels and sometimes the inner sensors of instrumentation must be included.

Finally, the network may correspond to the measures between magnets installed on different girders to be aligned with respect to each other. The metrology loop includes the measured points. The

complete STC analysis for the slow drifts of the whole accelerator must incorporate the girders, the pedestals of girders, and also the concrete slab on which it is installed.

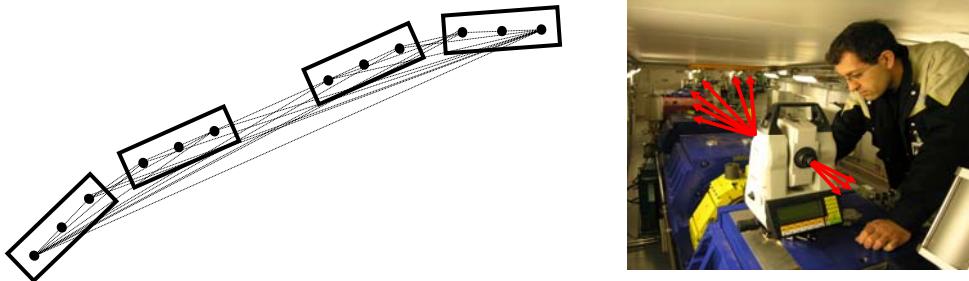


Fig. 34: Geodetic network and tacheometer TDA5005

7 Reversal layout

The error separation principle is used in any field of measurement whatever the trade, mechanics, physics, magnetism, alignment, etc. Thus that family of methods includes many different configurations but they all use the same frame called Error Separation Layout (ESL). Since this concept is of prime importance in DM, we introduce it progressively step by step.

7.1 The simple reversal

Any rotation of 180° makes it possible to eliminate offsets of centring systems. It is a powerful principle of bias reduction. The whole kinematical assembly that is reversed will be free of centring error. It is important to note that only the systematic part of the centring errors disappears and not the random errors, the number of which remains as many as there are reversed subsets. If the metrology loop is a series type, the errors are composed quadratically.

7.1.1 *Measure of the location of a bore by means of the reversal method*

A target is used to determine the planimetric location of a theodolite triback with respect to an external reference (Fig. 35), typically for its positioning on the alignment defined by two targets of reference. Since the bore called ‘A’ is not directly usable for an optical measurement, a spherical target is laid on a spacer. The theodolite aims at the target for an angle measurement.

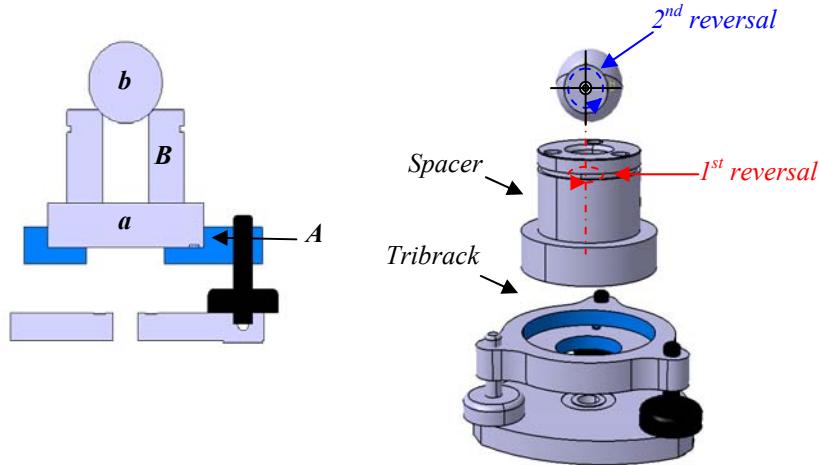


Fig. 35: Reversal method for centring systems

This spacer is machined at its ends as follows:

- At the bottom: a shaft called ‘*a*’ with the same diameter as that of the bore ‘*A*’ (with the functional clearance).
- At the top: a cone ‘*B*’ accepting the sphere with the target ‘*b*’.

Two ‘bias’ errors can exist, related to the mechanical machining:

- The offset between the shaft axis ‘*a*’ and the cone ‘*B*’ of the spacer.
- The offset between target and sphere centres.

A set of angle measurements is carried out:

- R_1 in an unspecified configuration.
- R_2 after the reversal of the spacer by 180° .

The average of this set of measurements eliminates the first bias: $R = \frac{R_1 + R_2}{2}$. In terms of referencing, the cone *B* is linked to bore *A* of the tribrack.

By applying the same principle of reversal to the sphere with respect to its centre, the second bias is eliminated. The target is then linked to the tribrack bore; the position of the target thus becomes representative of that of the bore. The error combination of centring due to the functional clearances gives: $\sigma_{tot}^2 = \sigma_1^2 + \sigma_2^2$.

7.1.2 Orientation measurement of a plane surface using autocollimation

The orientation of a plane materialized by a machined steel surface is obtained by autocollimation on an intermediate mirror (Fig. 36). This mirror is here a tool; it belongs to the system of measurement.

The first measurement is taken in an unspecified position of the mirror provided that it is in good contact with the steel surface. A second measure is carried out after reversal by 180° around the normal direction to the plane. The average of measurements gives the good orientation of the plane:

$$R = \frac{R_1 + R_2}{2}$$

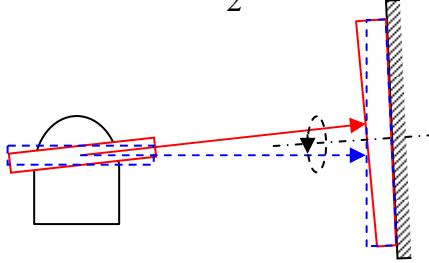


Fig. 36: Reversal method in autocollimation

The plane is now linked to the zero of the angular sensor of the instrument: the mirror disappears in terms of bias. If the instrument makes it possible to measure in the two angular directions (H and V), the principle of the reversal of the mirror is still valid.

7.1.3 Inclination measurement of a plane using inclinometry

One wants to measure the angle of inclination in a given direction of a plane materialized by a machined surface. An inclinometer is laid on the plan surface. The inclinometer functions with gravity which is its reference of measurement. The instrument shows a constant error (offset): the mechanical stands of the instrument define a horizontal plane for a reading different from zero.

A first reading is done, then another one after having reversed the instrument. The value of slope will be given by

$$R = \frac{R_1 - R_2}{2} \quad (9)$$

That formula corresponds to the average formula with a negative sign applied to the reading R_2 since the instrument and its zero are reversed and not the object to be measured. The instrumental error (offset) is given by application of the average of measurements:

$$\text{Offset} = \frac{R_1 + R_2}{2} \quad (10)$$

7.2 Multi-reversal measurement

The reversal principle can be generalized at various degrees. An important stage of this generalization is known as the multi-reversal method [4], [5].

It was developed for the dimensional measurement of parts having symmetry of order N . A mechanical part is entirely measured N times by a coordinate CMM in the N positions of the part after each rotation of $360^\circ/N$ around its axis of symmetry (Figs. 37, 38). At each step of the part rotation, a full rotation of the CMM sensor is carried out for rounding error measurements.

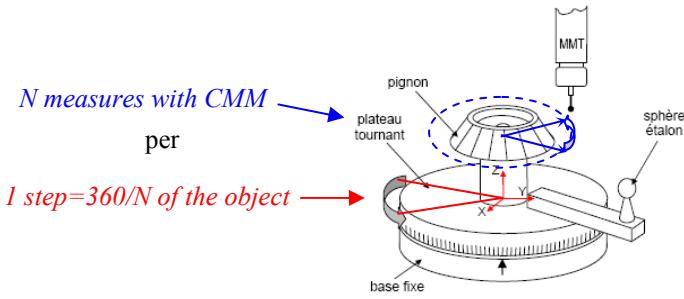


Fig. 37: Schematic of multi-reversal method [6]

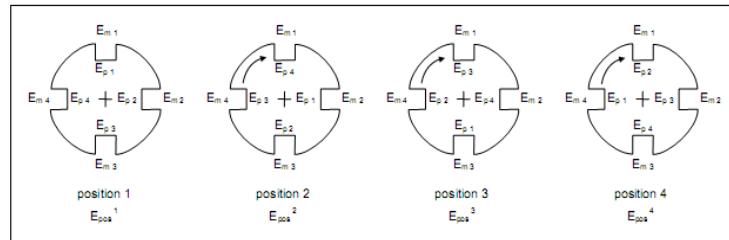


Fig. 38: Multi-reversal with order 4 symmetry [6]

Thus, each point of the part ‘sees’ successively the defects of the system of measurement and each position of measurement of the CMM ‘sees’ successively the defects of the part. The total error in each point is the sum of the error of the part (E_A) and of the measurement system (E_M):

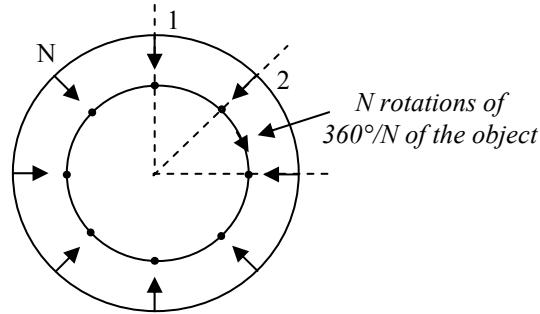
$$E_t = E_A + E_M .$$

Repositioning the part in the reference frame of the CMM after a rotation induces an additional unknown common to the corresponding set of measurements (E_P).

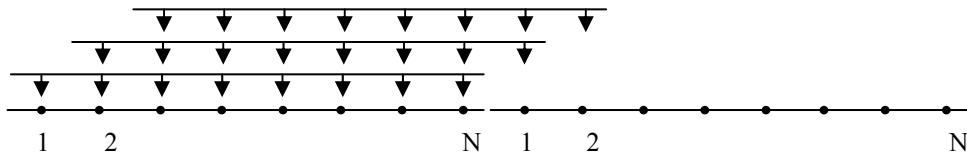
$$E_t = E_A + E_M + E_P .$$

7.3 More layouts

The CMM can be functionally described as a whole set of N sensors interdependent with the same structure and measuring the parts N times in parallel in N points of measurement (Fig. 39). Each sensor shows its own error equivalent to the ones of the CMM; these errors vary with the CMM geometry of its inner mechanics, the defects due to guidance units of the measuring head. The CMM errors calculated at the N points of the part are a combination of them. Conversely, the study of all the error vectors after calculation can lead to a determination of guidance unit defects [6].

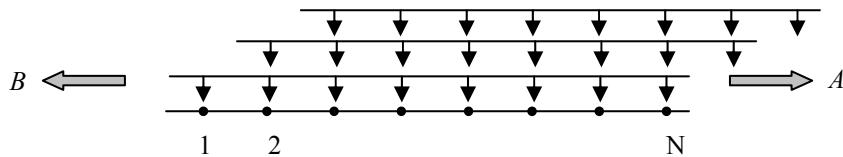
**Fig. 39:** ESL in circular configuration

The rotation of $360^\circ/N$ is relative between part and system of measurement: any part can turn or can remain fixed, according to hardware configurations. This circular architecture can be analysed linearly if its circumference is considered (Fig. 40). The functional diagram is as follows.

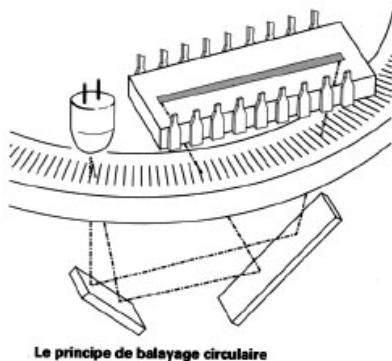
**Fig. 40:** ESL in circular configuration (circumference)

It reveals that the method can be applied to linear structures as well, without any particular symmetry, neither for the part, nor for the protocol of measurements. What is important is the ability of such an arrangement of both part and instrumentation, which allows the separation of several errors. In the rest of this lecture, we use the term ‘Error Separation Layout’ (ESL).

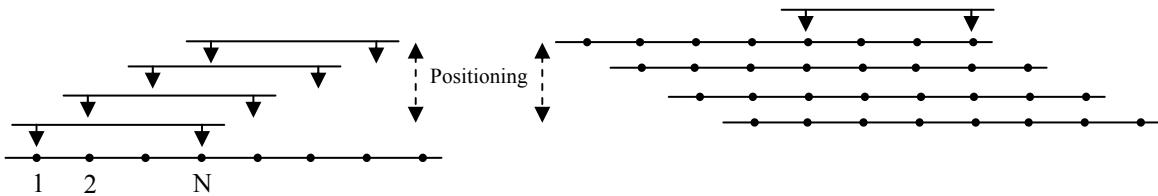
Figure 41 shows the corresponding diagram for a linear structure of N parallel measurements: since the symmetry does not exist any more, it is better to move the measurement structure in both directions.

**Fig. 41:** ESL in linear configuration

The sensors can deliver physical quantities other than length. As an example, it could be the output of an angular encoder at different points of the graduated circle of a theodolite (Fig. 42). The information is no longer a radial circle variation but a tangential variation of the spacing between graduations. This variation, reported to the radius of the graduated circle is interpreted in terms of angular quantity.

**Fig. 42:** Angular encoder of theodolite (^{1M}Leica)

Imagine that the gear wheel is now placed on a simple rotation stage equipped with an angular encoder. The roundness error measurement can be carried out by means of displacement sensors. N sensors for N measured points is expensive (but exhaustive). Using few of them or even only one is possible (Fig. 43).

**Fig. 43:** ESL with parse layout

Note that in Fig. 43, both layouts are equivalent regarding the positioning part: the number of unknowns is the same.

8 Modelling of an Error Separation Layout (ESL)

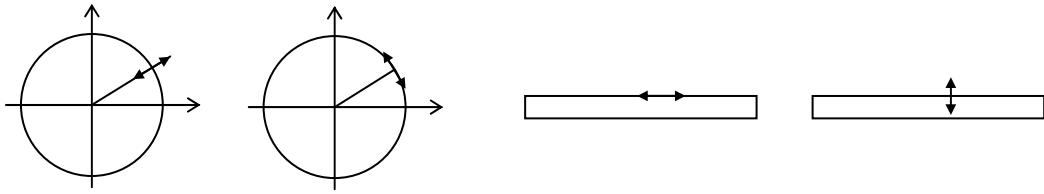
Any ESL can be summarized with the following variables:

(A, M, P, I) .

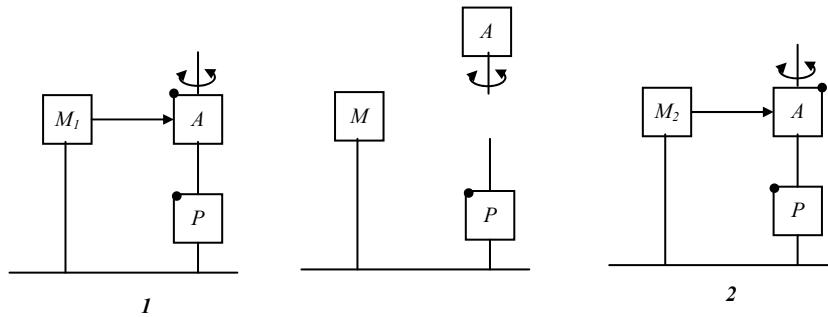
- the part called artefact : A
- the part corresponding to the measurement system : M
- the part or the operation for positioning : P
- the part corresponding to the instrument set-up : I

Different layouts can exist. The type of the information can be radial or tangential on circular or linear supports (Fig. 44). The dimensional information can come from the probes (typically displacement sensors) or from the artefact as with the angular encoders or graduated rules.

There are typically two families of unknowns between A , M and P parts. In addition, the instrument set-up unknowns are either supposed to be negligible or not.

**Fig. 44:** Type of information

However, a particular case is of prime importance due to the wide range of applications and to its theoretical interest: ESL for circular measurements. The relationship between parts is shown in Fig. 45.

**Fig. 45:** A first attempt at ESL in circular measurements

The measuring probe is supposed to be free of any systematic errors. Then, only artefact and positioning parts errors have to be assessed. We also assume that all these errors (or variations) are repeatable whatever the number of turns the system can do. The errors are only a function of θ .

The artefact is usually fixed on a spindle (or any rotation stage) whose rotation is not perfect regarding the level of accuracy to be reached on the artefact. That spindle corresponds to the positioning part of the ESL.

A unique probe for measuring the roundness errors of the artefact is set on the table. N points are measured with configuration no.1 of Fig. 45. Then, the artefact is dismounted and reversed on the spindle and the N points are measured again with configuration no. 2. Unfortunately, the reversal of the artefact on the spindle is not a true reversal with respect to the measuring probe: measurements will give partial information, in contrast to the positioning part where reversing it leads to (quasi-)complete information (see Section 8.4).

$$M_1(\theta) = A(\theta) + P(\theta) .$$

$$M_2(\theta) = A(\theta - \pi) + P(\theta) \neq -A(\theta) + P(\theta) .$$

Another ‘object’ appears, superimposed on the spindle parasitic movements and the shape of the artefact: the centring of the artefact on the spindle. If an eccentricity exists, it is not possible to fully reverse it. It creates a $\sin \theta$ curve superimposed on the measurements in the probe output. It is included in the instrument set-up ‘I’.

8.1 The positive ESL model

The complete model is shown on Fig. 46. It includes the following families of objects:

- Artefact: only a variable per θ , typically the radius $r(\theta)$ for radial measurements, or the angle $\alpha(\theta)$ for tangential measurements.
- Positioning: two variables per θ : $X(\theta), Y(\theta)$.
- Instrument set-up: (dX, dY) for the eccentricity, its effect is as $\sin \theta$, corresponding to the harmonic $n = 1$ in Fourier analysis.
- Measuring probes: $e(o)$, o being the probe output.

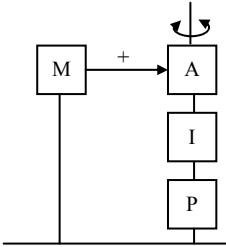


Fig. 46: The ESL^+ model

The direction of the arrow symbolizing the measurement will make sense later on (see Section 8.10). We propose to define the sign ‘+’ as ‘direct measurement’ for such an arrangement. The use of angular encoders will be with the negative sense.

The nature of artefact and positioning are clearly different.

Note that these objects can be extended to linear layouts with the corresponding variables. For example, the artefact can be described by $r(l)$ for straightness errors, $L(l)$ for graduation errors (Fig. 44).

There are usually only two families of unknowns: A and P in the example of the roundness and spindle errors measurements, A and M in the case of the gear wheel on the CMM.

8.2 The symmetry of true reversal

Evans et al. [7] showed that the reversal is mathematically similar to symmetry, usually obtained from a physical π radians rotation of one part of the layout to be assessed with respect to the measurement system, for example, the artefact or the positioning part. The study of the invariant allows the detection of reversal layout:

- i) **Artefact:** If the probe does not move, the invariant of a circular artefact is, say, the mean circle. Changing the direction of the error curve is clearly impossible (Fig. 47, index 1). The π radians rotation of the artefact is not a reversal because the required symmetry is not related to the centre of rotation of the artefact but to its mean circle. Another layout allowing a reversal condition is to rotate artefact and probe, then change the direction of sensitivity of the probe (Fig. 47, index 2). The invariant of the symmetry is here a point defined by the probe itself. Finally, a rotation of the artefact around a horizontal axis with the change of the sensitive direction is also possible (with the change of direction of θ rotation), as in straightness measurements (Fig. 48) of a slideway with respect to a datum straight edge.
- ii) **Positioning part:** the invariant is a point in 2D, the π radians rotation of the complete spindle (rotor + stator) is a reversal layout (Fig. 47).

- iii) **Eccentricity:** as for the trajectory of the positioning part, the invariant for symmetry is a point in 2D. In any case, repeating it after dismounting the artefact at micrometre accuracy or obtaining its reversal configuration with a high level of precision are major issues. As a consequence, the eccentricity cannot be separate from the *A* and *P* parts (Fig. 49).
- iv) **Measuring part:** In some cases, the parity of the linearity curve of probes can lead to symmetry.

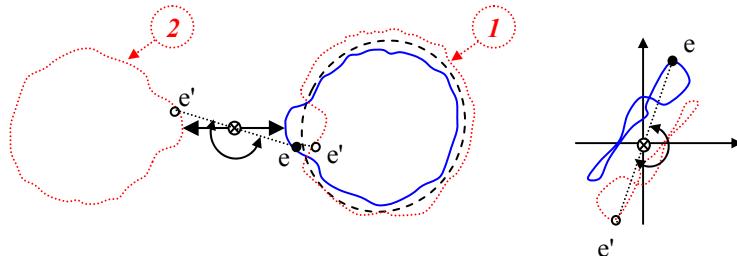


Fig. 47: Invariant of artefact and positioning part for reversal in circular measurements

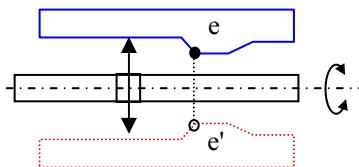


Fig. 48: Invariant of artefact for reversal in linear measurements

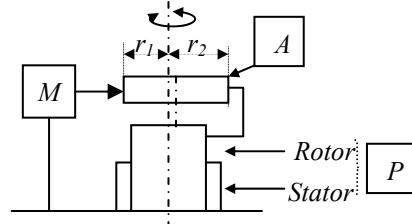


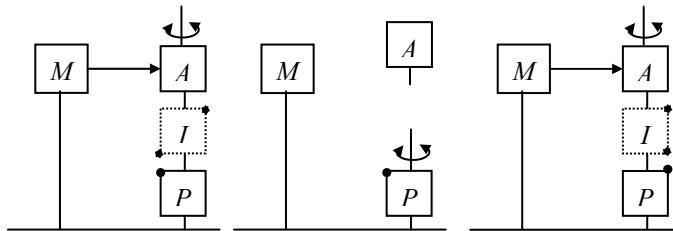
Fig. 49: Eccentricity of artefact/spindle

8.3 Extension to 2D and 3D problems: set of points remaining invariant

Considering the π radians rotation and its invariants is a first step in a more global approach applied to 2D and 3D problems. Self-calibration methods of stages dedicated to lithography in the semiconductor industry are based on the use of a lattice of invariant points defined by its symmetries [8].

8.4 The Donaldson reversal

When the positioning part (i.e., the full set rotor + stator of the spindle) can be reversed with respect to the probes and to the artefact, the situation is perfect: measuring all the N points versus θ and measuring again after the rotation of the spindle, leads to artefact and positioning errors in the sensitive direction of the probe (Fig. 50) without any loss of information except for harmonic $n = 1$ because of the lack of knowledge of the instrument set-up:

**Fig. 50:** Donaldson's reversal

$$M_1(\theta) = A(\theta) + P(\theta) \quad (11)$$

$$M_2(\theta) = A(\theta) - P(\theta) \quad (12)$$

$$A(\theta) = \frac{M1(\theta) + M2(\theta)}{2} \quad (13)$$

$$P(\theta) = \frac{M1(\theta) - M2(\theta)}{2} \quad (14)$$

A second probe, orthogonal to the first one must be set if 2D knowledge of the positioning part (parasitic movement of a spindle) is required.

Note that this method is strictly similar to the layout with an invariant point defined by the probe as shown in Fig. 47, index 2: the difference is the fact that both artefact and measuring are now reversed and not the positioning part.

However, these conditions are sometimes impossible to get, only the artefact can be rotated, or even neither part. In addition the remounting operation of the artefact may induce additional errors. Finally, the parasitic movement of the positioning part may be not perfectly repeatable: the curve of the trajectory is not exactly identical over several turns, creating a band instead of a unique curve. The mean shape of the band corresponds to the fixed or synchronous component of the error and its width corresponds to the variable or asynchronous component of the error. In these cases, using multiprobe or multistep methods is necessary.

8.5 Fourier analysis and generalized diameter

Let us introduce the Fourier analysis since the methods exposed below induce harmonic losses in the results of assessments. Several probes set around the artefact constitute the measuring part of the system, whatever the type of the delivered information. The probe output is the sum of artefact and positioning errors for the N measured points as in Donaldson's method. One can assume that the artefact shape or error curve can be described as a sum of harmonics, n being integer and N is the number of measurement points:

$$A(\theta) \approx \sum_{n=0}^{N-1} C_n e^{jn\theta} = \sum_{n=0}^{N-1} (A_n \cos(n\theta) + B_n \sin(n\theta)) = A_0 + \sum_{n=1}^{N-1} (A_n \cos(n\theta) + B_n \sin(n\theta)) \quad (15)$$

A_n , B_n , being the Fourier coefficients of the n^{th} harmonic.

Consequently, the artefact error curve is a combination of basic shapes whose values $n = 3$ and $n = 4$ are shown in Fig. 51. The harmonic $n = 0$ is a constant and corresponds to the mean radius of the artefact: $A_0 = R_0$. All harmonics are centred on zero, the sum of any, over θ is null:

$$\sum_{\theta=0}^{2\pi} (A_n \cos(n\theta) + B_n \sin(n\theta)) = 0, (\forall n \neq 0) \quad (16)$$

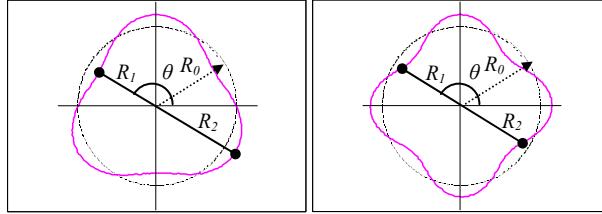


Fig. 51: Basic shapes for harmonics $n = 3$ and $n = 4$ and the corresponding generalized diameter for two probes: $D_{3/2}(\theta) = \text{Const}$, $D_{4/2}(\theta)$, varies as the shape

A generalized definition of the diameter of a circular figure is useful for what follows. The diameter of a circle can be written $D = 2R$, but it is also $D = R_1 + R_2 = \text{Const}$ with $R_1 = R_2, \forall \theta$.

Applying that definition to the first harmonic of the artefact decomposition (Fig. 51) leads to $D(\theta) = R_1(\theta) + R_2(\theta) = \text{Const}$ in the case of a circle. A simple chart will show that it is still true for any odd-order shape. Conversely, the even-order shapes present a diameter varying with θ as the shape.

Suppose now that there are three probes equally spaced at $k2\pi/3$ around the artefact. The relationship $D(\theta) = R_1(\theta) + R_2(\theta) + R_3(\theta) = \text{Const } \forall \theta$ is true for any n shapes with $n \neq 3k$ (Fig. 52) and $D \neq \text{Const}$ if $n = 3k$.

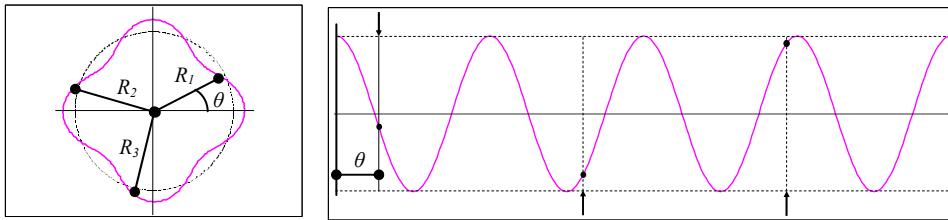


Fig. 52: $D_{4/3}(\theta) = \text{Const}, \forall \theta$

Let the generalized diameter (GD) of an n -symmetrical figure and for M equally spaced directions be:

$$D_{n/M}(\theta) = \sum_{m=1}^M R_m(\theta) . \quad (17)$$

The main interesting properties of the GD are

$$D_{n/M}(\theta) = \text{Const } \forall \theta \text{ for } n \neq kM \quad (18)$$

$$\frac{1}{M} D_{n/M}(\theta) \text{ varies as the shape itself for } n = kM \quad (19)$$

$$\frac{1}{2\pi} \sum_{\theta=0}^{2\pi} D_{n/M}(\theta) = R_0, \forall(n, M) \quad (20)$$

If necessary, and considering the artefact roundness being digitized by N points of measurements from the probes, the angle θ will be replaced by the number $i : 0 \rightarrow N-1$ where $\theta = i2\pi/N$.

$$\frac{1}{N} \sum_{i=0}^{N-1} D_{n/m}(i) = R_0, \forall(n, m) \quad (21)$$

Note that since the GD describes an error function, here of roundness, it can also be applied to an error function of tangential information. No particular condition concerning radial information has been expressed.

The concept of generalized diameter cannot replace Fourier analysis, but just offers a model for a qualitative approach. A thorough description of the analytic tools used in circular measurements has been synthesized by R. Probst in the appendix to his paper [9].

8.6 The multiprobe method (MP^+)

Constituting the average of evenly spaced probe outputs leads to the assessment of the variations of the GD of the artefact. But it cannot measure harmonics different from $n = kM, k \geq 0$.

Let us now try to manage the positioning part errors due to the spindle rotation. A displacement vector of the spindle rotor is seen by all the probes by the cosine for radial probes and by the sine for tangential probes of the angle $\varphi_m - \alpha(\theta)$ (Fig. 53). But the sum of all its contributions is always equal to zero at each step of θ because

$$\sum_{m=1}^M \cos(\varphi_m - \alpha(\theta)) = 0; \quad \sum_{m=1}^M \sin(\varphi_m - \alpha(\theta)) = 0, \quad (22)$$

with $\varphi_m = 2\pi m/M$

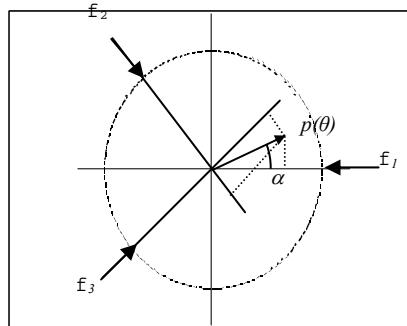


Fig. 53: Effect of a displacement vector $p(\theta)$ of the positioning part

In other words, the positioning errors are invisible on the GD of the artefact when averaging the probes, outputs, and the Fourier transform of the transfer function of the multiprobe method as in Fig. 54. Then, an assessment of A is

$$A(\theta) \approx \frac{1}{M} D_{n/M}(\theta) \approx \frac{1}{M} \sum_{m=1}^M M_m(\theta) , \quad (23)$$

with:

$$\frac{1}{N} \sum_{i=0}^{N-1} A(i) = A_0 = R_0 . \quad (24)$$

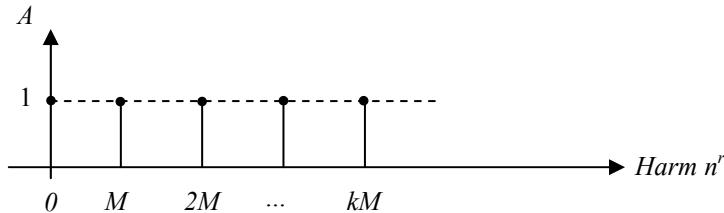


Fig. 54: Fourier transform (amplitude) of the MP⁺ method with n evenly spaced probes transfer function for the artefact roundness error

8.7 The multiprobe method with asymmetrical layout

The basic idea is to avoid the symmetry of the probes' angles since it is the origin of the harmonic losses. The probes are set at any φ_m angles (Fig. 55). Using the GD extended to that kind of layout, it is clear that the condition $D_{n/M}(\theta) \neq \text{Const}$ is easier to reach than with the equally spaced probes.

A clever choice of the φ_m can reduce drastically the harmonic losses in rejecting them far in the high frequencies (take care of which band is of interest).

However, the condition ‘forced to be zero’ on the positioning part is still necessary. Therefore a weighting GD is defined depending on the following conditions on the φ_m angles:

$$\sum_{m=1}^M a_m \cos(\varphi_m - \alpha(\theta)) = 0 ; \quad \sum_{m=1}^M a_m \sin(\varphi_m - \alpha(\theta)) = 0 .$$

The GD (its variations) is used by summing the probes' outputs M_m :

$$D_{n/M}(\theta) = \sum_{m=1}^M a_m M_m(\theta) .$$

Since the weighting coefficients distort the GD, it is necessary to correct its harmonic content. See more details about the method in D. Martin's lecture [10].

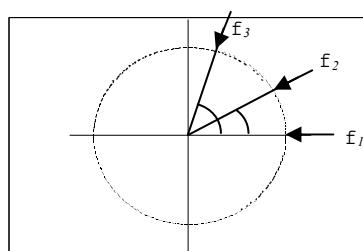


Fig. 55: The multiprobe method layout with asymmetrical angles

8.8 The multistep method (MS^+)

The multistep method involves only one probe and, therefore, only one sensitive direction as in the Donaldson method. Since a rotation of π rad of the artefact is not a true reversal, the idea is to rotate it M times with an increment of $2\pi/M$ between sequences of N measurement points (Fig. 56). Consequently, the artefact roundness error is shifted by $\varphi = 2\pi m/M$ at each sequence. The probe measures

$$M_m(\theta) = A(\theta - m\varphi) + P(\theta), \quad m = 1 \rightarrow M. \quad (25)$$

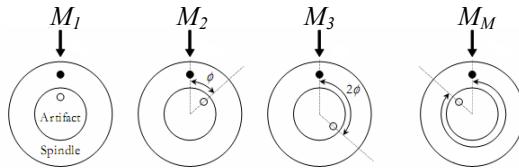


Fig. 56: Schematic of multistep method [11]

The example of the gear wheel measured by the CMM is actually a multistep layout with $M = N$.

The step $\varphi = 2\pi m/M$ being constant, the GD can be used: a radius at each step is measured. The GD (its variations) is calculated by adding the $M_m(\theta)$ together:

$$D_{n/M}(\theta) = \sum_{m=1}^M R_m(\theta) = \sum_{m=1}^M M_m(\theta).$$

As for the multiprobe method, $D_{n/M} = \text{Const}$ for all the harmonics such as $n \neq kM$. On the other hand, the error due to the positioning part is measured M times in the sensitive direction and in the same position (no rotation applied on the spindle). Then, constituting the GD as above and dividing it by M gives the positioning part for all harmonics except for $n = kM$ for which it is not possible to separate from the artefact roundness error:

$$P(\theta) = \frac{1}{M} \sum_{m=1}^M M_m(\theta). \quad (26)$$

The artefact error function is obtained by subtracting $P(\theta)$ from $M_1(\theta)$:

$$A(\theta) = M_1(\theta) - P(\theta). \quad (27)$$

The Fourier transform of the transfer function for the multistep method is shown in Fig. 57.

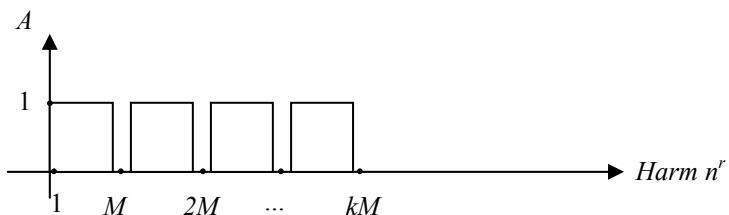


Fig. 57: Fourier transform (amplitude) of the MS^+ method transfer function of the artefact with $\varphi = 2\pi m/M$ increment. The harmonic $n = 1$ is chosen equal to zero.

The eccentricity of the artefact related to the spindle induces the same kind of uncertainty on the harmonic $n = 1$ as for the multiprobe method.

Note that the multiprobe and multistep methods are fully complementary [12], adding both results leads to a true reversal (except for $n = 1$).

8.9 Errors due to instrument set-up in roundness or spindle error measurements

The relative positions between measuring and positioning parts and artefact mislead about the parasitic amplitude on some harmonics. The case seen above is the eccentricity between artefact and spindle whose influence is on the harmonic $n = 1$. Note that it is possible to calculate it (if it is only the origin of the first harmonic) by using directly the Fourier coefficients $dx = A_1$, $dy = -B_1$.

The tilt of the artefact (parallelism error between the plane containing the N points of the artefact to be measured and the one containing the probes) can also influence the harmonic $n = 2$. But its influence is generally less sensitive than the effect of the eccentricity. The instrument set-up errors ($n = 1$, $n = 2$) cannot be separate from artefact or positioning parts.

8.10 Angular encoders: negative ESL and tangential measurements

Angular encoders are made of a glass circle whose circumference is accurately engraved up to 700 000 graduations (Fig. 42). The diameter of the circle may reach 140 mm.

The layout is the same as for the roundness error of an artefact. The only change is the type of information on the artefact which is tangential instead of radial. As for the radial measurement, the generalized diameter can be defined for the artefact error function. The positioning part of the system still exists with the two components $X(\theta)$, $Y(\theta)$, the read-heads correspond to the probes of the measuring part. They record the nominal value of an angle θ , the error due to graduation defects, the one due to the positioning part between probes and circle and instrument set-up influence (Fig. 58):

$$M(\theta) = \theta + A(\theta) + P(\theta) + I(\theta) . \quad (28)$$

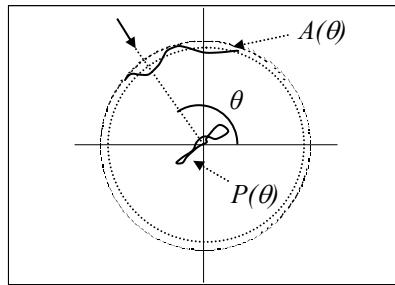


Fig. 58: Tangential measurements layout, $I(\theta)$ not shown

The following fundamental changes have to be noticed:

- The information goes from the artefact part (the circle) to the measuring one (the read-heads), contrary to the roundness error measurements (Fig. 59). This last remark is of the greatest importance because the transfer functions of the ESL methods do not have the same significance when using that layout for angular measurements: the transfer function of the multiprobe method with equally spaced probes appears to be very poor for roundness error measurement (Fig. 54). Since the roles are now inverted, that transfer function can be interesting for angular measurements to reduce the effect of the graduation errors on the angle assessment.

- The use of the system is then inverted: in the previous examples of roundness measurements, rotating the artefact over a turn was necessary for the assessment of the whole error function. When using angular encoders, the error function is not known and it is necessary to reduce its effect: an angle measurement is obtained by reading one time on the circle, this latter does not rotate over a turn. This is a negative ESL. The Fourier decomposition of the curve error still remains applicable even for unitary measurements (Fig. 60).
- The average of the error function of the graduations over θ is null: there is no concept of ‘radius’ anymore. Then, from Eq. (21):

$$\frac{1}{N} \sum_{i=0}^{N-1} A(i) = A_0 = 0 . \quad (29)$$

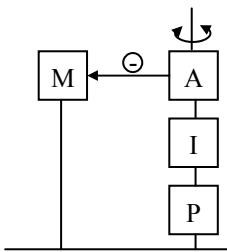


Fig. 59: The ESL model of angular encoder

A layout with two opposite probes (Fig. 60) using the average of the read-heads outputs ($MP2^-$) eliminates errors due to eccentricity between the circle (A) and probe support (M), all the odd harmonic errors of graduations, and finally the positioning part variations since it cannot be seen by the MP layout [Eq. 22)].

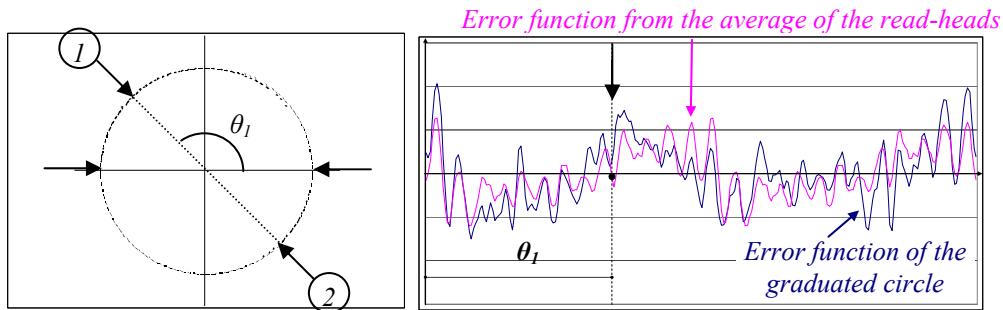


Fig. 60: θ_I angle obtained from a pair of opposite read-heads ($MP2^-$)

The four-probes configuration ($MP4^-, \varphi = k\pi/2$) is shown in Fig. 61. Both eccentricity ($n = 1$) and tilt ($n = 2$) due to instrument set-up are eliminated with the average of the outputs. Therefore, only $A(\theta)$ at $n = 4k$ harmonics affect the assessment of the angle θ .

We insist on the fact that the benefit from the MP transfer function on angular measurements is only due to the inversion of the information direction and not to its tangential nature.

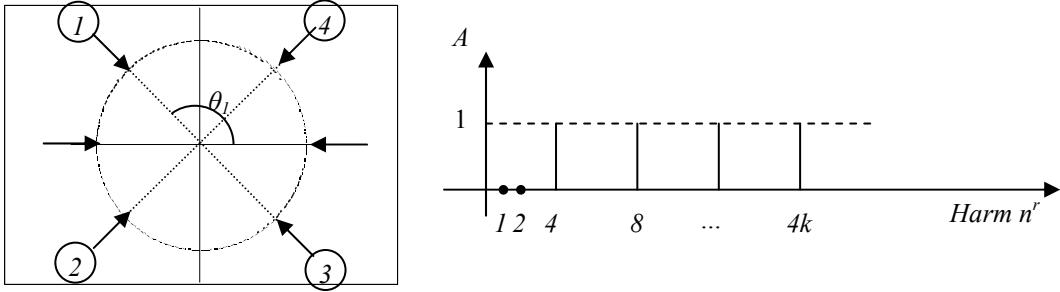


Fig. 61: Fourier transform (amplitude) of the transfer function for four read-heads of an angular encoder (MP4'): $n = 1$ (eccentricity) and $n = 2$ (tilt) are eliminated

8.11 Continuous measurement with angular encoders

A first approach for reducing the error function would consist in an increase in the number of probes. But we quickly face a limit due to cost and the physical layout. The second one consists in rotating the circle over a turn. The probes register the whole error function whose average is null [(Eq. 29)]:

$$\sum_{i=0}^{N-1} M(i) = N\theta + \frac{1}{N} \sum_{i=0}^{N-1} A(i) = N\theta \Rightarrow \theta = \frac{1}{N} \sum_{i=0}^{N-1} M(i). \quad (30)$$

In order to cancel the instrument set-up errors, the principle must be applied with the MP2⁻ or MP4⁻ layouts of angular encoders. The term $M(i)$ in Eq. (30) is replaced by the average of a MP layout.

An angle α as shown in Fig. 62 is computed with the difference of two MP2⁻ layouts: the circle rotates over θ in the MP2₁⁻ configuration, then the measuring part supporting the two read-heads is rotated by α and the circle rotates again over θ in the MP2₂⁻ configuration. The angle assessment is

$$\alpha = \frac{1}{N} \sum_{i=0}^{N-1} [MP2_2^-(i) - MP2_1^-(i)]. \quad (31)$$

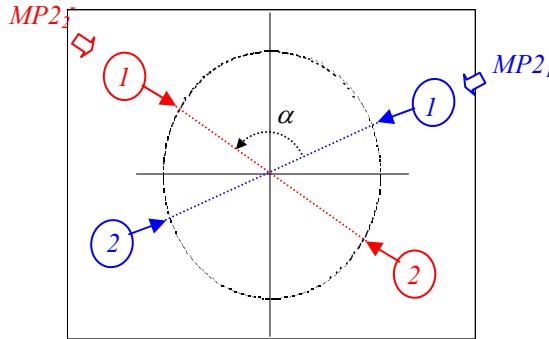


Fig. 62: An angle measurement without any error of graduated circle

That is why the dynamic sensors of accurate theodolites were introduced around 1980. The Wild T2000 has a circle graduated with only 1024 divisions [13]. The circle describes a full revolution for each angle measurement and is scanned by two read-heads, one fixed as a reference and the other mobile and mechanically linked to the telescope (Fig. 63). The phase shift between the two signals

makes it possible to calculate the angle formed by the two read-heads. It is the average value of all phase shifts which gives the precise measurement.

Actually, the T2000 is equipped with two pairs of opposite heads to eliminate the variable effect of the eccentricity (not shown in Fig. 63).

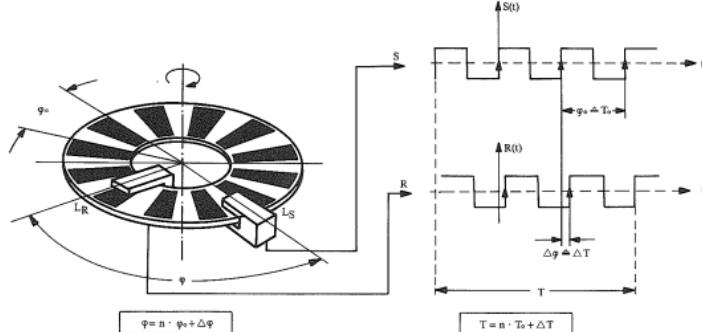


Fig. 63: Dynamic angular encoder of the Wild T2000

Similar assemblies exist for the realization of rotating plates of very high degree of accuracy (few 0.01") [4], [10], and [14]. The principle of the ESL is the same as for the T2000. Such rotating plates differ because they are equipped with a set of two encoders (circle + read-heads) mechanically mounted in juxtaposition to each other (Fig. 64). The circles are interdependent, turn together, and constitute part A of the ESL, each encoder (set of read-heads) being the MP⁺₁ and MP⁺₂ layouts of the probes.

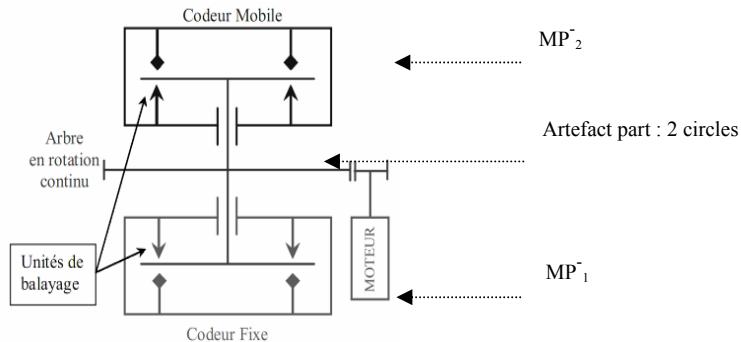


Fig. 64: Double encoder [4]

8.12 Calibration of graduated circles

Another use of the rotating circle of angular encoders is very similar to roundness error assessment of an artefact: it is sometimes necessary to assess the error function of the graduations in order to use it for checking manufacturing or for an *a posteriori* calibration use. In that way, the ESL problem (whatever the approach for calculation) is exactly the same as for the roundness measurements; the sense of measurement is direct (MP⁺, MS⁺): the wider range of the transfer function has to be found. In the so-called equal-division-averaged (EDA) method, two circles are compared and both layouts are used in parallel: MP⁺ for the reference circle and MP⁺ for the circle to be calibrated [15].

8.13 Multistep layout for angular measurements

Measuring an angle α with the multistep method is possible. The errors of the circle are reduced if the angle measurement is iterated with an increment of $360^\circ/M$ of the circle: for $M = 4$ (Fig. 65), the engraved circle is rotated by $\varphi = \pi/4$ before a set of measurements is carried out in the two directions defining the angle α .

One can prove that using the average of a pair of opposite read-heads and taking the M averages corresponding to the M steps, for calculating the GD as for the radial measurements [(Eq. 26)], gives exactly the same transfer function as for the multiprobe layout (Fig. 57). In this case, Eq. (27) is not used anymore.

But the multistep method assumes that the positioning part and the instrument set-up errors are synchronous. This assumption is not enough for very precise encoders. That is why the multiprobe layout is always preferred for high-precision angular encoders.

Before the advent of the current modern theodolites, geodetic networks campaigns could include up to 16 iterations for an instrument equipped with a pair of opposite read-heads [16].

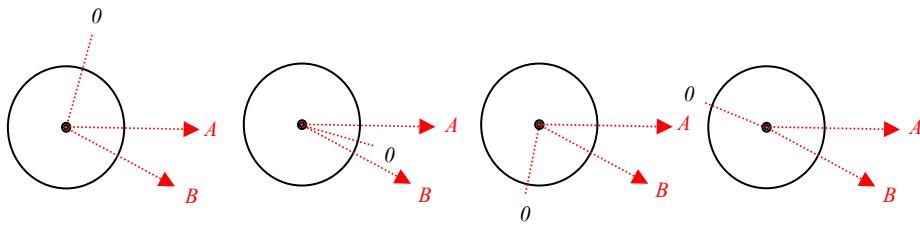


Fig. 65: Theodolite measurement in MS4⁺ layout

8.14 Matrix-based approach

Another way to solve the ESL problems is to define a set of linear equations describing the layout. The first step is to detect the unknowns involved in the layout. As an example, a MP3⁺ configuration with N measured points is described by N unknowns $A(\theta)$ for the error function of the artefact and $2N$ unknowns for the 2D curve of the positioning part: $X(\theta)$, $Y(\theta)$. On the other hand, the set of three probes gives $3N$ data. The set of equations does not allow redundancy in this example but a unique solution exists.

In the MP3⁺ method of radial measurements, the following equations describe the system shown in Fig. 66:

$$A(\theta) + P(\theta) = M(\theta) \Leftrightarrow N.X = M \quad (32)$$

where
$$A = \begin{bmatrix} Id & Id & 0 \\ Id & \cos(\varphi_2).(Id - \varphi_2) & \sin(\varphi_2).(Id - \varphi_2) \\ Id & \cos(\varphi_3).(Id - \varphi_3) & \sin(\varphi_3).(Id - \varphi_3) \end{bmatrix} \text{ with } \dim(A)=3N. \quad (33)$$

M is a vector built with the probes outputs. X is the vector of the unknowns. Id is the matrix *Identity*. $Id - \varphi_i$ are matrices built with the circular permutation of the columns of Id by a number corresponding to $\text{int}\left(\frac{\varphi_i}{2\pi}.N\right)$.

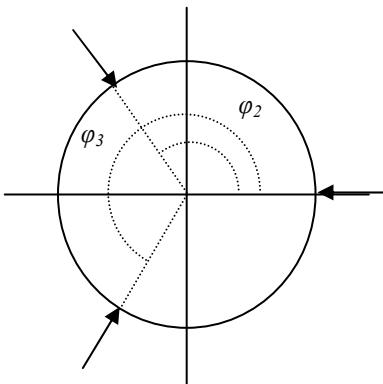


Fig. 66: MP3⁺ at $k2\pi/3$ layout

The general solution is obtained by solving the following equation according to the least-squares principle:

$$X = (A^T \cdot A)^{-1} \cdot A^T M . \quad (34)$$

In the case of a MP3 with three equally spaced probes, the result of matrix-based calculation is the **exact** solution (except for the harmonics involved in the instrument set-up) providing the data are also exact, i.e., without any noise. Conversely, the computation based on the GD always presents losses (see Section 8.5). However, the result may suffer from the same problem, depending on the level of noise on the measurement. Simulations show that the amplitude of losses is often lower in matrix-based computation than in the GD approach in the presence of white noise on the probe outputs but they also show that the accurate knowledge of the φ_i angles of the probe location is then more sensitive.

Since the system is rank-deficient, because that kind of matrix shows $\text{Det}(N) = 0$, inverting the matrix N leads to instabilities. Another way, which is far better, is to use the pseudo-inverse matrix of N . The equation is then

$$X = \text{pinv}(N) \cdot M \quad (35)$$

Figure 67 shows an example of the comparison on a set of data in a MP3⁺ layout of simulated roundness-error measurements and calculated with TMMatlab and using the **pinv** function. Here are the corresponding layout parameters:

- MP3⁺ layout
- $\varphi_i = k2\pi/3$
- $A(\theta)$ is a broadband signal.
- $N = 256$ measurements
- Any $P(\theta)$ is used
- The level of noise applied to the probe outputs is less than 1% of the $A(\theta)$ amplitude
- No $d\varphi_i$ is applied here

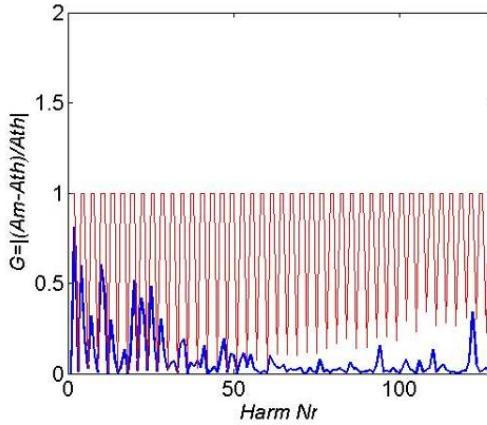


Fig. 67: Comparison of GD- and matrix-based calculations of the artefact for a symmetric MP3⁺ layout: noise on output probe effect

The error function gain is calculated for both solutions: $G = (A_{mes} - A_{theo})/A_{theo}$ where the A_i are the amplitudes of the harmonics. $G = 0$ means that the harmonic is fully determined, $G = 1$ means its amplitude is null. The theoretical Fourier transform of symmetric MP3⁺ appears clearly on the GD-based curve (red line) and the comparison is in favour of the matrix-based calculation (blue line). These results must be interpreted as the ultimate capability of the matrix-based approach because the effect of the uncertainty on the true location of the probes of about $d\varphi_i = 2\pi/N$ may be unacceptable (Fig. 68).

The proposed example does not lead to very good results when using it to assess the roundness; the asymmetric MP3⁺ is better. In that last case, GD- and matrix-based approaches tend to have the same behaviour in presence of a $d\varphi_i$ shift.

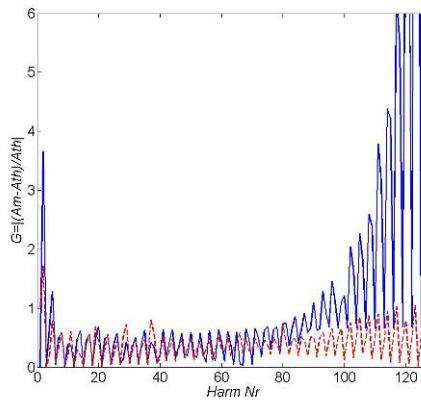


Fig. 68: Comparison of GD- and matrix-based calculations of the artefact for a symmetric MP3⁺ layout: noise on output probe + $d\varphi_i$ shift effect

The matrix-based approach in ESL can be delicate. The pseudo-inverse and the Singular Value Decomposition (SVD) may be interesting ways. Note that some authors have already proposed a novel approach with Prony decomposition involving SVD without harmonic losses [17].

In any case, applying matrix calculation for angular encoders is not necessary; the GD-based approach with engraved circle fulfils the challenge of precision, with or without rotation of the circle.

8.15 The ESL approach in magnetic measurements

Rotating coils are intensively used in the field of particle accelerators for the magnetic characterization of the magnets. The level of accuracy in terms of dimensional metrology can reach 0.01mm and 10 ppm in terms of magnetic parameters. That is why some care must be taken in the design and the procedure of magnetic measurements.

Since the bench dedicated to this use often includes a rotation part, coils of Hall probes, it is possible to find a correspondence between ESL and the techniques of the magician. The basic equation describing the magnetic field in a multipole magnet is as follows:

$$B(z) = \sum_{n=1}^{N(\infty)} C_n \left(\frac{z}{R_r} \right)^{n-1}, \quad (36)$$

where $z = x + jy = re^{j\theta}$ is the affix of any point located in the field and with $C_n = B_n + jA_n$.

Here B_n, A_n are the harmonic coefficients obtained by Fourier transform¹. R_r is an arbitrary reference radius acting as a normalization of the calculations.

One has to remember that the magnets we deal with are $2m$ -poles. The corresponding field harmonics are $m = 2n$ (Fig. 69).

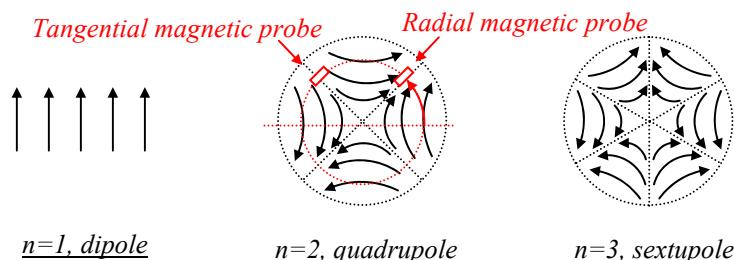


Fig. 69: The three first harmonics of $2m$ -pole magnets

The magnetic probes can run tangentially or radially. The set ‘probe-support’ is known as the detector. Both rotate as shown in Fig. 69 (red lines on Qpole chart), measuring the flux $\phi(r, \theta) = \phi(\theta)$, $0 < \theta < 2\pi$, since $r = \text{Const}$ allows the calculations of the harmonic coefficients of the field $B(z)$ through its Fourier transform.

Even if the typical Fourier transform of the multipole magnet fields is simple (only one main harmonic, the fundamental corresponding to the kind of the magnet), the undesired harmonics to be measured are less than 10^{-4} of the fundamental.

In terms of ESL, one can define the artefact part as the magnetic field or, rather, the flux measured by the probe. It corresponds to a signal being a function of the probe rotation θ . The measuring part corresponds to the probe (one or more) and the positioning part to the synchronous parasitic displacement of the probe when rotating. In that layout, the P part is applied to the measuring part instead of the artefact. Figure 70 shows the proposed model compared to the one of roundness measurements.

*In magnetism, the coefficients A_i and B_i are inverted with respect to the introduction to Fourier analysis (Section 8.5).

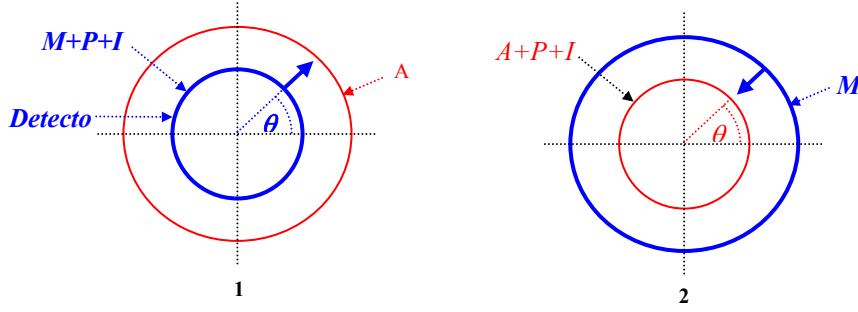


Fig. 70: ESL comparison: 1: rotating magnetic probe; 2: roundness measurements

That arrangement with a detector of only one probe can measure all the harmonics of the field (MPI^+ on Fig. 71). It is a simple version which is actually very common. There are N unknowns of the artefact part defined by an increment of $\theta = 2\pi / N$. Another ESL is present on the bench with the use of a two-read-heads angular encoder and linked to the detector shaft (MP2 on Fig. 71) to control the rotation θ .

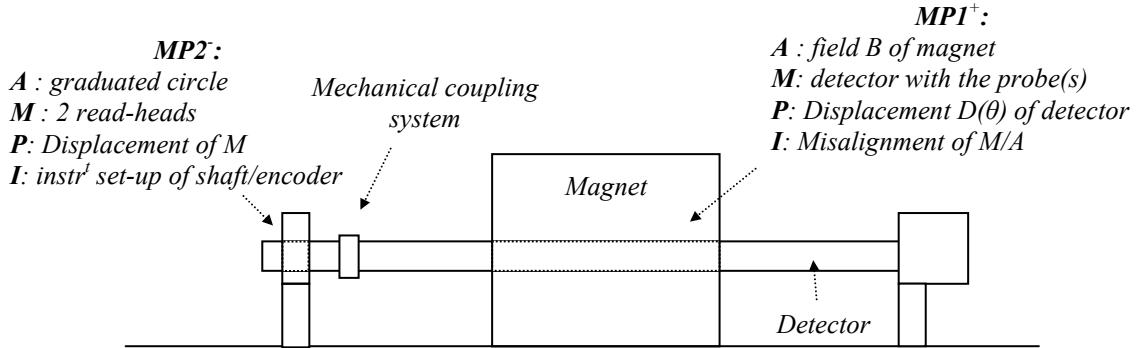


Fig. 71: The ESL on a bench for magnetic measurements

Instrument set-up errors also exist in such magnetic benches: the flux measurements are affected by the misalignment of the detector with respect to the centre of the magnet field. If the offset between the centre z_c of the magnet, typically a quadrupole, and the centre z_m of the detector rotation is $d = (z_c - z_m)$, then every Fourier coefficient $C_{n_{\text{meas}}}$ measured by the detector is a function of

$$\sum_{k=n}^{\infty} C_k \frac{(k-1)!}{(k-n)!(n-1)!} \left(\frac{d}{R_{\text{ref}}} \right)^{k-n} \quad (\text{Ref. [18]}).$$

That function highlights the nature of the artefact which is different from dimensional radial or tangential measurements: an eccentricity modifies only the first harmonic in the dimensional domain but all of them in magnetism. It is due to the artefact in magnetism (field \vec{B}) whose harmonic spectrum varies as r^{n-1} instead of r^1 in the dimensional domain. However, applying the formula to $n = 2$ corresponds to the dimensional metrology case, and also to the quadrupole field. The formula gives

$$dx \approx \frac{R_r}{2} \frac{B_1}{B_2}, \quad dy \approx -\frac{R_r}{2} \frac{A_1}{B_2}. \quad (37)$$

Therefore, the first harmonic is supposed to be due to the misalignment offset only, and the measured coefficients B_1 and A_1 are then used to give the co-ordinates of $(z_c - z_m)$ as in dimensional metrology (see Section 8.9).

Using the first harmonic to assess the misalignment as for the roundness error measurements assumes that there is no artefact information in it, i.e., no dipole field created by the quadrupole. That hypothesis is usually reliable.

The angular encoder may show a similar offset: it can be not coaxial with the shaft axis of the detector. The effect is of the same kind as between detector and quadrupole: the encoder delivers an output affected by an error $e_\theta \propto \sin \theta$, creating a harmonic $n = 1$ (dipole) in the measured spectrum of the field. The parallelism defect between encoder and shaft axes induces an error on the harmonic $n = 2$ similar to the one in ESL due to the tilt effect (see Section 8.9).

Even if the magnetic benches have excellent quality of manufacturing, especially for the shaft of the detector, parasitic displacements may remain during its rotation. The problem is well known in dimensional metrology; that is the positioning part of the ESL. It has been proved that the main part of the effect appears on the harmonic ($n = 1$) of the measured spectrum of a $2m$ -pole magnet [18]: such displacements can be interpreted as an instantaneous misalignment of the detector, depending on θ . The eccentricity E is fixed over θ , the displacement $D(\theta)$ is a function of θ .

A possible improvement could be the design of a three radial coils detector with a dissymmetrical layout (MP3⁺ dissymmetric, see Section 8.7) for the quadrupoles' magnetic measurements since they are quasi similar to the dimensional case and since their misalignment is sensitive. Two more coils could control the parasitic displacement of the M part during its rotation if required. The level of rejection to higher harmonics that shows the dissymmetrical layout would be easily favourable as regards the harmonic distribution of quadrupolar field. That proposal has to be tested to be confirmed, especially because the probes in magnetism measure the flux and not the field and therefore its harmonic spectrum varies as r^n . The probes do not run like in dimensional measurements, even for a quadrupolar magnet:

$$\Phi(\theta) \propto \int_r B(r, \theta) dr = f(r^n) \quad [18].$$

The approach used so far for the magnetic measurements ESL analysis can be completed by the following examples of what could be a parallel to ESL in that domain. It is sometimes necessary to measure only an harmonic, say $n = 5$, corresponding to a decapole component of the field. A solution is to use a detector with five probes (coils or Hall probes) equispaced with $\varphi_i = k2\pi/5$ (Fig. 72) and the harmonics $n = 5k$ are measured. Such a detector has been designed in the LHC project for static measurements (no rotation of the detector) [19] but it is sometimes used as a MP5⁺ arrangement with rotation. The same is true with the use of two identical coils in electric opposition [20]: the odd harmonics are rejected, especially the one of the dipole ($n = 1$). In these two examples, the positioning part of ESL is supposed to be null; it is not a true error separation situation.

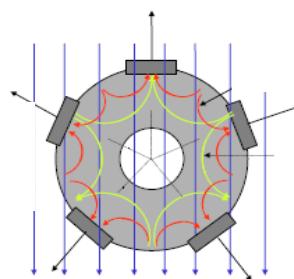


Fig. 72: Five equispaced Hall probes \Leftrightarrow MP5⁺ [17]

8.16 Conclusions

The GD-based approach cannot replace the Fourier analysis tools which are far more complete. But what we want to do is to enlarge the model qualitatively. The ESL model could be a synthetic point of view common to many applications in very different domains such as surveying, mechanical metrology or magnetism. Keeping in mind the ESL model would help any designer or metrologist in any domain of the measure. Further investigations would be of great interest for applying ESL in the domain of rotating coils.

9 Differential measurements

The principle of differential measurement is extremely common in all fields of metrology. Instead of measuring a distance between two points with a ruler, by applying an extremity to a point and by reading the graduation corresponding to the other one, the ruler is shifted and two readings on the ruler are carried out in front of the two points (Fig. 73). The length is the difference between the readings.

The first method, known as ‘direct measurement’, shows a regular problem due to the accessibility of the ruler’s origin. The use of a measuring instrument often needs a material definition, a mechanical reference which is supposed to be representative of the origin of measurement itself. Unfortunately, a good precision of the zero adjustment with respect to a mechanical reference is seldom possible. Calculating the difference between two measurements eliminates it:

$$d = (L_2 + e) - (L_1 + e) = L_2 - L_1 . \quad (38)$$

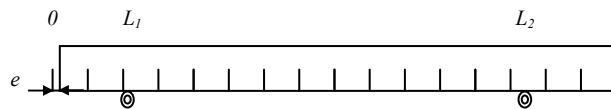


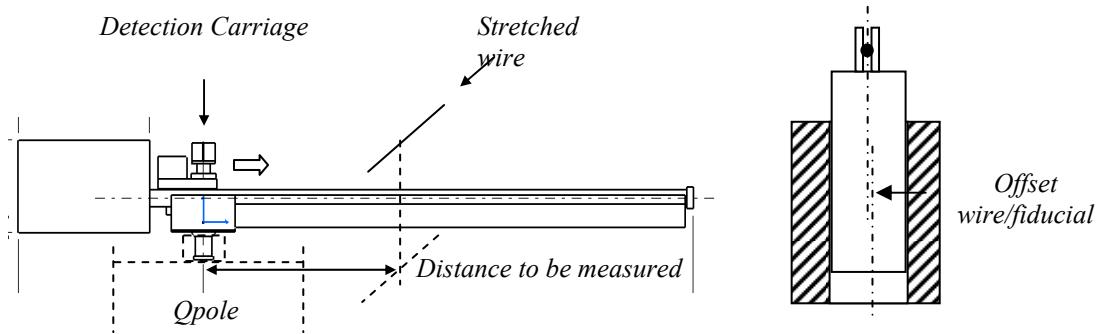
Fig. 73: Differential measurement

This example, rather commonplace in the field of length measurements, is evident in the angle measurements. Indeed, an angle measured with an angular encoder is inevitably a difference of two directions. However, that kind of differential measurement cannot correct the linearity errors of a measurement system, only a calibration can solve it.

The application of the principle can be considered in a 2D or 3D configuration, the example of the ruler or the angular sensor being in 1D.

9.1 Ecartometry measurements

A wire ecartometer is used to measure the misalignments of components by means of their fiducials. The instrument (Fig. 74), centred on the fiducials of the component to be aligned, measures the distance from this point to the line defined by a wire stretched between the fiducials at both ends (Fig. 76). The adjustment by the zero measurement in coincidence with the mechanical reference of the instrument is difficult to realize. Each reading L_i is affected of the zero offset e of the instrument: $L_i = D_i + e$. In addition, centring the wire precisely on the two fiducials of extremity is very expensive in terms of accuracy and offsets also exist on the wire position (Fig. 75). The elimination of these errors is carried out by measuring with respect to a wire stretched between two other points which are not the fiducials and by changing the base of calculation: thus one creates a fictitious wire and measures without offset.



Figs. 74, 75: Wire ecartometer and centring error

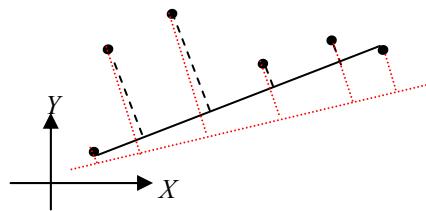


Fig. 76: Differential measurements in ecartometry

9.2 V_0 in geometry

The use of a tacheometer (measurements of both angles and distances) to determine the Cartesian coordinates of the points C and D while knowing those of the points A and B is simplified by adding a variable in calculations V_0 which corresponds to the direction of the zero of the angular encoder of the instrument (Fig. 77). The variable V_0 is calculated from the known co-ordinates of points A and B :

$$V_0 = V_{AB} - L_B .$$

Then, the directions of the vectors AC and AD are calculated with respect to the y -axis (direction) from the V_0 : $V_i = V_0 - L_i$.

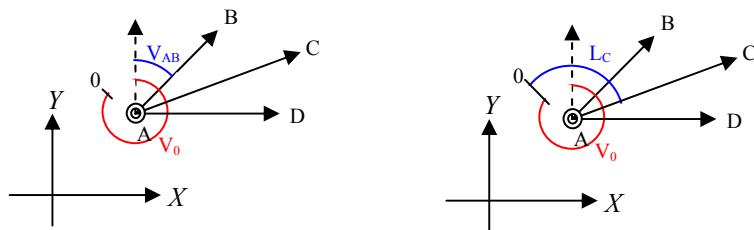
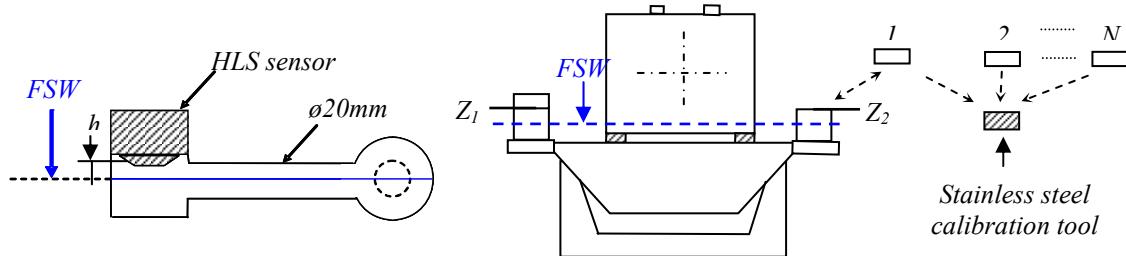


Fig. 77: V_0 in planimetry

The V_0 has the same role as the ecartometer offset e of the previous example. The zero of the circle is not known and justifies its use.

9.3 Network of a Hydrostatic Levelling System (HLS)

The HLS is a capacitive non-contact displacement sensor which runs with electric conductors and especially with free surface of water (FSW) defining a perfect horizontal plane (Fig. 78). It enables measurements and adjustments of relative position between structures in the vertical direction. When a hundredth of mm accuracy is required, the determination of the offset e of the sensor due to the zero drift with respect to its mechanical reference is critical. We present an example using as many sensors as the N points to be controlled (Fig. 79).



Figs. 78, 79: HLS sensor and HLS network and calibration tool

N sensor offsets are to be determined. The principle of differential measurements intervenes first, on the relative positions $dZ = Z_1 - Z_2$ of the zeros which are to be determined, not their true position h with respect to the surface of water, which only represents a precise reference of horizontality. The differences in the readings dL from the sensors give the position information dZ . But each sensor HLS_i is affected with an error e_i . Since the configuration is differential, then the principle also applies to offsets. The problem is solved by the use of a stainless-steel calibration tool (Fig. 79) which is measured by all the sensors to compare their offsets: the common part of the offsets disappears: $h_1 - L_1 + e_1; h_2 - L_2 + e_2; de = e_1 - e_2$ is measured with the tool. Then $dZ = h_1 - h_2 = L_1 - L_2 + de$.

9.4 Interferometry

The interferometry principle is a very powerful tool in DM whatever the scale of the measurements. It is employed in microscopy and astronomy. What is interesting here is that the differential principle exists at the physical level. The fringes of two signals, one from the measuring channel and the other from the reference, are combined together. The scale of the phenomenon is given by the wavelength of the emitted light; it means that the sensitivity can be very important; the level of a nanometre is reached in many cases.

The most common arrangement in DM is achieved with the Michelson interferometer (Fig. 80) where a half-transparent mirror (beam-splitter) splits a laser source into two beams, one acting as a ‘reference’ and the other one being the ‘measuring’ one. The two beams are recombined, creating a pattern of fringes and then processed after each of them has been reflected onto a mirror.

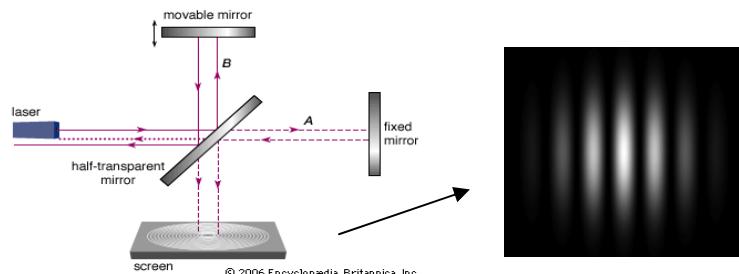


Fig. 80: Michelson interferometer applied to linear displacement measurement

If a mirror moves, typically the ‘measuring’ one, while the other mirror stays fixed, a fringe pattern scrolling arises, depending on the distance d of displacement. Counting the fringe leads to the distance d .

$S_1 = a_1 \cos(2\pi f t)$ is the signal coming from the reference, while $S_2 = a_2 \cos(2\pi f t + \varphi)$ is the one coming from the moving mirror. φ is the phase difference due to the displacement:

$$\varphi = k \frac{2\pi}{\lambda} 2d . \quad (39)$$

λ is the wavelength of the light source in the air. If $d = \lambda/2$ and $k = 1$, then φ describes 2π in phase. That is why $\lambda/2$ is considered as the basic increment of such methods. Counting k , the number of fringes leads to the distance $d = k \frac{\lambda}{2}$.

The mirrors can be replaced by retro-reflectors which have the advantage of reflecting accurately the beam parallel to the incident one. However, flat mirrors are always used for ultimate accuracy.

Several models of interferometer exist. The heterodyne interferometer using two frequencies is very common in metrology laboratories (Fig. 81). The laser source includes two frequencies f_a and f_b with in quadrature polarities. The signal $f_b - f_a - \Delta f$ is analysed, where Δf is the shift frequency due to the translation of the ‘measuring’ retro-reflector.

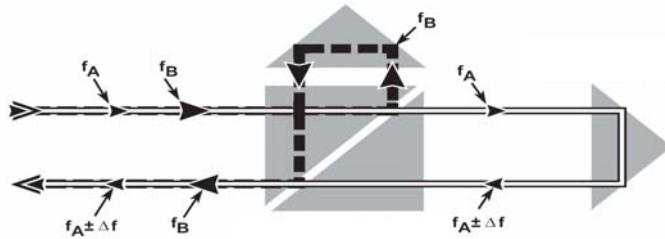


Fig. 81: Heterodyne interferometer: layout for linear measurements (TM Agilent)

Interferometry can be applied in other domains of DM such as angle measurements or straightness measurements. Figure 82 shows the schematic of differential linear measurement applied to angles: the so-called ‘reference’ reflector of the linear set-up is now linked to the measurement one. Any rotation of the angular reflector induces a path difference between f_a and f_b corresponding to a relative displacement d . If the spacing H of the two retro-reflectors is accurately known, the angle displacement can be computed: $\sin(\alpha) = d/H$.

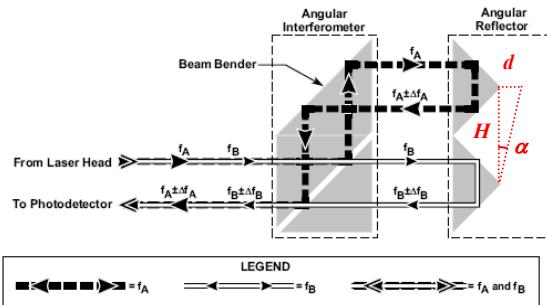


Fig. 82: Heterodyne interferometer for angular measurements (TM Agilent)

In order to reach ultimate accuracy (few nm), the differential principle is applied at many levels. Thus more precise optics need to mix both frequencies on the whole optical path so that they ‘see’ exactly the same errors (Fig. 83).

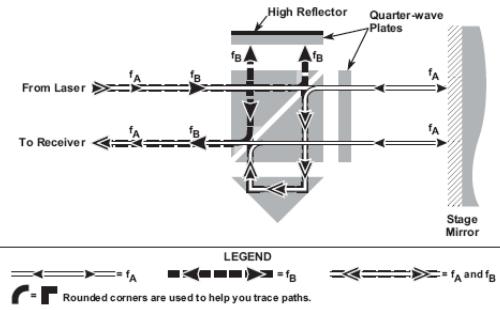


Fig. 83: High-accuracy optics for linear measurements (™Agilent)

9.5 Differential measurements with magnification: a link with reversal layout

Some layouts allow the measurement of the double of a physical quantity to be measured. These are typically the reversal layouts where the error is as $2e = R_1 - R_2$ (Section 7.1). Some physical arrangements can reach such effects. They also correspond to a reversal arrangement, the physical π radians rotation being always present. The use of a retro-reflector shifts a beam offset (Fig. 84). Using a plane mirror for autocollimation is exactly the same situation: the sensitivity of the method is multiplied by two (Fig. 36). That principle is also applied for reading at opposite read-heads of an angular encoder [21] or viewing the extremities of a levelling machine bubble [Fig. 85(b)].

In the field of electronics, quad cells for photon beams as well as BPMs for electron beams run identically [(Fig. 85(c))]. The beam position variations in z , for instance, are computed from the four electrode signals of the BPM [22]:

$$z \propto \frac{(I_A + I_B) - (I_C + I_D)}{(I_A + I_B + I_C + I_D)}. \quad (40)$$

The sum of the current in the denominator is constant for small displacements of the beam and acts as normalization with respect to the beam current. But the structure of a reversal is still present even in Eq. (40). The four electrodes are used in the horizontal direction as well (x), and two symmetry axes exist for the π radians rotation of the reversal layout.

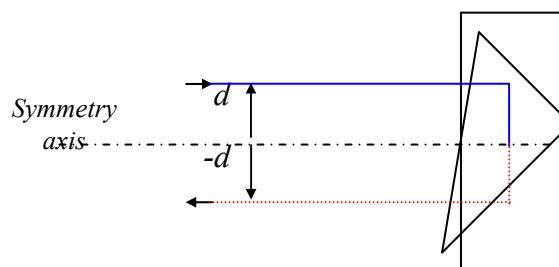


Fig. 84: Symmetry in a retro-reflector

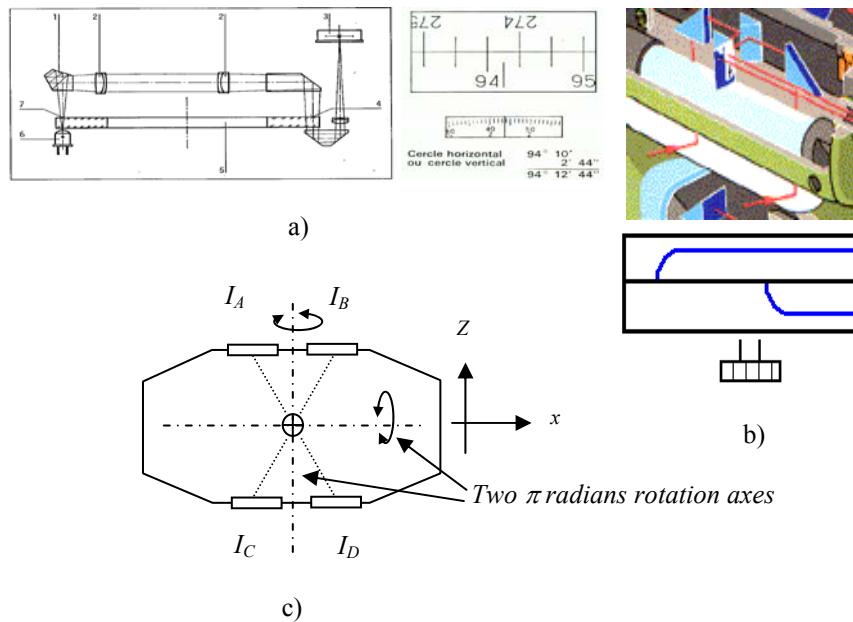


Fig. 85: Instrument components showing differential function: (a) opposite read-heads of an optical angular encoder (theodolite), (b) bubble of a levelling machine, (c) BPM

Furthermore, magnification can be greater than two. The classical example of the use of the reticle pattern with two sloped lines aiming at a target is interesting. The pattern has an angle α and the upper and lower lines intersect the target differently. Figure 86 (which is drawn to scale) shows a vertical displacement h of only 2% of the width graduation and induces a dx shift of 28% corresponding to a magnification $G \cong 13$. The corresponding formula is

$$dx = 2h \cdot \cot(\alpha/2) . \quad (41)$$

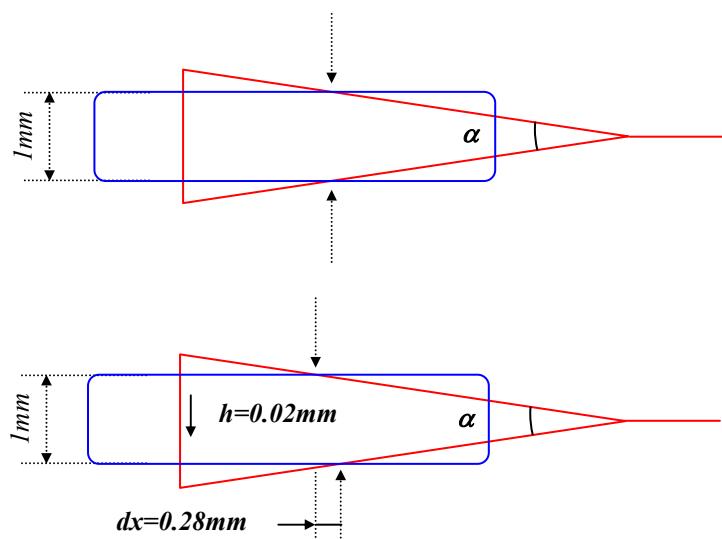


Fig. 86: Differential measurement with magnification (the proportions are correct))

10 Case study: the orbit definition of a synchrotron storage ring

10.1 Presentation

The beam orbit of a storage ring is fully defined by the location of its quadrupole magnets (Qpoles) (Fig. 87). The beam quality depends on the quality of the orbit. Before the alignment operation of these magnets, it is necessary to materialize their magnetic axis. In terms of metrology loop, it consists in linking the magnetic definition to the fiducialization. Then, magnets can be adjusted with each other by means of mechanical and/or optical methods. The magnetic axes of all the magnets are then linked together to define the orbit. The first step of the alignment is the magnetic axis detection of the Qpoles. A magnetic bench is dedicated to that operation (Figs. 87 and 93) from which two kinds of fiducials are adjusted (for mechanical alignment on girders) or measured (for optical adjustment in the tunnel). These latter are surveyed with a special structure called ‘Qpole comparator’ with respect to the cylindrical support of the rotating coil.

The Qpoles are then mounted on their girder and laser measurements are carried out to check their mechanical alignment.

The last step is the precise alignment of all Qpoles together to define the storage ring orbit with wire ecartometers for the planimetry and HLS for the altimetry.

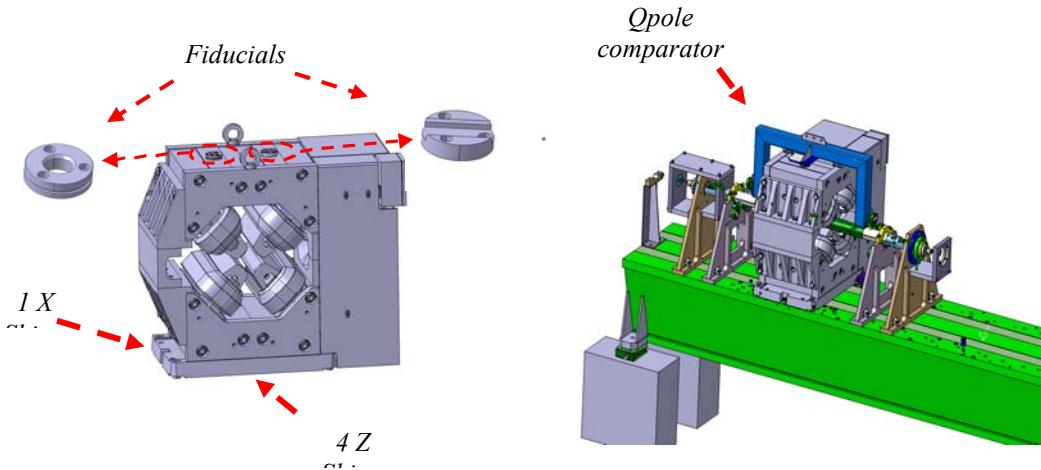


Fig. 87: Storage ring quadrupole at SOLEIL and bench for magnetic measurements

10.2 Detection with rotating coil

The description of the magnetic bench dedicated to the SOLEIL quadrupoles is given in detail in Arnaud Madur’s thesis [22].

Rotating coil measurements allow knowledge of the magnet magnetic field \mathbf{B} . The coil rotates and then measures the flux coming from the magnet. The coil is set radially or tangentially. In the present case study, a radial coil is used (Fig. 89). A Fourier transform is applied to the flux signal. The magnetic axis of the Qpole is then located (Fig. 88) by applying the following formulas:

$$dx \approx \frac{r_0}{2} \frac{B_1}{B_2}, \quad dz \approx -\frac{r_0}{2} \frac{A_1}{B_2} \quad \tan(2\theta) = -\frac{A_2}{B_2},$$

B_n , A_n , being the n^{th} real and imaginary harmonics of the magnetic field $\overrightarrow{\mathbf{B}}$. Note that these coefficients depend on the geometry of the coil. Here r_0 is a constant and θ the magnetic tilt.

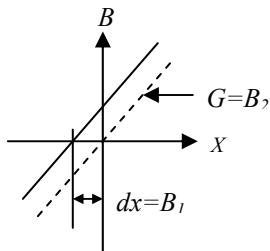
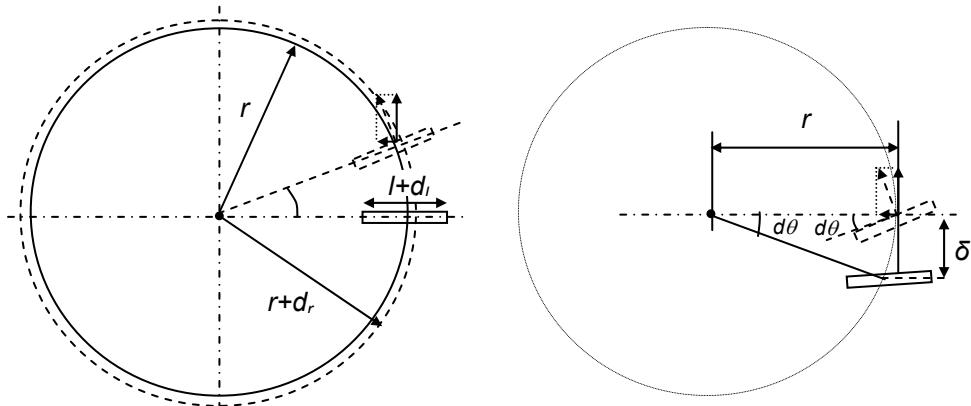


Fig. 88: Principle of zero detection applied to the x direction

Some imperfections due to the realization of the coil and of its support can occur: the geometry may not be exactly the one expected (Fig. 89). The lack of precision on the realization of the coil is not of first importance in the zero detection because of the symmetry of revolution of its influence. It should induce a scale factor to be applied to the residual dx and dz still existing after having set the definitive shims. Since these latter are small, that effect can be neglected on the alignment of the magnet.

The rotation axis of the coil may be outside the plane including the coil itself (Fig. 90). In that case, the flux seen by the coil is not strictly radial and a tilt error $d\theta = \delta/r$ appears on the orientation of the Qpole magnetic definition. Note that the case of an offset of the coil in the Z direction of about δ is strictly equivalent. That error can be estimated by rotating a magnet around its vertical axis and by using the average of the two corresponding tilt values.



Figs. 89 and 90: Radial coil and geometrical defects

In the proposed case study, we shall suppose that the true axis of rotation is fixed and confused with the geometrical axis of the coil. In other words, there is no radial run-out due to the bearings.

Figure 91 shows the bench in detail: components and metrology loop (red dashed line). The rotation of the sensor is managed with an angular encoder. Its zero is defined by an inclinometer fixed on the sensor. It will be possible to link the angular readings to gravity and thus, measure the magnetic tilt of the magnet.

The magnetic sensor is fragile and a change during the period of two or three months when measuring all the Qpoles must be envisaged. Any other part of the bench could slightly change in position too, especially the pin which defines the X reference (10). As a conclusion, it is necessary to link the magnetic axis of the coil with respect to the X pin and to the girder surface which is the Z reference. A reference tool (3) has been designed: a permanent Qpole magnet with eight mechanical faces to position it on a stand fixed to the bench and in contact with the X pin and the Z surface (Fig. 91).

The eight positions allow a good redundancy: the measurements could describe a perfect octagon and their average is used as a final result. In addition, if the size of the reference tool is accurately known, it would be possible to use another tool in case of destruction of the first one. An ESL configuration appears here. The magnetic axis of the tool is defined with respect to its eight faces and to the coil. The use of the reference tool must be analysed in terms of STC: it allows avoiding $STC = (\mu\text{m}, \infty)$ for the whole metrology loop which is very difficult to achieve (no change possible during the campaign). If both tool and bench mechanics are supposed to be perfectly machined at the nominal sizes, then the result should be zero. It would be possible only if every part was machined at a high level of accuracy (better than $10 \mu\text{m}$), that is very demanding work.

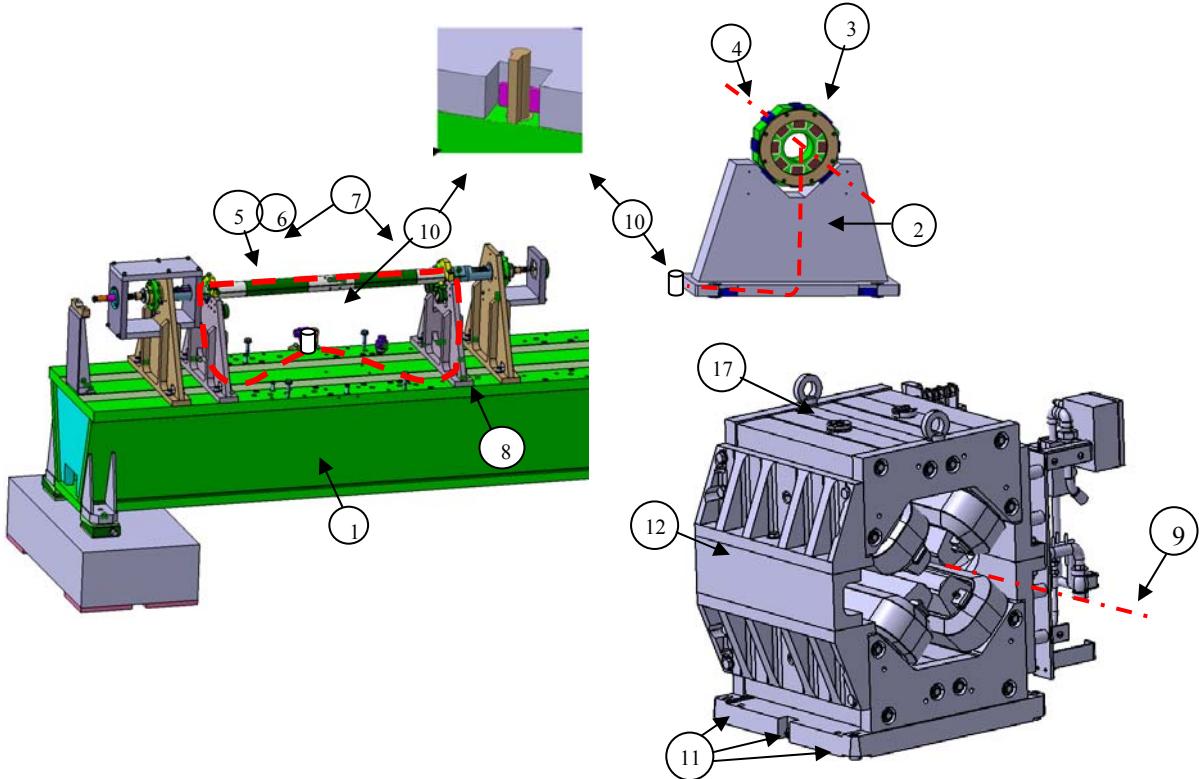


Fig. 91: Quadrupole bench for magnetic measurements: details and metrology loop (red dashed line)

The bench STC can be extrapolated by using its periodic survey with the same reference Qpole shown in Fig. 92 [23]. The magnet comparator has the following STC (according to the DOF), including its mechanical and electronic (sensors) stabilities (Table 1).

Table 1: Magnet comparator STC

X	(few μm , 5 min)	rotation of the system
Z	(few μm , 3 months)	no rotation possible
Θ_s	(10 μrad , 5min)	rotation of the system

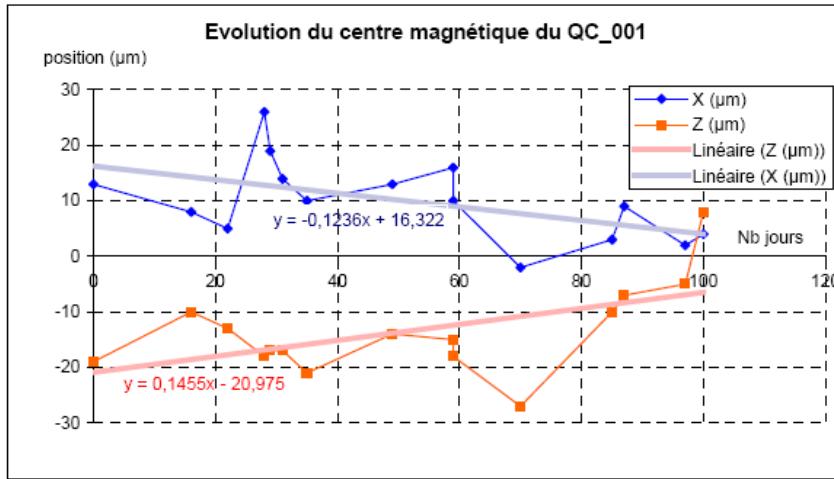


Fig. 92: Periodic magnetic checking of the BMS with a reference Qpole

Table 2: STC analysis of the bench

No .	Component	Action	σ (μm)	Bias	STC = ($\mu\text{m}, t$)	STC = Easy/Diff.
1	Bench	<i>Mech.*</i>			($, \infty$)	E
2	Tool stand					
		<i>Contact* * </i>		ESL	(,min)	E
3	Tool ext. face					
		<i>Mech.</i>			($, \infty$), (,min) = ESL	D,E
4	Tool mag. axis					
		<i>Meas.</i>	5			
5	Coil mag. axis					
		<i>Mech.</i>			($, \infty$)	E
6	Coil rotat. axis					
		<i>Rotat.***</i>	5		($, \infty$)	E
7	Ball bearings					
		<i>Contact</i>			($, \infty$)	E
8	Coil stand					
		<i>Mech.</i>			($, \infty$)	E
1	Bench					

*Mech: mechanical machining

**Contact: mechanical contact

***Rotat: Movement in rotation

10.3 Detection of the Qpole axis

Each Qpole is laid on the bench in order to be measured by the sensor. The whole set of magnets should ‘see’ the sensor and then they can be compared even if bench offsets still exist. The differential layout eliminates them. The shims are changed until the sensor measures a null field as it is on the magnetic axis of the magnet. The corresponding shim is the final one and will equip the Qpole.

The STC analysis must be done for the duration to measure all the magnets, i.e., equivalent to infinity. It shows that two weak points exist: the sensor and the radial pin. The first point is solved by the use of the reference tool. The second one comes from a lack of mechanical design: the pin has to withstand contact with magnets, whose weight is as much as 500 kg, approximately 200 times. Reaching $STC = (\mu\text{m}, \infty)$ is then very difficult. It is important to reliably link it to the bench.

The use of an inclinometer requires the definition of a referential for both STC analysis and calculation. In other words, the bench itself must not vary in inclination over a period of 2 or 3 months. For that, a second inclinometer must be fixed on the girder of the bench. All the inclinometry measurements will be normalized with respect to the slope variations of the bench.

10.4 Fiducialization

The fiducialization is materialized by a kinematical centring system (dot-line-plane) fixed on the upper face of the magnet. The set of functions ‘dot–line’ defines the X, Z, tilt references. The plane function helps with the centring of instruments and is defined by the yoke itself. The survey of the fiducials with respect to the magnetic definition of the Qpole is obtained on the magnetic bench by means of a stainless-steel structure called the ‘Qpole comparator’ (QC) (Fig. 93). This structure presents four electronic dial gauges, a pair at each end, one set in the vertical direction, the other one in the horizontal direction. In addition, an inclinometer is fixed on it. It allows the measurement of the magnetic tilt coming from the bench measurements through the Z shims of the magnet and the control of the lever arm of 365 mm inducing errors in the X direction of the fiducialization results.

Table 3: STC analysis of the Qpole axis detection

No .	Component	Action	σ (μm)	Bias	STC = ($\mu\text{m}, t$)	STC = Easy/Diff.
9	Qpole mag. axis					
		<i>Meas.</i>	5			
5	Coil mag. axis					
		<i>Rotat.</i>	5	(∞)	E	
7	Ball bearings					
		<i>Contact</i>		(∞)	E	
8	Coil stand				($, ToolI^*$)	
		<i>Mech.</i>		(∞)	E	
1	Bench					
		<i>Mech.</i>	5	(∞)	D	
10	Bench pin (surface)					
		<i>Contact</i>		(∞)	E	
11	Qpole shim(s)					
		<i>Mech.</i>	5	(∞)	E	
12	Qpole yokes					
		<i>Mech.</i>		(∞)	E	
9	Qpole mag. axis					

* $ToolI$ covers the items 5, 7, 8, 1, 10: Period between two uses of the Reference tool.

When the QC is centred on the magnet fiducials, the four dial gauges can touch the cylindrical support of the magnetic sensor. The sensor rotates a full revolution and the average of the 64 readings per dial gauge are stored, plus the inclinometer reading. The QC is then rotated by half of the revolution around its vertical axis, realizing a change of the horizontal dial gauges direction and of the

inclinometer zero (note that the direction of the vertical dial gauges is not inverted). The measurements are once again carried out after permutation.

There is currently an ESL configuration: 2×64 points on the coil, 4 dial gauges and 1 inclinometer. These measurements are processed by using the average of the data between the two ends of the QC: it is actually, the magnetic centre of the magnet we want to define and not its real axis. The final result is a set of three offsets per Qpole: dx , dz and $dtilt$.

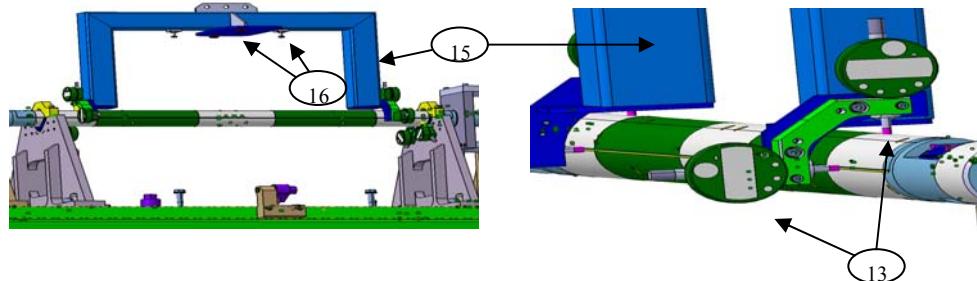


Fig. 93: Qpole comparator

The STC analysis raises the following sensitive points:

- i) The Z measurements are not vertically permuted. As a consequence it is necessary to have a tool in order to check the vertical stability of the QC. A small and simple bench has been designed. It simulates the several contacts of the QC: centering system and dial gauges on an extremely rigid false Qpole. By using it, we can check the Z measurement of the QC including the electronic zero of the dial gauges at any moment. Practically speaking, this checking would be rather cumbersome on a regular basis. Since the stability of the structure is not so bad, we prefer to check only its link with the dial gauges before each use of the QC. The zero is far more unstable. A special wedge is dedicated to that operation.
- ii) The QC is centred on the magnet by means of truncated spheres in contact with magnet fiducials. The distance between the two spheres must not change during the measurement campaign, i.e., the STC is $(\mu\text{m}, \infty)$. A dl variation of that distance induces a $dl/2$ error on the results in the X direction.
- iii) The bench tilt has to be measured each time the QC is used.

Table 4: STC analysis of fiducialization

No .	Component	Action	σ (μm)	Bias	STC = ($\mu\text{m}, t$)	STC = Easy/Diff.
9	Qpole mag. axis					
		<i>Meas.</i>	5		(,30 min)	E
5	Coil mag. axis					
		<i>Rotat.</i>	5	ESL		
13	Dial gauge					
		<i>Tool2*(14)</i>	5		(,30 min)	E
15	QC structure					
		<i>X: Mech. Z: Tool3**</i>			(,30 min) (,days)	E E
16	Trunc. spheres					

		<i>Contact</i>	5		(,30 min)	E
17	Qpole fiducials					
		<i>Mech.</i>			(, ∞)	E
12	Qpole yoke					
		<i>Mech.</i>			(, ∞)	E
9	Qpole mag. axis					

*Tool2: Wedge

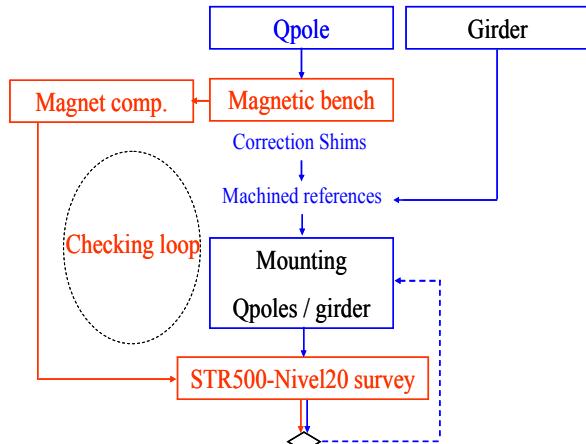
**Tool3: Z bench

10.5 Laser ecartometry of Qpoles mounted on a girder

The Qpoles are mechanically aligned by the contact of their shims with the girder: its upper surface is accurately machined for an altimetric reference (Z , *tilt*) and pins define the X position of the magnets. That operation is a checking of the previous steps: by using the fiducials and their offsets measured with the QC, the Qpoles set should show a perfect alignment. It means that the metrology loop including the shimming operation, the fiducialization, and the laser survey with STR500 should be close to zero (Fig. 95).

Thus the result of this survey gives a good estimate of the quality of the work. The achieved results at SOLEIL were respectively in X and Z : 15 μm and 11 μm . A laser source is set at an extremity of the girder (Fig. 94). The retro-reflector sends back the beam to the telescope and its difference compared to the beam reference gives the dx and dz of the retro-reflector with respect to the beam. An inclinometer completes the survey to manage the lever arms. It is used twice with a rotation of 180°.

Since the measurements are carried out twice, at both ends for the laser, similar to a reversal, there is an ESL configuration for the laser ecartometer measurements.



Figs. 94 and 95: Laser alignment and checking loop

Table 5: STC analysis of laser ecartometry

No .	Component	Action	σ (μm)	Bias	STC = ($\mu\text{m}, t$)	STC = Easy/Diff.
18	Laser beam	\Leftrightarrow^*	\Leftrightarrow	\Leftrightarrow	(,2 min) \Leftrightarrow	\Leftrightarrow
		<i>Meas.</i>	5	ESL		
19	Retro-reflector					
		<i>Contact</i>			(, ∞)	E
17	Qpole fiducials					

		<i>Fiduc.</i> **	10		(∞)	E
9	Qpole mag. axis					
		<i>Bench</i> ***	10		(∞)	E
11	Qpole shim					
		<i>Contact</i>	5		(∞)	E
20	Girder pin (surface)					
		<i>Mech.</i>	5		(∞)	E
21	Girder	\Leftrightarrow	\Leftrightarrow	\Leftrightarrow	(∞) \Leftrightarrow	\Leftrightarrow

* \Leftrightarrow : Common to all magnets on a girder

**Fiduc. : Fiducialization => offsets

***Bench : Detection of the Qpole axis => Shim

10.6 Planimetric Qpole alignment (orbit definition)

The alignment of the Qpoles in the tunnel is carried out by iterations and leads to the final step, the accurate alignment with wire ecartometer and inclinometer. The proposed method is based on differential measurements to eliminate offsets error from the instrument. For more details, see Section 9.1, first example. The measures are common to two girders; it is a way to link them together in terms of a metrology loop.

A redundancy does not appear in the next table: all the girders are measured twice, with their two neighbours. The final least-squares calculation includes measurements from the laser ecartometry and also the TDA5005 measurements (accurate angles + distances). Since the orbit involves all the Qpoles, the metrology loop includes the girder stands and the concrete slab of the tunnel. Note that keeping $STC = (\mu m, \infty)$ for them is not possible, then a new survey and alignment operation whose frequency is called ‘SA’ in Table 6, will be necessary and depends on the level of the acceptable misalignment by the machine physicists.

Table 6: STC analysis of planimetric Qpole alignment

No .	Component	Action	σ (μm)	Bias	STC = ($\mu m, t$)	STC = Easy/Diff.
	Wire	\Leftrightarrow	\Leftrightarrow	\Leftrightarrow	($, 10$ min) \Leftrightarrow	\Leftrightarrow
		<i>Meas.</i>	10			
22	Ecartometer centering					
		<i>Contact</i>	5		($, 10$ min)	E
17	Qpole fiducials					
		<i>Fiduc.</i>	10		(∞)	E
9	Qpole mag. axis					
		<i>Bench</i>	10		(∞)	E
11	Qpole shim					
		<i>Contact</i>	5		(∞)	E
19	Girder pin					
		<i>Contact</i>	5		(∞)	E
21	Girder					
		<i>Mech.</i>			($50, SA$)*	D
22	Concrete slab	\Leftrightarrow	\Leftrightarrow	\Leftrightarrow	($50, SA$)*	D

*SA: Period between two Survey and Alignment operations

10.7 Altimetric Qpole alignment (orbit definition)

Altimetry is obtained with HLS sensors whose reference is a free surface of water available all along the storage ring (Figs. 96 and 97). Linking the Qpole magnetic axis to that surface is a very sensitive operation. It requires the following steps in addition to the fiducialization:

- altimetric measurements from fiducials to HLS vessels,
- linking all the zero sensors together.

The first operation consists in measuring the Qpole fiducials of a girder and the three HLS vessels on the girder with the laser ecartometer and an inclinometer. A long rod laid on the vessels allows one to get the same approximate level for all the points. The laser ecartometer then measures only the vertical direction. The laser ecartometer is used twice, in two symmetrical positions related to the surveyed points. There is an ESL configuration for the laser ecartometer measurements. The STC requirement of the laser ecartometer is similar to the one for its use when checking the Qpole position on the girder.

The zero sensor link has already been described in Section 4.6, third example. Its frequency is once a year, the corresponding drift of sensor zeros is around 10 µm.

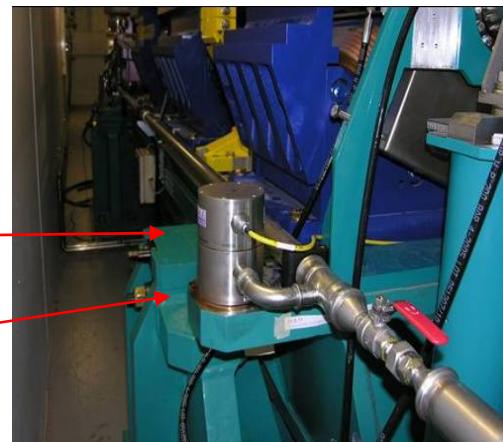


Fig. 96: HLS on girder

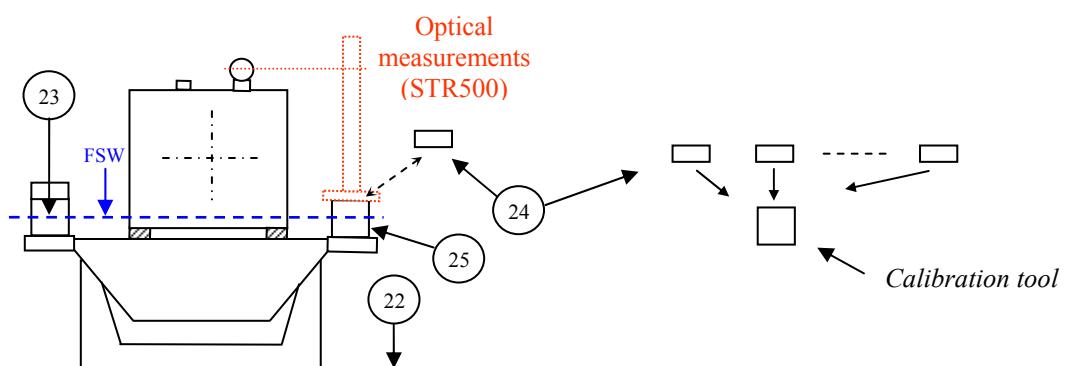


Fig. 97: HLS network and stainless-steel calibration tool

Table 7: STC analysis of altimetric Qpole alignment

No .	Component	Action	σ (μm)	Bias	STC = ($\mu\text{m}, t$)	STC = Easy/Diff.
22	Water	\Leftrightarrow	\Leftrightarrow	\Leftrightarrow	\Leftrightarrow	\Leftrightarrow
		<i>Meas.</i>	5		($, \infty$)	E
24	HLS zero sensor		10		(10, one year)*	E
		<i>Contact</i>			($, \infty$)	E
25	HLS vessel					
		<i>Laser*</i> *	10		($, \infty$)	E
17	Qpole fiducials					
		<i>Bench</i>	10		($, \infty$)	E
11	Qpole shims					
		<i>Contact</i>	5		($, \infty$)	E
22	Girder surface					
		<i>Mech.</i>	5		($, \infty$)	E
20	Girder					
		<i>Mech.</i>			(50,SA)***	D
22	Concrete slab	\Leftrightarrow	\Leftrightarrow	\Leftrightarrow	(50,SA)	D

* With the calibration tool.

** Laser: Laser ecartometer measurements between HLS vessels and fiducials (not described here).

*** SA: Period between two survey and alignment operations

10.8 Planimetric alignment with precise tacheometer (orbit definition)

A tacheometer is an instrument based on a theodolite and equipped with an Electronic Distance Meter (EDM). It measures polar coordinates between two points. The Leica TDA5005 is used to measure the network of points defined by all the Qpole fiducials and wallbrackets. Each instrument for alignment operations has its own range of use. The TDA5005 operates at middle- and long-range of distances for pure alignment on a straight line and for the general shape of the storage ring. The latter cannot be detected with ecartometry.

The TDA5005 is centred on the fiducials of Qpole and brackets fixed on the SR walls. Qpoles are used in preference to benefit from angle accuracy. $STC_{\theta} = (3 \cdot 10^{-4} \text{ deg.}, 10 \text{ min})$. Every detail in the preparation of the campaign needs to reach only 10 min.

The redundancy of the measurements (Fig. 98) leads to a bundle adjustment based on least-squares criteria. Note that the metrology loop includes the slab and girders as the ecartometry and HLS measurements; the corresponding $STC = (50 \mu\text{m}, \text{SA})$, with SA being the duration between two realignment campaigns.

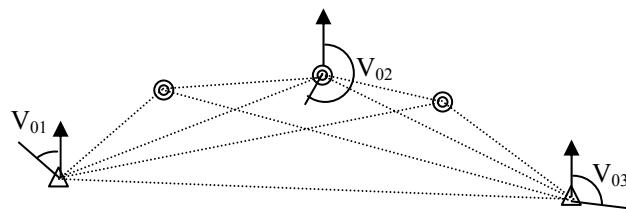
**Fig. 98:** Geodetic network

Table 8: STC analysis of planimetric survey

No .	Component	Action	σ (μm)	Bias	STC = ($\mu\text{m}, t$)	STC = Easy/Diff.
12	Qpole yoke_{TDA}					
		<i>Mech.</i>				
17	Qpole fiducials_{TDA}					
		<i>Contact</i>	5		(,10 min)	E
26	TDA centering					
		<i>Mech.</i>				
27	Zero (dist and angles)				(,10 min)	D _θ **
28	Air (dist and angles)	<i>Meas.</i>	0.12mm & $3 \cdot 10^{-4}$ deg			
19	Retro-reflector					
		<i>Contact</i>	5		(,10 min)	E
17	Qpole fiducials/_{refl}					
		<i>Mech.</i>				
12	Qpole yoke_{refl}					
		<i>Mech.*</i>			(50,SA)***	D
22	Slab				(50,SA)	D
		<i>Mech.*</i>			(50,SA)	D
12	Qpole yoke_{TDA}					

*: Including girders and stands

**: Difficult for angle zero

***SA: Period between two Survey and Alignment operations

Acknowledgement

I would like to thank Fabrice Marteau, member of the Magnetism and Insertion Group, SOLEIL, for his precious help on the topic of magnetic measurements.

References

- [1] B. Schatz, Contrôles des angles, *Techniques de l'ingénieur*, R 1300.
- [2] C. Lasseur, Prospects of close range digital photogrammetry in large physics installations, 5th International Workshop on Accelerator Alignment, Argonne, 1997.
- [3] L. Lahousse, Contribution à la construction de machines de grande précision géométrique, ENSAM thesis, 2005.
- [4] S. Leleu, Contribution à l'évaluation des angles, ENSAM thesis, 2000.
- [5] F. Hennebelle, Détermination des incertitudes de mesures sur machines à mesurer tridimensionnelles, ENSAM thesis, 2007.
- [6] C. Baudoin, Contribution à la qualification d'un processus de fabrication par une approche dimensionnelle : application au pignon forgé conique dit « Net Shape », ENSAM thesis, 2006.
- [7] C. J. Evans, R. J. Hocken and W. T. Estler, Self-calibration: reversal, redundancy, error separation, and ‘absolute testing’, Annals of the CIRP **45** no. 2 (1996).
- [8] M. R. Raugh, Two-dimensional stage self-calibration: role of symmetry and invariant sets of points, J. Vac. Sci. Technol. B **15** (1997) 2139.

- [9] R. Probst, Self-calibration of divided circles on the basis of a prime factor algorithm, *Meas. Sci. Technol.* **19** (2008) 015101.
- [10] D. Martin, The analysis of parasitic movements on a high precision rotation table, MEDSI 2006, Himeji, Japan.
- [11] R. D. Grejda, Use of ultraprecision axes rotation with nanometer level metrology, Pennsylvania State University thesis, 2002.
- [12] D. J. Whitehouse, Some theoretical aspects of error separation techniques in surface metrology, *J. Phys. E: Sci. Instrum.* **9** (1976) 531.
- [13] O. Katowski and W. Salzmann, The angle-measurement in the Wild Theomat™ T2000 (Wild Heerbrugg, 1985).
- [14] E. W. Palmer, Goniometer with continuously rotating gratings for use as an angle standard, *Precision Engineering—Journal of the American Society for Precision Engineering*, **10** no. 3 (1988) 147–152.
- [15] T. Masuda and M. Kajitani, An automatic calibration system for angular encoders, *Precis. Eng.* **11** (1989) 95–100.
- [16] G. Durbec, *Cours de topométrie générale*, T.I. (Eyrolles, Paris, 1969).
- [17] S. Tong, Two-step method without harmonics suppression in error separation, *Meas. Sci. Technol.* **7** (1996) 1563–1568.
- [18] A. K. Jain, Harmonic coils, CAS, Measurement and Alignment of Accelerator and Detector Magnets, Anacapri, Italy, 1977, CERN 98-05, p. 175.
- [19] A. Masi, Advanced measurement systems based on digital processing techniques for superconducting LHC magnets, Universita' Degli Studi Di Napoli Federico II thesis, 2005.
- [20] L. Walckiers, Harmonic coil measurement method, CAS, Bruges, June 2009.
- [21] K.H. Münch, *The Kern E2 precision theodolite* (Kern & Co Ltd, 1985).
- [22] A. Madur, Contribution to the magnetic metrology of accelerator multipole magnets: Synchrotron SOLEIL storage ring quadrupoles, ENSEM thesis, 2006.
- [23] G. Kube, Sensitivity estimation for the PETRA-III beam position monitors based on a boundary element method, DESY Technical Note 2007-01.

Dielectric insulation and high-voltage issues

D. Tommasini

CERN, Geneva, Switzerland

Abstract

Electrical faults are in most cases dramatic events for magnets, due to the large stored energy which is potentially available to be dissipated at the fault location. After a reminder of the principles of electrostatics in Section 1, the basic mechanisms of conduction and breakdown in dielectrics are summarized in Section 2. Section 3 introduces the types and function of the electrical insulation in magnets, and Section 4 its relevant failure mechanisms. Section 5 deals with ageing and, finally, Section 6 gives some principles for testing. Though the School specifically dealt with warm magnets, for completeness some principles of dielectric insulation for superconducting accelerator magnets are briefly summarized in a dedicated appendix.

1 Basic principles

1.1 The electric field

This section will summarize the basic principles of electrostatics in a minimalist approach tailored to an introduction to the dielectric insulation of magnets. The reader is invited to check reference textbooks such as, for example, Ref. [1] and the recommended ‘further reading’.

We first introduce the *electric field* \mathbf{E} as a physical entity capable of producing a force \mathbf{F} on an electrical charge q :

$$\mathbf{F} = q \cdot \mathbf{E}$$

An electric field can be generated in different ways, in particular by electric charges producing a distribution of *electrical potential* V and/or by a time varying *magnetic field vector potential* \mathbf{A} according to the Maxwell equation:

$$\mathbf{E} = -\mathbf{v}V - \frac{\partial \mathbf{A}}{\partial t}.$$

As the Earth’s surface is electrically equipotential, it is convenient to set its electrical potential to zero (electrical ground).

It is important to remark, and this is the main basis of a dielectric insulation system, that what produces forces on electrical charges, responsible for creating electrical currents, is the electric field and not the absolute value of the electric potential. A bird can have a rest on a high-voltage transmission line because the electric field is not high enough to trigger an electrical discharge.

1.2 Definition of ‘dielectric’

The word *dielectric* comes from the Greek ‘dia = through’ + ‘electric’, which was condensed into ‘dielectric’ for ease of pronunciation. In 1836 Faraday [2] discovered that electric charges created by a

high-voltage generator could not create an electric field inside a room enclosed by a metallic envelope (what is since called a Faraday cage). In practice, electric field lines do not ‘pass through’ an electrical conductor, in opposition to what happens with any material not carrying electricity (like glass or air). Faraday thus needed a new term to define such ‘non-electrical-conducting’ materials allowing the electric field to pass through and consulted William Whewell who, in December 1836, invented the term *dielectric*.

It is interesting to note that Whewell [3] coined an impressive series of other important scientific terms, in certain cases by analogy to other words (*scientist*, *physicist* in analogy to the word *artist*), in other cases to help friends facing specific scientific issues (among others Lyell with the terms *Eocene*, *Miocene* and *Piocene* and Faraday with the terms *anode*, *cathode*, *dielectric*, *diamagnetic*, *paramagnetic*, *ion*, *electrode*).

1.3 A basic reminder of dielectric properties of matter

In a dielectric material the conduction and valence bands are separated by a large energy gap so that there are no electrons available for electrical conduction: the interaction between an external electric field and the charges present in the dielectric is described by the so-called *polarization*. This phenomenon is the basis of the explanation Maxwell gave to the ability of a capacitor of storing electrical energy and then providing this energy in the form of electrical current. According to Maxwell, the electric field present in a capacitor deforms the distribution of charges (in some way stretches the atoms/molecules) of the material between the electrodes. When the electric field changes in time, the distribution of charges changes, resulting in what he called a *displacement current*. The total current density flowing through a material can then be expressed as the sum of a conduction and a displacement current density according to the formula

$$J_{tot} = \sigma \cdot E + \frac{\partial D}{\partial t},$$

where σ is the electrical conductivity and D is the total polarization effect or *electric displacement field*, which can be expressed as

$$D = \epsilon \cdot E,$$

where ϵ is the dielectric permittivity of the material.

The dielectric permittivity describes the ability of a material to polarize when submitted to an electrical field, and is typically referred to that of vacuum according to the relationship

$$\epsilon = \epsilon_0 \cdot \epsilon_r,$$

where $\epsilon_0 = 8.854 \times 10^{-12}$ F/m is the ‘dielectric constant of free space’ and ϵ_r the relative dielectric constant of the material.

The interested reader is invited to revisit standard textbooks on the theory of dielectrics. We conclude here by recalling the main modes of polarization.

- Electronic polarization: the electric field modifies the electron density. This mode is important in certain crystals such as Si and in round noble gases. The response to the applied field is very fast.
- Orientation polarization: it is important, for example, in water in which the molecules form electric dipoles randomly oriented. The orientation of these dipoles by an electric field is disturbed by thermal noise viscosity: the response to E is then slower than the previous one and more dissipative.

- Ionic polarization: it mainly includes bulk and interfacial effects. It is slow and dissipative, and typical of ionic crystals (such as NaCl) and heterogenous systems.

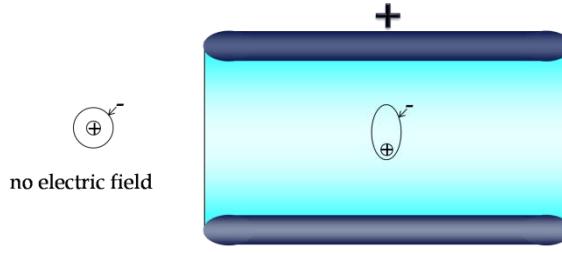


Fig. 1: Stretched electron density relevant to electronic polarization

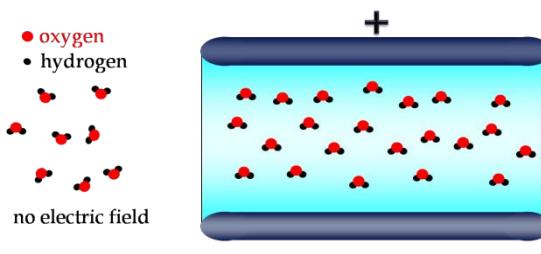


Fig. 2: Orientation polarization of water molecules. Electrons in the covalent bond tend to get slightly denser close to the oxygen atom, which is more electronegative than the hydrogen ones

All these modes are dissipative: changing a charge distribution in a dielectric dissipates energy, resulting in heating the dielectric. In the theory of dielectrics these effects are described as dielectric losses and depend on several parameters, in particular on frequency.

Let us take the example of the orientation polarization of water. If the magnet field changes extremely rapidly in time, the molecules are not capable of following the orientation, and the dielectric losses associated with the field change will be small. Very slow changes in time will also produce small losses because the dissipative cycle is repeated only a few times in the unit of time. In the case of water, at ordinary temperatures the maximum of the dielectric losses are at about 10 GHz, which would suggest using this frequency for microwave cooking. In reality, for this application, the frequency is lowered to 2.45 GHz in order to avoid heating only the first layers of water, allowing time to uniformly heat the food.

1.4 Electrostatics for high-voltage engineering

A treatment of the theory of electrostatics is beyond the scope of this lecture. We recall the essential concepts to underline the importance of geometries and interfaces in dielectric insulating systems.

As we have understood from the above sections, what counts to exert forces on electrical charges is the electric field, capable of producing conduction and displacement currents. From the Maxwell equation in Section 1.1 the electric field distribution E is computed from the electric potential V and the magnetic potential A . Considering the static case, we just need the distribution of the electric potential V , which is computed from the *Poisson* equation

$$\nabla^2 V = -\frac{\rho}{\epsilon},$$

where ρ is the electrical free charge density. In the absence of free charges the *Laplace* equation holds:

$$\nabla^2 V = 0.$$

Let us consider a ‘plane capacitor’ configuration, with infinite parallel electrodes separated by a distance d , at potentials V_1 and V_2 , similar to the one in Fig. 3.

Owing to the plane symmetry, the Laplace equation is mono-dimensional and can be solved as

$$\frac{\partial^2 V}{\partial x^2} = 0 \rightarrow V(x) = c_1 x + c_2 \rightarrow c_2 = V_1; c_1 = \frac{V_2 - V_1}{d}$$

and the electric field is

$$\mathbf{E} = \frac{\partial V}{\partial x} = \frac{V_2 - V_1}{d}.$$

This means that, in case of parallel large electrodes the electric field is constant and uniformly distributed between the electrodes. In real cases, as will be shown with a few examples, the electric field will be higher than that obtained with this formula, for different reasons:

- the shape of the electrode is often far from an ideal case of parallel large electrodes (Fig. 3);
- the material between the electrodes is not homogeneous (Fig. 4).

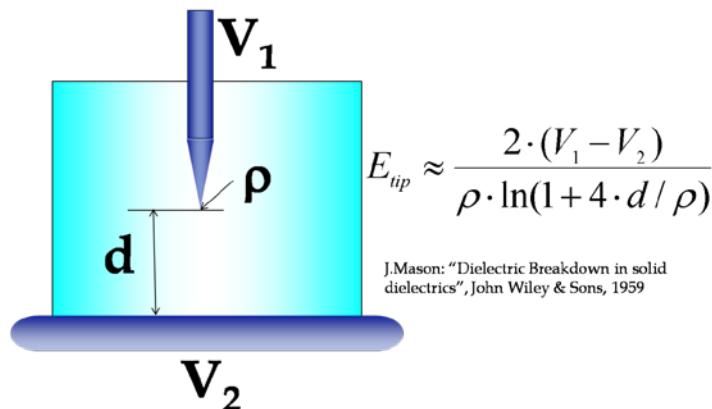


Fig. 3: The electric field of a ‘needle-to-plane’ electrodes configuration at the needle tip

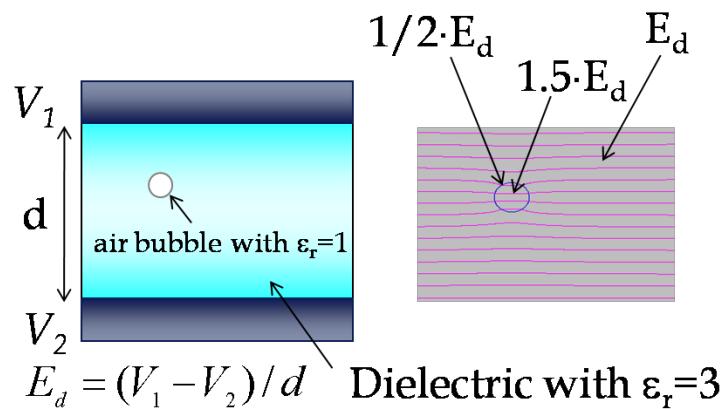


Fig. 4: The electric field is enhanced where the dielectric constant is lower, as in air bubbles possibly present in a dielectric insulator

2 Breakdown in dielectrics

We have seen in Sections 1.2 and 1.3 that a dielectric is a *non-electrical conducting* material, in which macroscopic electrical currents are mainly due to the displacement current.

However, when the electric field is strong enough, the dielectric material may suddenly lose its property of non-conduction, permanently or temporarily, showing an *electrical breakdown*.

In practice we can define the electrical breakdown as an abrupt rise of electrical current under the effect of an electric field. Its causes depend on the medium, the environmental conditions, the geometry, type and material of the electrodes, and on the type and amplitude of the electric field.

The maximum electric field achievable in a dielectric without the occurrence of an electrical breakdown is called *dielectric strength*, typically expressed in kV/mm.

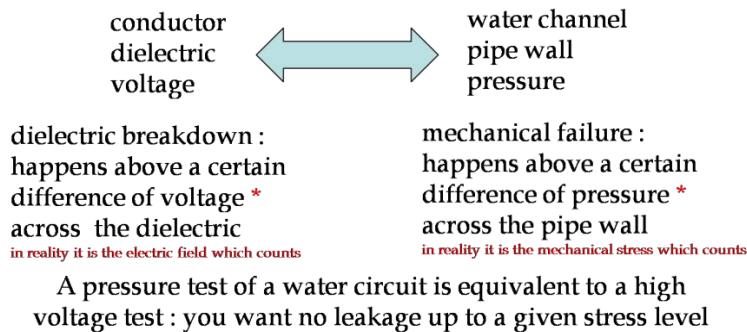


Fig. 5: A non-rigorous but effective analogy between dielectric insulation and water hose wall

2.1 Breakdown in gas

In a gas, free electrical charges under a sufficiently high force can produce ionization and avalanche breakdown by hitting other atoms. At a given temperature, the relationship between the voltage difference V_B between the electrodes at which a breakdown occurs, their inter-distance d and the pressure p of the gas follows the so-called *Paschen law* [4]:

$$V_B = \frac{a \cdot (p \cdot d)}{\ln(p \cdot d) + b}$$

where the constants a and b depend mainly on the gas.

For each gas, at a given temperature, the curve described by this law has a minimum voltage difference between the electrodes below which no breakdown can occur whatever the product between inter-electrode distance and gas pressure. On the other hand, above that voltage difference, one is likely to find combinations of inter-electrode distance and gas pressure at which an electrical breakdown can happen. This is why using gas as a dielectric, though in general very efficient at high pressures or at very low pressures (high vacuum), can become unsafe if during the operation the gas pressure can vary over a large range. Figure 6 shows the so-called Paschen curve for dry air at 20°C, extrapolated from several experiments (from T.W. Dakin, Electra N°32, 1974). At high pressures a further increase of pressure increases the density of the gas but also decreases the mean free path: though the probability of collisions increases, the lower collision energy due to the shorter mean free path is dominant and provides an increase of V_B . At low pressures a further decrease of pressure increases the mean free path but also decreases the density of the gas: though the collision energy increases due to the longer mean free path, the lower probability of ionizing collisions provides an

increase of V_B . In summary, at high pressures the mean free path is dominant, at low pressures the probability of collision is dominant, and there is a product $p \cdot d$ at which V_B has a minimum.

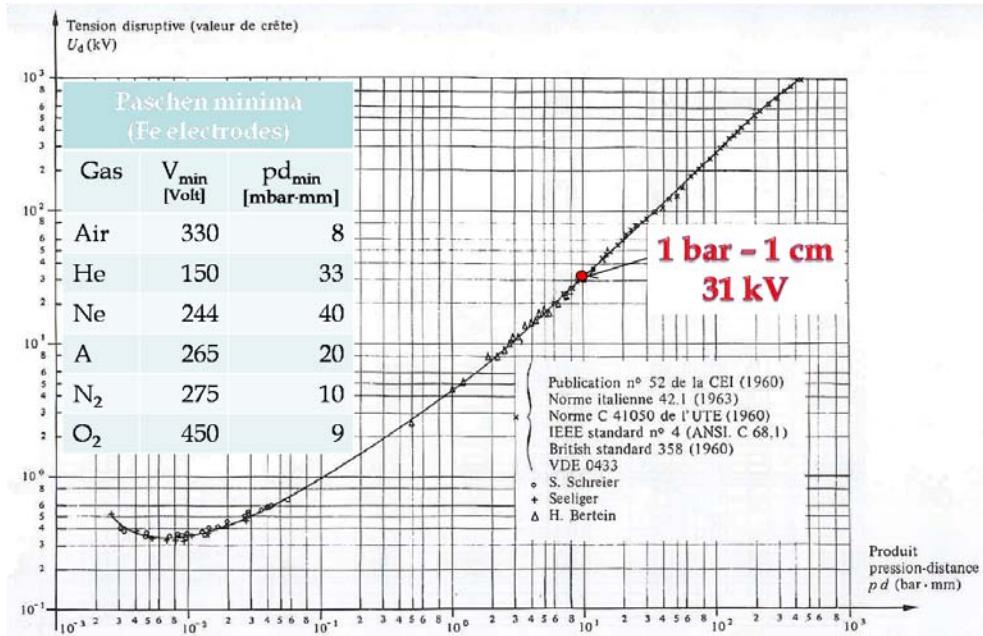


Fig. 6: Paschen curve for dry air at 20°C and Paschen minima for different dry gases

2.2 Breakdown in liquids

Electrical insulating liquids, in general oils, are used in high-voltage transformers, capacitors, switches, and circuit breakers. In many cases they act as both dielectric insulation and coolant. In accelerator components they are used, for example, in fast pulsers for voltages above 30–50 kV. The most important aspect to consider when dealing with liquid insulation is its purity. When an electric field is applied to an electrical insulating liquid, current is initially dominated by impurities. At fields higher than 100 kV/cm electron emission can start at the interfaces of the impurities triggering an ionization process possibly leading to an electrical breakdown.

The breakdown mechanisms in liquids are not yet fully understood. Experimentally the relationship between the voltage difference V_B between the electrodes and their inter-distance d takes the form

$$V_B = A \cdot d^n$$

where $n < 1$ (increasing the distance between the electrodes increases the breakdown voltage less than linearly: this can be justified with statistical considerations on discharge path), and A is a constant.

2.3 Breakdown in vacuum

Does an absence of particles mean no breakdown? In principle yes, but practical vacuum still has particles, or residual gases: in these terms breakdown in vacuum is studied together with the breakdown of gases (remember the Paschen law). In practice, only pressures lower than 10^{-2} mbar can be considered to provide a real dielectric insulation. At higher pressures, especially in the mbar range, such vacuum can even create ideal conditions for electrical discharge. Particular care must be taken to avoid ‘higher pressure’ spots for example due to local degassing. In the worst cases, conditions for sputtering can be created, possibly with the help of the magnetic field produced by a magnet, depositing metallic coating on the surface of insulators leading to the catastrophic failure of the insulation. In practice, at low pressures, many parameters become extremely important, such as the

pressure itself, the distance between electrodes, the applied voltage, the electrode materials, the type and cleanness of the insulating materials.

2.4 Breakdown in solids

The father of the theories of dielectric breakdown in solids is Herbert Fröhlich, a German physicist who, due to the anti-Semitic laws, performed most of his work on dielectrics at the University of Bristol in England. Fröhlich was a theoretical physicist who decided to engage himself in the physical understanding of phenomena like the dielectric breakdown which appeared at that time governed by empirical facts. The review ‘Dielectric breakdown in solids’ of 1939 [5] still remains a milestone for both scientific merit and clarity. We can’t help providing an extraction of his brilliant introduction below:

Owing to its great technical importance, the dielectric breakdown in solids has for many years been a subject of experimental and theoretical investigations. Nevertheless, only in recent years has it been possible to come to a closer understanding of this phenomenon. It is the aim of this article to give an account of these recent developments. One of the most important results of recent research has been the experimental proof of the existence of an intrinsic electric strength. This means that at a given temperature a maximum breakdown strength exists for each dielectric substance which is a constant of this substance, and which is obtained under ideal conditions (homogeneous field, uniform material without weak spots. etc.). Therefore it should be, and has been, possible to calculate this intrinsic electric strength from simple physical constants of the material. In this Report we shall deal mainly with the intrinsic electric strength. There exists also a quite different type of breakdown, the so-called breakdown through thermal instability. This type of breakdown is of importance only in special conditions such as high temperature. We shall give only a very short account of the thermal breakdown, since several reports on it have been published already. In addition, there are the various complex forms of breakdown which occur in industrial insulation; but these, although of great engineering importance, usually reduce finally to one of the two fundamental types considered. The description of industrial breakdowns would be lengthy and is related to the particular properties of complicated substances. Since, also, there is an extensive engineering literature on the subject, it will not be treated here.

The two fundamental types of breakdown mentioned by Fröhlich are:

- the electronic breakdown. With a sufficient energy (above a critical electric field) electrons can cross the forbidden gap from the valency to the conduction band, eventually producing collisions with other electrons and leading to breakdown;
- the avalanche breakdown. As in gases, with sufficient energy (above a critical field) conduction electrons gain enough energy to liberate electrons from the lattice atoms by collisions.

In both cases the breakdown permanently modifies the matter of the failing path.

In practice the electrical breakdown appears below, sometimes much below, the intrinsic limit of the material. Parameters and phenomena affecting the dielectric strength are:

- material type and characteristics (of course)
- type and duration of the applied electric field

- mechanical integrity: in most cases it is a mechanical failure, like a crack, which triggers a fault
- radiation damage: leading to a mechanical failure or to a change of dielectric properties
- chemical actions (like oxidation): typically triggered or accelerated by radiation, temperature
- hydrolysis: in particular when the insulating system is not water-tight
- contamination
- interfaces: in many cases they are a favourite path for a breakdown. At an interface the electric field can be increased due to variations of the dielectric constant, for example, in case of presence of air bubbles, the dielectric strength can be reduced, for example, due to presence of humidity or free ions, as a mechanical discontinuity and interface may lead to a mechanical failure
- progressive erosion of the material creating a breakdown path: starting from air bubbles or propagating along surfaces/interfaces, having different forms (treeing, tracking, partial discharges)
- ageing: some of the above phenomena/parameters produce a progressive transformation over time

3 Electrical insulation in magnets

In an electrical machine, the electrical insulation ensures that current flows only along the conductors and not between individual conductors or between the conductors and the other part of the magnet. We distinguish electrical insulation:

- between coil turns
- between different active parts
- between active parts and ground

A weak electrical insulation may produce:

- current leak with local heating up to melting and possible fire
- progressive damage of the leakage path up to a short circuit
- unbalance of circulating current (possibly with magnetic field distortion)
- autotransformer effect with reduction of magnetic field
- incorrect functioning of protections

We shall keep in mind that the energy stored in magnet circuits is available for any catastrophe!

Solid dielectric materials can be distinguished in three main classes:

- inorganic materials: ceramics, glass, quartz, cements and minerals such as mica, etc.
- organic materials: thermoplastic and thermosetting
- composites: fully organic (aramidic fibres-epoxy tapes) or mixed (epoxy-mica tapes)

We will not go through the hundreds of materials possibly suitable for an electrical insulation: we invite the reader to check textbooks and catalogues and remain open to search for the best solution. We are insisting on this because the knowledge and use of the appropriate material and manufacturing technology is one of the most difficult issues in the science and technology of dielectric insulation.

Too often an easy way is chosen, copying the most popular solutions even in cases for which a different specific material would be more appropriate. Unfortunately, it is a matter of fact that the experience gained in the technologies used for large electrical machines and in other high-voltage devices is not always exploited in magnet manufacture. An example, to make just one, is electrical tracking, which will be treated later, requiring the use of specific material in case the risk of contamination/humidity is important. The same magnet, depending on its operation mode, environmental conditions including radiation and humidity, type and amplitude of the supply voltage, mechanical stresses, may require different materials and technologies for its electrical insulation. It is also important to remember that there are no ‘definitive’ solutions. For example, cables in nuclear plants or accelerator magnets close to target areas are submitted to very high radiation doses and are typically insulated with MgO powders. Their insulation, being mineral, is extremely resistant to radiations, however, it is also very hygroscopic and, though it is typically sealed, it is in general not appropriate for voltages above 1000 V.

4 Failure mechanisms

A dielectric insulation is stressed by several factors, among others: electric (field strength and type), thermal, mechanical, chemical (including oxidation), radiation, and of course contamination.

These factors can produce short and/or long term degradation.

Also, and this is unfortunately in many cases not considered with enough attention, environmental conditions can modify the dielectric system leading to its failure not really because of degradation or ageing, but because in operation the properties of the dielectric system are different from the ones the designer has considered. A typical example is the dielectric insulation in proximity to the hydraulic connection posts, which should in principle be capable of providing the required dielectric strength even in presence of humidity.

That said, in magnets the causes of dielectric failures (electrical short circuit) can be basically schematized in three main groups:

- *insufficient dielectric properties (wrong design)*
- *mechanical failures* breaking the dielectric integrity
- *modification of the bulk or surface dielectric properties*

In case the above events are due to an irreversible, progressive modification of the properties of the dielectric system due to external stresses we speak of *ageing*.

It is not easy to universally define the concept of ageing with respect to nominal characteristics of a material. In this sense, we will give here two examples of different nature:

- the difficulty in defining an ‘intrinsic’ dielectric strength of materials. When an electrical field is applied, if this is above a certain value, even within very short times electric stress can progressively modify the bulk of a dielectric leading to an electrical breakdown. This is why the ‘intrinsic’ dielectric strength, which should be the one the material can withstand before being modified, is sometimes measured with very short pulses, even just a few nanoseconds, obtaining for certain materials values up to several MV/mm. However, shall we really speak of ageing when the degradation takes place over such a short time? Wouldn’t it be better to define an operating dielectric strength, and ageing be defined with respect to that value? This is the practical approach used by most International Standards, and ageing in practice becomes the result of a process taking a reasonably long time and a macroscopically reasonable progressive modification of properties.

- Let us now consider a short circuit that is due to the deposition, along the surface of a dielectric, of humidity. The overall dielectric properties of the insulating system change with respect to a dry situation, however, we believe this is certainly not ageing at least because the modification of the property is not irreversible and, most of all, it is not a modification of the dielectric itself but it is a modification of the insulation system. In case humidity would progressively penetrate in the dielectric the opinions may be more easily shared between ageing and a modification of the system due to an additional element (the water).

A comprehensive overview of failure mechanisms, degradation, and ageing is far beyond the scope of this note and can be found in specialized literature.

We prefer here to briefly recall in the next paragraphs some remarkable examples of failure mechanisms relevant to magnets insulation, which in certain cases are progressive and also fall into the class of ageing. In the next section we recall the basics of the theory of ageing of dielectric materials and insulation systems.

4.1 Mechanical failure

A mechanical failure, even partial or very localized like a hole or a crack, is often the reason for an electrical short circuit. Mechanical failures may be produced in many different ways, from the embrittlement of the material due to radiations to mechanical fatigue. Here we want to stress the importance of the quality of the manufacture by discussing a photo (Fig. 7) taken by the author on a sample of resin-fibreglass composite after a flexural fatigue test.

In the picture it is clearly visible how, on that specific specimen, the adhesion between glass fibres and resin was not good, leading under a mechanical flexural strain to local debonding and consequent concentration of stresses, responsible for premature breakage of certain lots of samples. In a dielectric system, for example, a coil insulation on magnets, the quality and treatment of the fibres to provide a good adhesion to the resin, the type and quality of the resin, the curing cycle, and all procedures to avoid contaminants, discontinuities, critical interfaces, air bubbles, are of extreme importance. Depending on the application, the environmental conditions, the required viscosity during the manufacture, the geometries, different preparation of a similar resin compound may be envisaged, acting for example on the base resin (aromatic, cycloaliphatic, novolak or phenolic), the hardener (amine, anhydride), the accelerator, the flexibilizer, the fillers (Al_2O_3 , MgO , quartz, dolomite) and plenty of other additives that modern technology makes available to us.

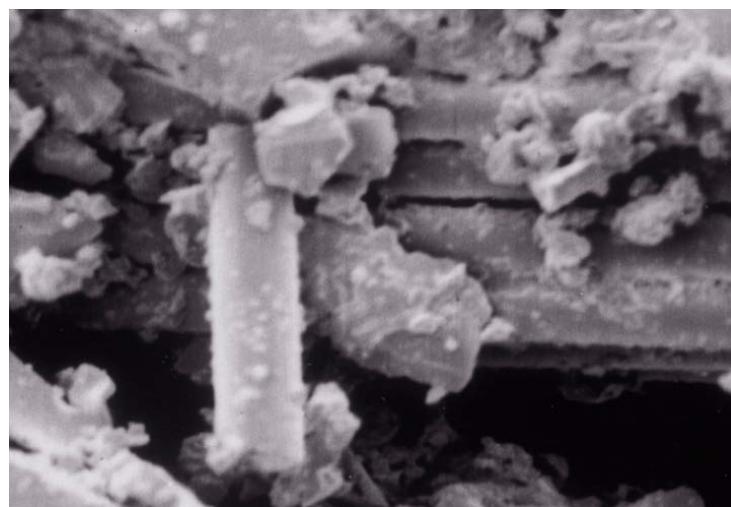


Fig. 7: Electron microscopy of a resin-fibreglass composite after a flexural fatigue test (from: G. Liberti and D. Tommasini, 'Fatigue on fibreglass/resin composites', ICCM, Milan, May 1988)

4.2 Partial discharges

Partial discharges (PD) constitute one of the main mechanisms of ageing, leading to failure of an electrical insulation in continuous operation. In Fig. 4 we already illustrated how interfaces between air and dielectric such as air bubbles and delamination represent volumes where the electric field is higher with respect to that of the surrounding dielectric. This is due to the difference of dielectric permittivity of the insulating material and that of air. But, in these air regions, in addition to the electric field being higher, the dielectric strength of air is much lower than that of the dielectric. If in the air bubble the electric field exceeds the dielectric strength of air, which is about 31 kV/cm for dry air, and lower in presence of contaminants or humidity, then an electrical discharge appears in the bubble, called *partial discharge* because it does not link the two electrodes. In principle a partial discharge would not cause a problem, because it is not a short circuit. For example in aerial transmission lines partial discharges, in the form of the so-called *corona* effect, can be present permanently during operation, producing some dissipation of energy but no or minor degradation of the system. But this is in free air, where the molecules interested by the PD are continuously ‘refreshed’ by new ones without the risk of accumulating discharging elements possibly crossing the two electrodes and producing a real short circuit. For information, we remark that corona can be visible in two different colours:

- a positive voltage creates a uniformly distributed bluish-white cloud,
- a negative voltage creates reddish, pulsing spots.

In a solid dielectric the situation may become critical. The PD in the bubble produce ionization capable of breaking the chemical bonds of the dielectric at the boundary of the bubble, will heat the dielectric, will produce carbonization, will increase the air pressure in the bubble, may trigger chemical degradation: in summary PD will progressively erode the internal surface of the bubble either uniformly or by forming channels until (Fig. 8) a complete electrical short circuit is produced.

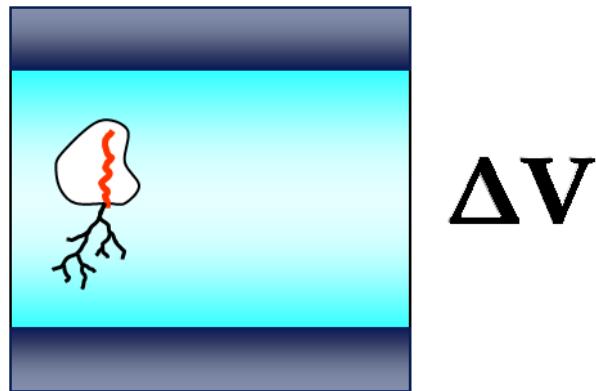


Fig. 8: Schematic representation of the propagation of partial discharges in an air bubble

4.3 Treeing

Treeing constitutes a progressive propagation of a discharge path produced by partial discharges in case of diverging electrical fields, as for example schematically represented in Fig. 3, where the electric field at the needle tip can be orders of magnitude higher than the one corresponding to the plane electrodes geometry at the same inter-electrode distance and the same voltage difference. Treeing typically affects high-voltage cables, in particular in coaxial geometries, where the electric field is higher close to the central conductor than at the outer cable side. If, in proximity to the central conductor, the electric fields exceeds the dielectric strength of the material (in particular if air is present at the interface between the central conductor and the dielectric, triggering partial discharges), then a branched conductive path will progressively extend from the inner electrode to the outer one. It

often happens that high-voltage cables, after several years of operation, show visible electrical treeing which did not yet produce a short circuit.

4.4 Tracking

The progressive extension of a conductive path from one electrode to the other with a spread of sparking channels may happen along a surface: in this case we speak of *tracking* (Fig. 9). This phenomenon is typically triggered by a contamination of the surface of the dielectric, which can be for example humidity, creating quasi-equipotential regions between the electrodes in case the contamination is even slightly electrically conducting, increasing then locally the electric field, possibly further enhanced by the difference of dielectric permittivity between the dielectric and air. A proper choice of the dielectric material depending on the operating conditions is crucial every time tracking may possibly become an issue, as for example in magnet connection posts (Fig. 10).

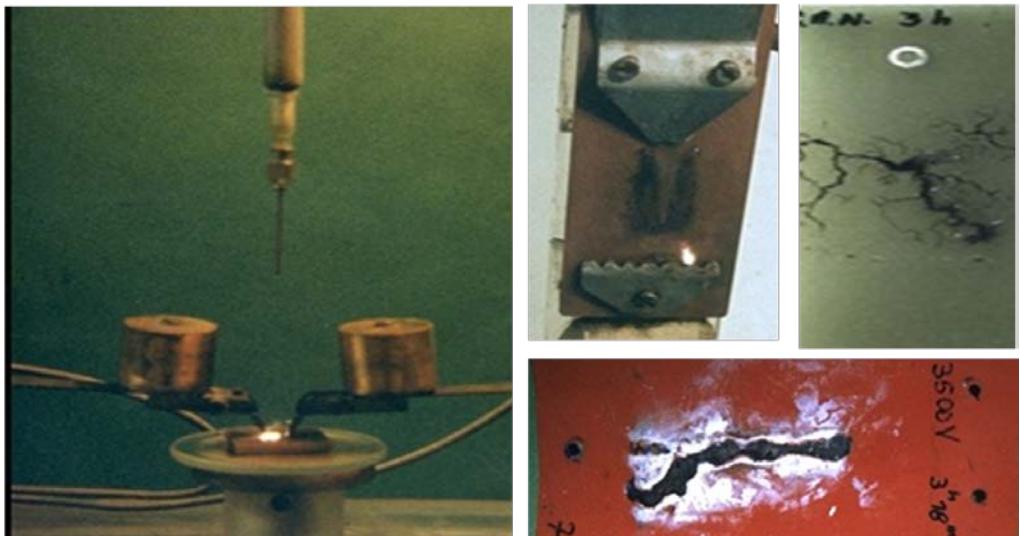


Fig. 9: Examples of tracking test set-up and tracking patterns



Fig. 10: A failure at a magnet connection post due to tracking triggered by humidity

4.5 Radiation

We can distinguish radiation from:

- charged and neutral light particles
- charged and neutral heavy particles
- electromagnetic radiation

These can excite electrons and, in case they have enough energy (typically more than 10 eV) they may produce ionization. The ICNRP (International Commission of Non-Ionizing Radiation Protection) defines electromagnetic radiations, from radiofrequency to ultraviolet, as well as electric and magnetic fields, as *non-ionizing radiations*. In particle accelerators we are typically confronted with *ionizing radiations*, for which we recall two basic concepts: the *range*, consisting in the distance over which a particle of given type and energy loses most of its energy, and the *stopping power* which represents the energy loss per unit of distance covered by the particle. While for electromagnetic radiation and for light particles the stopping power decreases over the path in matter, for hadrons it has a peak, called *Bragg peak*, towards the end of the range. This property is being used in recent medical accelerators to kill cancerous cells (hadron therapy).

Concerning the effects in dielectrics, the electrons produced by the ionization can excite molecules and break bonds forming free radicals, very reactive especially when oxygen is present. New cross-links can also be formed, possibly leading to a modification of the structural properties of the material, which in general becomes stiffer but also more brittle. The covalent bond, typical of polymeric materials, is very sensitive to ionizing radiations.

As in many cases the effect of ionizing radiations depends mostly on the total radiation energy effectively transmitted to the material, we define the *absorbed dose*, expressed in *Gray*, as the energy effectively absorbed by a material per unit of weight:

$$1 \text{ Gy} = 1 \text{ J/kg} (= 100 \text{ rad}) .$$

In general, relatively to radiation, failure of a dielectric insulation is due to loss of mechanical properties (in particular embrittlement) or to the evolution of gases inside the material.

The collection of tables in Fig. 11, extracted from several CERN publications, reports a classification of the resistance of materials with respect to radiations. The tables are barely readable, their intent is to invite the reader to check directly these publications, which provide details on the measurement methods and the criteria used to compile this classification.

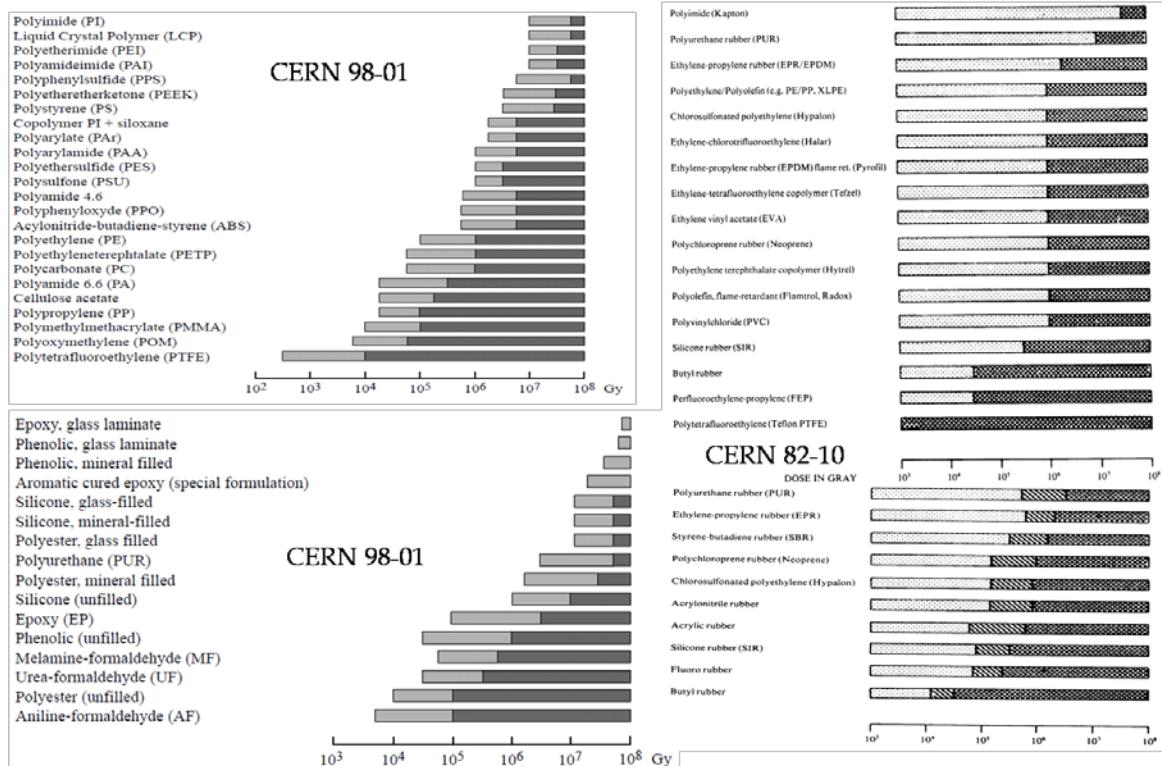


Fig. 11: Compilation of radiation damage data in accelerator materials and components

However, we also know that in certain cases, especially for heavy ions, the electrical properties can be heavily affected by ionizing radiation because of formation of free charges or change of energetic levels of the matter.

Figure 12 refers to the results of an experiment recently carried out in the frame of a common scientific activity between CERN and GSI to understand the difference of electrical damage produced, at given absorbed doses, by different ionizing radiation on polyimide tapes. The effect of heavy-ion radiation on the dielectric strength is noticeable already at very low radiation doses.

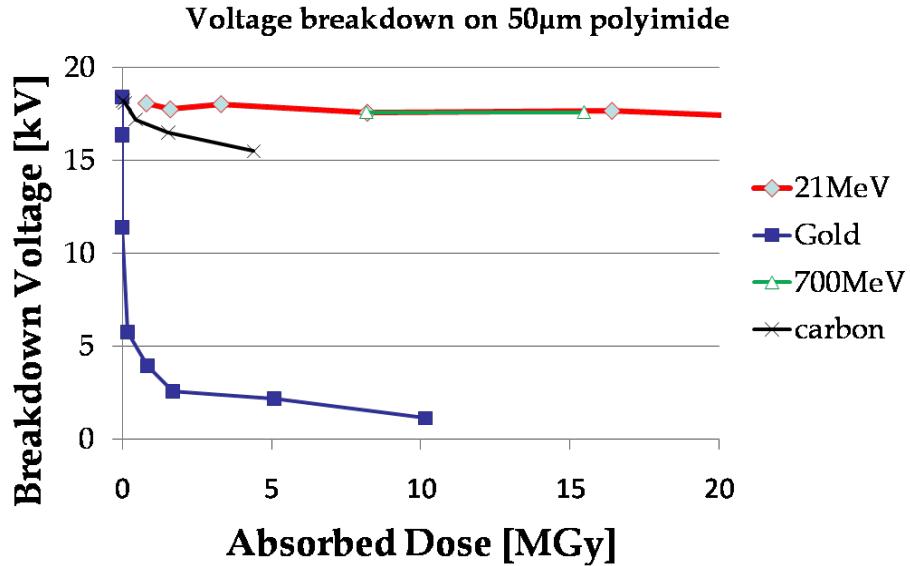


Fig. 12: Dielectric strength of irradiated polyimide [courtesy R. Lopez, CERN, and T. Seidl, GSI]

To conclude this short introduction to radiation damage of dielectrics, we recall situations where radiation doses are extremely high [above 10–100 MGy]. In these cases the technologies used in the magnets have to be robust towards the effects of such high radiation levels, and also the techniques to be used for a possible replacement of a defective magnet have to be conceived at an early stage of the magnet design to reduce the exposure of people to radiations.

Such high levels of radiation require in general the use of special materials:

- inorganic insulations (issues: bonding and moisture absorption)
 - cements and minerals (concrete, mica, quartz)
 - glasses
 - ceramics (oxides of aluminum, magnesium, beryllium)
- new compounds (issues: mechanical properties still to be fully confirmed over a long time)
 - cyanatene ester
 - blends epoxy/cyanatene ester

We recall here the CERN report 82-05 (Fig. 13) for mineral materials and recent developments being carried out in the frame of the ITER project for new compounds.

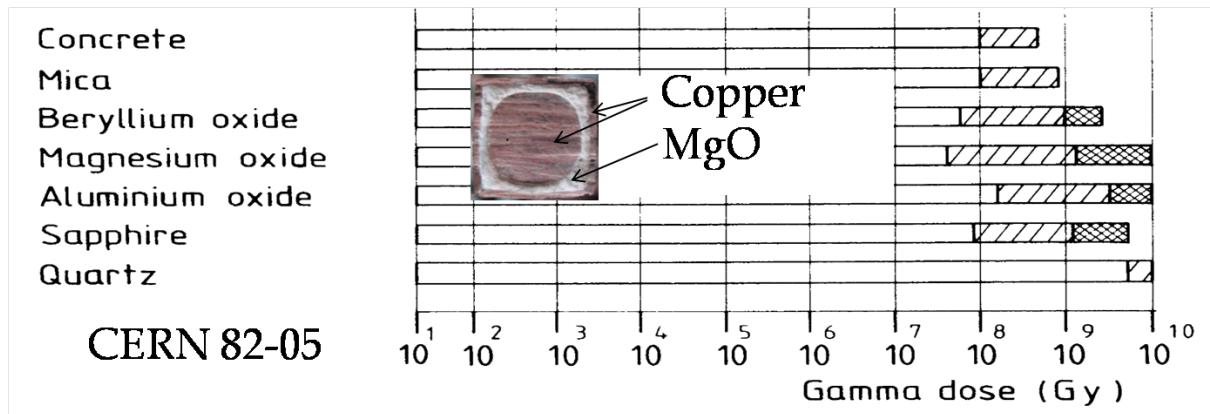


Fig. 13: Radiation damage data for mineral insulations; cross section of a MgO insulated cable.

5 Ageing

We already discussed the definition of ageing, and we have also introduced in the previous sections some of the processes producing ageing. In this section we will recall the basic principles of the theory of ageing, taking as example the thermal ageing, possibly triggering a fault as a consequence of the modification of dielectric properties, and the electrical ageing.

The purpose of this approach is to introduce the reasoning principles applicable to the design and evaluation of dielectric systems for magnets.

5.1 Thermal ageing

The modification of the properties of the material is due in this case to the effects of temperature over time. The effects of thermal ageing were identified and empirically quantified by Montsinger [6], who found that the degradation speed of paper oil insulated transformers doubles every 8°C of temperature increase above the operating temperature of the device.

This finding is generally known as the *Montsinger's rule*, stating that:

a temperature rise of 10°C halves the expected lifetime of an electrical system

Montsinger was also convinced that the mechanism of thermal ageing led to a mechanical failure of the dielectric, which at a certain moment develops a crack representing a path for an electrical discharge. Furthermore, he was not convinced of the idea expressed by Steinmez and Lamme in 1913 [7] that below a certain temperature threshold there is no ageing: according to his idea, as the lifetime halves every temperature increase of 8–10°C, it should also double at every temperature decrease of the same amount.

The real basis of thermal ageing were assessed by Dakin in 1948 [8], who proposed that thermal ageing is due to chemical reactions, like oxidation, modifying the chemical structure of the material. In that case, their speed shall be governed by the Arrhenius law:

$$V_R = A_R \cdot e^{-\frac{E}{KT}}$$

where V_R is the velocity of the chemical process, E is its activation energy, T the absolute temperature, K the Boltzman constant, and A_R a constant.

Starting from this assumption, Dakin could show that the *lifetime L* of an electrical system submitted to thermal ageing, defined as the time of operation at a given temperature at which a given

property (for example the dielectric strength) goes below a given threshold can be expressed in the simple form:

$$L = C \cdot e^{\frac{B}{T}}$$

where C and B are constant.

This allows one to plot the so-called ‘lifetime curves’ as a function of temperature which, in an appropriate scale, are straight lines: ‘Arrhenius lines or Arrhenius plot’:

$$\ln L = \ln C + \frac{B}{T}.$$

This formalization opened the way to accelerated tests, because, in principle, once the two constants C and B are determined by measuring the lifetime at two temperatures (higher than the operating one, to complete the test in a short time), it is possible to obtain the lifetime at any other temperature (Fig. 14).

Dakin understood immediately the potential of his model, but also correctly warned that this should be used within a reasonable range of temperatures, in order that the process remain the same.

For example, below temperatures not capable of activating the chemical process responsible for ageing there is no modification of the material properties, solving the contradiction of Montsinger’s approach towards the absence of thresholds. On the other hand, at very high temperatures, instead of for example a progressive oxidation we may directly burn the material.

In practice Dakin’s model is used to define the so-called *temperature classes*. The temperature class of an electrical material or system is in general defined as the temperature below which the lifetime exceeds 20 000 hours (IEC 60216-1).

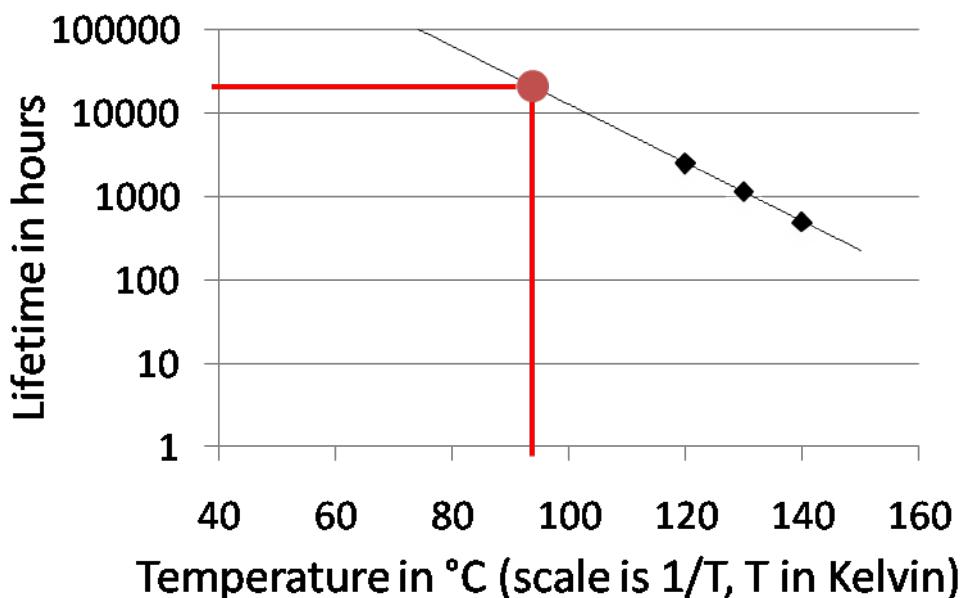


Fig. 14: Arrhenius plot and its use to extrapolate lifetime. The three black squares are experimental lifetime measurements after ageing at three different temperatures, the red lines refer to the extrapolation at a lifetime of 20 000 hours. In this example the temperature class is 90°C.

5.2 Electrical ageing

In this case the modification of the properties of the material is due to the effect of the electric field. Electrical ageing is in general triggered by defects, for example treeing in high-voltage cables, in particular in ac where the alternating electric field may also induce dielectric heating.

The study of ‘pure electrical’ ageing is complicated and requires very high electric fields, typically produced by divergent fields as ‘needle-plane’ geometries, in practice falling into electrical treeing.

In general the lifetime L of a material undergoing electrical ageing can be empirically described by the relationship:

$$L = C \cdot (E - E_0)^{-m}$$

where C and m are constant, E is the applied electric field and E_0 a threshold electric field.

This relationship recalls that of mechanical fatigue, where the lifetime can be replaced by the number of cycles and the electric field by the mechanical stress, such that we may speak of ‘electrical fatigue’.

6 Testing

An electrical insulation shall be tested in such a way as to be confident that it is capable of providing the required insulation levels over specific operation and faulty conditions and during a given period of time. This is rarely possible, because the relationship between dielectric strength and the many parameters entering in the time and conditions of the operation cannot be condensed in a single test.

It is common practice to test electrical components by applying a test voltage equal to the double of the operational voltage plus one kV: for example, a magnet operating at 500 V to ground may be tested at a voltage of 2 kV to ground. However, this general guideline has to be adjusted to the specific application and to the operational and faulty types of possible electrical stresses.

Magnets operating in accelerators are submitted, in addition to electrical stresses, to many environmental and mechanical stresses, which can be tolerated in case the magnet meets at least two conditions:

- it is correctly designed, with the appropriate materials, technologies and geometries;
- it is correctly manufactured.

The electrical tests intended to qualify the design and possibly the manufacture procedures are called *type tests* and, in many cases, they can be destructive. An example of type test are the well-known *crash tests* for cars, which aim at demonstrating that a specific design manufactured according to a specific process with given materials meets certain safety requirements.

Here we will focus on *acceptance tests for magnets*, intended to qualify individual objects, with two examples: the high-voltage test of a conductor to ground and the partial discharge measurements.

6.1 High-voltage test of a conductor to ground

Let us consider a magnet operated at 200 V in dc. We also assume its ground insulation scheme is designed with a minimum insulation thickness of 2 mm. By applying the rule of testing at two times the maximum voltage plus one kV, we would obtain a test voltage of 1.4 kV, which would seem high with respect to the applied voltage.

Can we state that this test voltage is appropriate to ensure that the insulation to ground is well made?

With 1.4 kV to ground the test would pass even if the insulation is cracked: the air in the crack, if dry, would need at least 2 kV and in ideal conditions even more than 3 kV (the dielectric rigidity of dry air, for 1 mm) to fail.

But, who cares if there is such a defect: in any case the magnet can withstand the required voltage!

Unfortunately, real life will care:

- the crack may dramatically propagate during time;
- a wet film may cover the crack surface, due to a water leak or just for ambient conditions.

In both cases the magnet would easily fail.

I leave it to the reader to state whether this test voltage is appropriate or not.

6.2 Partial discharge measurements

Partial discharge may be representative of defects or degradation of the insulation system.

PD can be studied with respect to:

- voltage level (to ground) triggering a flow of charges (PD inception voltage)
- voltage level stopping a flow of charges (PD extinction voltage)
- amount of charge involved in a single discharge
- energy dissipated in a single discharge
- time needed to produce inception or discharge at a given voltage
- characteristics of charge flow (in particular in ac)

In practice, for complex systems like magnet insulation, irrelevant defects may give signals not relevant to a correct diagnosis. On the other hand, certain defects, for example the ones produced by mechanical failure, may remain silent until shortly before failure.

PD techniques are thus important for large numbers (for example transformer windings), because one can count on statistical basis and exclude the components showing abnormal behaviour, or for repetitive and relatively simple geometries such as high-voltage cables.

7 Conclusions

Failure of the dielectric insulation of an electrical machine is often not dominated by electrical stresses but, in most cases,

- by a mechanical failure
- by an unsafe design with respect to environment or operation

A good designer ensures:

- the design correctly considers operation and fault conditions
- the test conditions are capable of identifying a defective manufacture.

References

- [1] J.D. Jackson, Classical Electrodynamics, 3rd ed. (Wiley, New York, 1999).
- [2] J.H. Gladstone, Michael Faraday (Macmillan, London, 1872).
- [3] R.R. Yeo, Defining Science: William Whewell, Natural Knowledge and Public Debate in Early Victorian Britain (Cambridge University Press, 1993).
- [4] F. Paschen, Wied. Ann. **37** (1889) 69 [Ann. Phys. **273** no. 5 (1889) 69–75].
- [5] H. Fröhlich, Dielectric breakdown in solids, Rep. Prog. Phys. **6** (1939) 411–43.
- [6] V.M. Montsinger, Loading transformers by temperature, AIEE Trans. **49** (1930) 776–792.
- [7] C.P. Steinmetz and B.G. Lamme, Temperature and electrical insulation, AIEE Trans. **32** (1913) 79–89.
- [8] T.W. Dakin, Electrical insulation deterioration treated as a chemical rate phenomenon, AIEE Trans. **67** (1948) 113–122.

Further reading

S.M. Naidu and V. Kamaraju, High Voltage Engineering, 3rd ed. (Tata McGraw-Hill, New Delhi, 2004).

M. Aguet and M. Ianoz, Haute tension, volume XXII, *in* Traité d'Electricité.

J.M. Meek and J.D. Craggs, Electrical Breakdown of Gases (Wiley, Chichester, 1978).

R. Arora and W. Mosch, High Voltage Insulation Engineering (Wiley Eastern, New Delhi, 1995).

H.M. Ryan, High Voltage Engineering and Testing [IEE Power Series, 17] (Peter Peregrinus, Stevenage, 1994).

Appendix A: Superconducting magnets

A.1 Introduction

As any electrical device, superconducting accelerator magnets require that active parts be dielectrically insulated from each other and from ground. Though the conductor resistance is null in the superconducting status, during transients, i.e. during magnet energization or during a quench, the voltage difference between adjacent cable turns and between coil and ground can rise to levels approaching or even exceeding the kV range. The large energy stored in superconducting magnets makes the event of an electrical short circuit a potential drama not only for the safety of the magnet itself but in certain cases for the whole accelerator. For example, the magnetic energy stored in a LHC dipole operating at 8.3 T is of about 7 MJ, corresponding to 1.5 kg of TNT (trinitrotoluene) [1]. Just a small fraction of this energy is capable of melting the way from the superconducting cable to an austenitic steel vacuum chamber. In the case of operation in a helium bath, providing a good dielectric insulation may appear easier than in air, thanks to the high breakdown voltage of liquid helium, one order of magnitude larger than that of dry air, i.e., about 30 kV/mm [2]. However, in presence of vapour, as for example following the conductor heating during a quench or for systems not fully immersed in liquid helium, the breakdown voltage of helium is about one tenth that of air at the same pressure, and at sub-atmospheric pressures it decreases till its Paschen minimum of about 150 V [3]. Furthermore the insulating materials have to keep their dielectric and mechanical properties at cryogenic temperatures and under typically large mechanical and radioactive stresses.

A.2 Dielectric strength

In principle the dielectric strength required for a superconducting cable insulation, typically of the order of a few hundred volts between adjacent coil turns, may appear very comfortable with respect to the breakdown voltage of the insulating materials adopted. The reference breakdown voltage of a 25-micron thick polyimide tape is 7800 V and that of fully reacted pre-preg tapes is about 5 times smaller, strongly dependent on the material and on the degree of reaction. The challenges to design and implement a dielectrically reliable cable insulation are:

- in case of non-impregnated coils using cables insulated by tapes, ensure that the topology of the insulating scheme leaves at any place a sufficiently long path along the insulating tapes from one cable turn to the adjacent one. As a guideline, to provide a 1 kV dielectric strength at the beginning of a quench in gaseous helium at 1 bar the required surface length is of the order of 5 mm;
- in case of impregnated coils, ensure before the impregnation that each turn is correctly spaced from the other. In certain cases an electrical discharge test may be indicated to identify a defective assembly;
- in all cases ensure that the insulation will not break during magnet assembly and operation. In most cases, the main issue of preserving the dielectric insulation characteristics consists in preserving its mechanical integrity. As an example, for a given total thickness of the insulation, it may be preferable to wrap in contact with the cable surface the most robust tape, which is less sensitive to damage at cable edges so that the next tapes will lay on a smoother surface. In case of use of pre-preg tapes, or other composites, perform a thorough validation of mechanical characteristics and ageing at cold to evaluate possible degradation during the magnet lifetime. For example, using resin impregnated fibreglass tapes with too low percentages of resin may lead to a quick degradation of the composite after only a few cycles of operation.

Finally, it is important to remark that, in particular in magnets using multilayer coils, the voltage difference between adjacent layers can be much higher than the typical voltage difference between adjacent turns. In these cases it may be preferable to add additional interlayer insulators.

A.3 Heat removal

Another important property of a superconducting cable insulation is its thermal conductivity. The lower the thermal conductivity, the higher the heat which can be removed from the cable to the coolant for a given temperature difference. This limitation of course does not affect internally cooled cables, in which the superconductor is directly in contact with the coolant.

At cryogenic temperatures the study of heat transfer through cable insulation is rather complex, in particular because it depends on temperature, geometry, the status of helium and, depending on the cooling regime, on the status of the material surface. Furthermore, at the boundary between a solid material, such as cable insulation, and liquid helium the temperature is not continuous but there is a temperature difference at the interface, described by a thermal boundary resistance by Kapitza in 1941 [4]. An excellent review on Kapitza resistance can be found in Ref. [5].

Owing to these complexities, summarized in a review by Baudouy et al. [6], there are only a few measurements of the thermal conductivity of polyimide at low temperatures, the one mostly used as reference being that of Lawrence et al. [7]. In case of epoxy fibreglass composites data are even more difficult to interpret because of the variety of the compounds and of the surface status. A good reference can be considered the work of L. Imbasciati et al. [8].

In superfluid helium a powerful mechanism of heat transport can be obtained by providing micro-channels through the cable insulation: this technique has been adopted for the main superconducting magnets for the LHC and allowed to increase by about a factor two the power which can be removed from the cable to helium compared to a similar insulation without microchannels. Recently a new cable insulation scheme has been developed for a possible use for the new inner triplets for the LHC luminosity upgrade, providing a further increase of the heat transfer of about a factor of 5 [9,10].

References for Appendix A

- [1] A. Beveridge, *Investigation of Explosions* (Taylor and Francis, London, 1998).
- [2] J. Gerhold, Dielectric breakdown of helium at low temperatures, *Cryogenics*, **12** (1972) 377–384.
- [3] J. Gerhold and T.W. Dakin, Paschen curve for helium, *Electra*, **52** (1977) 80–86.
- [4] P.L. Kapitza, *Zh. Eksp Teor. Fiz.* **11** (1941) 1.
- [5] G.L. Pollak, Kapitza resistance, *Rev. Mod. Phys.* **41** (1969) 48–81.
- [6] B. Baudouy, Kapitza resistance and thermal conductivity of Kapton in superfluid helium, *Cryogenics*, **43** (2003) 667–672.
- [7] J.L. Lawrence et al., The thermal conductivity of Kapton HN between 0.5 and 5 K, *Cryogenics*, **40** (2000) 203–207.
- [8] L. Imbasciati et al., Thermo-mechanical characterization of insulated and epoxy-impregnated Nb₃Sn composites, *IEEE Trans. Appl. Supercond.* **13** (2003) 1788–1791.
- [9] F. La China and D. Tommasini, Comparative study of heat transfer from Nb-Ti and Nb Sn-3 coils to He II, *Phys. Rev. Spec. Top. Accel. Beams*, **11** (2008) 082401.
- [10] P. Granieri, P. Fessia, D. Richter, and D. Tommasini, Heat transfer of an enhanced cable insulation scheme for the superconducting magnets of the LHC luminosity upgrade, MT-21, Hefei, China, October 2009.

Magnetic measurement with coils and wires

L. Walckiers

CERN, Geneva, Switzerland

Abstract

Accelerator magnets steer particle beams according to the field integrated along the trajectory over the magnet length. Purpose-wound coils measure these relevant parameters with high precision and complement efficiently point-like measurements performed with Hall plates or NMR probes. The rotating coil method gives a complete two-dimensional description of the magnetic field in a series of normal and skew multipoles. The more recent single stretched wire is a reference instrument to measure field integrals and to find the magnetic axis.

1 Introduction

The field of measurement of accelerator magnets has followed the requirements dictated by developments in accelerator technology. Synchrotron Light Sources require stringent magnetic axis alignment. The field quality of superconducting magnets for an accelerator like the Large Hadron Collider (LHC) has been measured to unprecedented precision, including the time and ramp rate dependent effects due to superconductor cables.

Measuring coils and more recently stretched wires are used extensively for accelerator magnets since they measure the parameters seen by particle beams: magnetic field components in the plane perpendicular to their trajectory and integrated over the length of the magnets. Coils and wires allow fast measurements with data already reduced to the requirements.

This field of physics has benefited from the development in electronic components. Fast acquisition voltmeters or recent 18-bit ADCs with sampling times in the range of microseconds increase the bandwidth and the precision of voltage integrated over time, the basis of this type of measurement.

Section 2 describes the classical method of coils flipped by half a turn inside dipole magnets or static coils in pulsed magnets. They are still in use to measure resistive dipoles having flat horizontal aperture and in particular for magnets having a small radius of curvature. In addition, the theory developed for these simple coils must be understood to avoid flaws with more sophisticated measurement methods. Section 3 introduces the concept of coil arrays to suppress the contribution from some harmonic components, a concept fundamental to the accurate use of the harmonic coil method.

Sections 4 to 7 group the Single Stretched Wire (SSW) based methods. They measure with high absolute precision the fundamental parameters of accelerator magnets. Section 4 introduces the equipment and demonstrates that the SSW gives an absolute value of the field integrated over the magnet length with a minimum of calibration concern. Section 5 details how to reference in all dimensions the position of the quadrupole magnet axis and field direction. Section 6 gives the most accurate method to quantify the integrated field gradient and address the issues related to wire deflection due to gravity and magnetic susceptibility. The vibrating wire method described in Section 7 is still subject to interesting developments in order to find the axis of individual magnets aligned on a girder. It is probably the only candidate method to measure small-aperture magnets under development for high-energy linear accelerators.

Sections 8 to 11 cover the harmonic or rotating coil method. This technique gives high resolution and measures in one coil revolution all relevant parameters of any accelerator magnet. Both theoretical and experimental developments allow one to confidently design sophisticated instruments measuring with high bandwidth and precision the full harmonic content of a magnet.

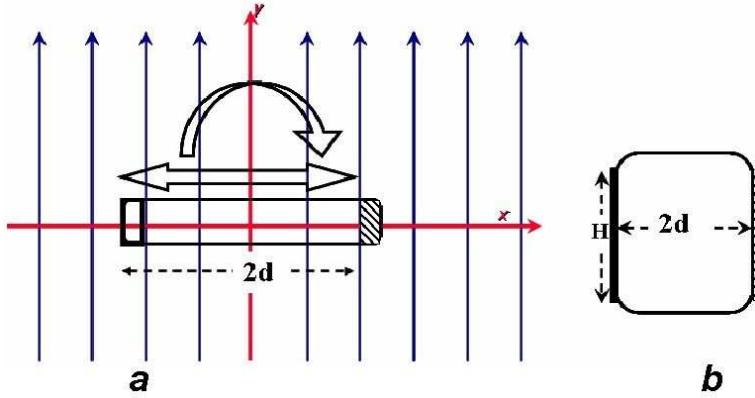


Fig. 1: (a) A simple coil, flipped in a dipole field to measure the central field, displaced laterally to measure the field quality. (b) Square coil with a single layer winding insensitive to the sextupole terms when $H = 2d$.

2 Coils to measure dipole magnets

The following section describes methods that measure only partially the 2D field along the axis of accelerator magnets. The rotating coil method (Section 8 and following) gives a complete measurement of any 2D field expressed in a formal way by a series of complex multipoles. It cannot, however, be used in the following cases which are relevant for most ‘accelerator magnets’ compared to storage rings:

- dipole magnets of small accelerators are bent,
- they mostly have wide horizontal apertures compared to the gap height,
- the rotating coil method does not (easily) measure pulsed fields, i.e., when $\Delta B/\Delta t$ cannot be neglected over one coil revolution period.

2.1 Flip coils for dipole magnet strength

The flux picked by the single-turn coil of width $2d$ sketched in Fig. 1(a) that is longer than the magnet of length L and that rotates by half a turn is

$$\Psi(\pi) - \Psi(0) = 2 \cdot \int_0^L \int_{-d}^d B_y(x) \cdot dx \cdot dl . \quad (1)$$

By assuming either a perfect dipole, i.e., $B_y(x, y)$ constant over the aperture or a coil of width $2d$ small compared to the field errors, i.e., the higher harmonics present in the magnet, the quantities relevant for the particle beam are deduced for the excitation current I in the magnet assumed to be constant during the time to flip the coil:

$$\text{Dipole strength} = \left[\int B_y \cdot dl \right] (I) \quad \text{in [T\cdot m].} \quad (2)$$

$$\text{Transfer function} = \left[\int B_y \cdot dl \right] (I)/I \quad \text{in [T\cdot m/A].} \quad (3)$$

The measuring coil has to be longer than the magnet. A rule of thumb says that the coil should extend outside both magnet ends by 2.5 times the aperture. To quantify the validity of this approximation, there is no other way than to perform a $B(z)$ scanning with a Hall plate or to make a full 3D calculation of the stray field in the magnet ends.

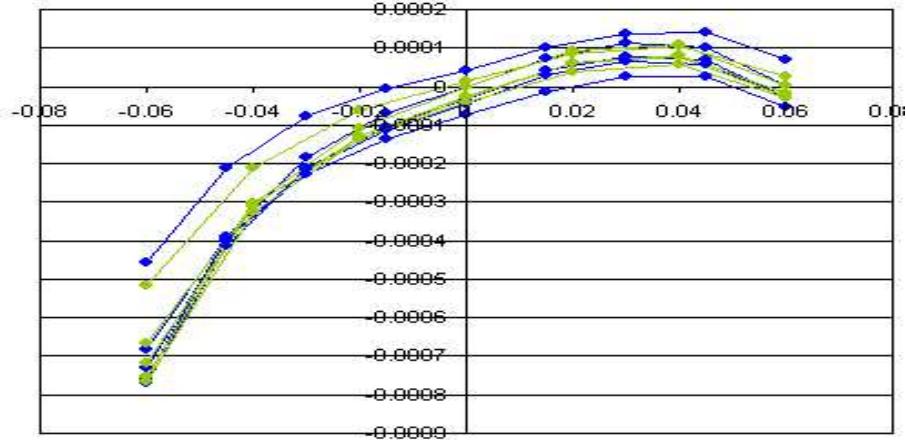


Fig. 2: $\Delta B_y(x)/\Delta x$ measured on 10 CNAO dipole magnets with the 11-coil fluxmeter

2.2 Coils displaced horizontally to measure the field quality

The first order imperfections to consider in a resistive magnet are the variation of the vertical field over the aperture width $\Delta B_y(x)$. The situation is more complex in superconducting magnets where a non-negligible horizontal field component $B_x(x)$ can be present in the horizontal mid-plane. The full 2D complex formalism has then to be applied to describe the field harmonics that can be measured by the rotating coil method covered from Section 8 onward. An horizontal displacement Δx of the coil of Fig. 1(a) leads per unit length along the magnet axis to

$$\frac{\Delta B_y(x)}{\Delta x} = \frac{\Psi(x + \Delta x) - \Psi(x)}{2d}. \quad (4)$$

Figure 2 gives the field quality in the horizontal symmetry plane of 10 dipole magnets for the CNAO facility (Section 3.3). These curves were measured, for efficiency reasons, in pulsed current mode by 11 different static coils rather than 10 horizontal displacements.

2.3 A real coil has finite dimensions for the winding

One way to increase the voltage amplitude at the coil terminals is to increase the width $2d$ therefore picking a vertical field non-constant over this width according to the field quality of the magnet. What is measured is no longer the field at the centre of the coil. The real part of Eq. (29) gives

$$B_y(x, y) = B_1 + B_2 \cdot x + B_3 \cdot (x^2 - y^2) + \dots . \quad (5)$$

Integrating over the coil width going from $[x - d]$ to $[x + d]$ leads, with the hypothesis of a negligible winding height, to

$$\Psi(x) = B_1 \cdot 2d + B_2 \cdot 2dx + B_3 \cdot (6dx^2 + 2d^3) + \dots . \quad (6)$$

This equation indicates that a measurement with a half-turn flip is never sensitive to the even harmonics (quadrupole, octupole, etc.) as long as the axis of rotation is centred in the middle of the coil. On the other hand, the sextupole component enters in a dipole strength measurement:

$$\Psi(\pi, x = 0) - \Psi(0, x = 0) = 2.(B_1 \cdot 2d + B_3 \cdot 2d^3) + \dots . \quad (7)$$

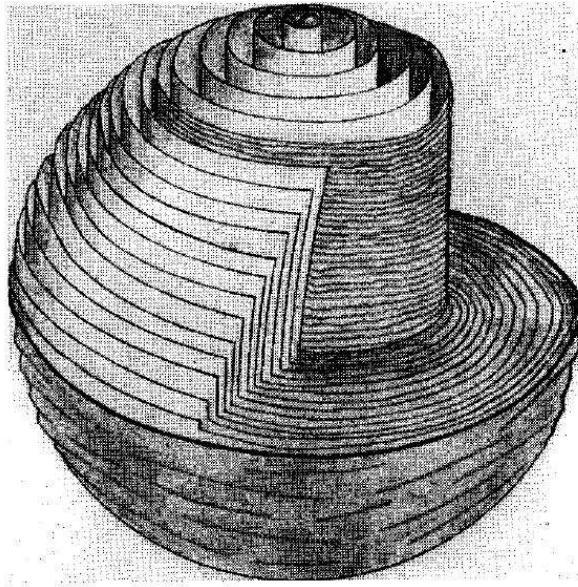


Fig. 3: A fluxball, composed of cylindrical windings, measures the field at the central point of the ball

All multipole terms enter with different and varying sensitivities when measuring the field quality by a displaced coil:

$$\Psi(x + \delta) - \Psi(x) = B_2 \cdot 2d\delta + B_3 \cdot 6d \cdot (2x\delta + \delta^2) + \dots . \quad (8)$$

Coils with several turns are used to further increase the sensitivity. The winding has then a non-zero cross-section. The coil of Fig. 1(b), with coil width equal to the winding height is easily calculated to have no sensitivity to the sextupole B_3 terms. The issue detailed by Eq. (7) for flip coil measurement is therefore limited to higher orders: starting with the decapole term present in the magnet.

This way of avoiding sensitivity to higher multipole terms was generalized in theory long ago with the fluxball [1]. This fluxball coil, as sketched in Fig. 3 seems difficult to manufacture but picks up a flux equal to the field at the centre of the ball and independent of the spatial harmonics present. A coil approximating this possibility of measuring the field central to a highly sensitive coil probe has been proposed in Ref. [2].

2.4 Static coils in pulsed fields

Most accelerator magnets have to be measured in fast current ramping conditions. The coil of Fig. 1(b), insensitive in that case to terms lower than the decapole, is an easy tool to use in fixed position in a field pulsed from 0 to nominal value. Modern integrators with large bandwidth and time resolution connected to such a coil can give the full $B_1(t)$ curve and measure saturation effects of the iron yoke. Hysteresis and eddy current effects can be separated by measurements at different ramp rates. One precaution deals with the remanent field of the magnet being measured, i.e., the field at zero current value. Three ways can be used to solve it:

- have a bipolar power supply to perform symmetric sweeps from negative to positive maximum current;
- demagnetize the magnet first, with either a bipolar supply or a supply with an inverter;
- measure the remanent field with a flip coil or Hall plates for instance, low accuracy is sufficient in most of the cases.

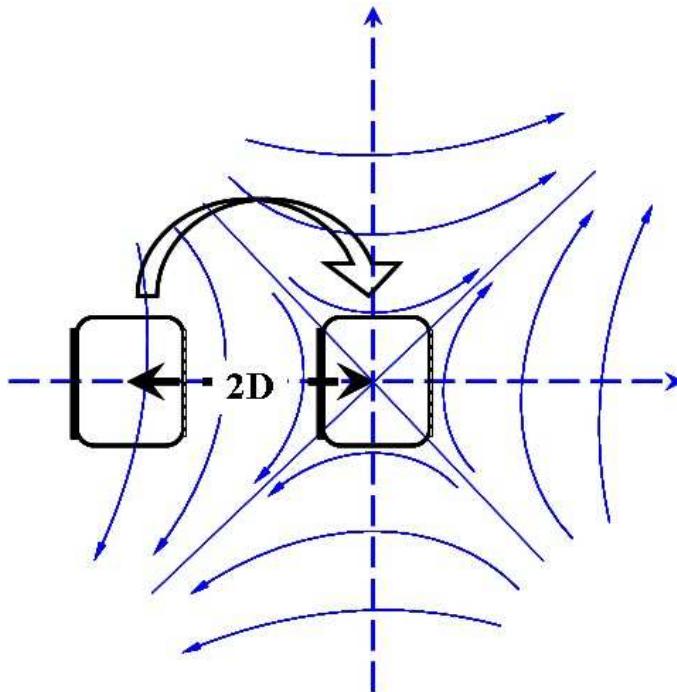


Fig. 4: Coil set to measure the gradient and field quality of a quadrupole magnet

3 Arrays of coils to measure quadrupoles or higher order multipoles

3.1 Two-coil array for quadrupole strength and field quality

The measurement of the strength of a quadrupole magnet could be done by the displaced coil method of Section 2.2. High accuracy requires high precision displacement of the coils and the measurement of the field quality, i.e., higher order multipoles, leads to a uselessly complex mathematical treatment. The array of two coils connected in electrical opposition and sketched in Fig. 4 is often used to give the quadrupole strength by flipping them about their symmetry axis. The coil array, i.e., the distance D , can be calibrated once in a known quadrupole. The field quality is deduced by displacing, along the x and y axis, the array in a quadrupole in the same way as the single coil displacement to measure the field quality of a dipole magnet Section 2.2). Equation (9) is valid as long as the two coils have an equal effective surface.

$$B_1(x + D) - B_1(x - D) = 2 \cdot D \cdot G(x) . \quad (9)$$

The idea of using two coils connected in opposition in order to cancel the contribution of the dipole component is appreciably more effectively used with the rotating coil method (Sections 9.3 and 9.5). In that case, the compensation scheme suppresses the main harmonic of the magnet. It is therefore much more powerful for accurate field quality measurements. The rotating coil equipment is more complex to manufacture but is preferred for quadrupole magnets that are straight and have a circular aspect ratio. The present method is, however, mentioned as useful for combined-function magnets and pulsed magnet measurements.

3.2 The Morgan coil for pulsed magnets

G. H. Morgan [3] proposed a complex coil array that can measure any multipole magnet, or magnet component, in pulsed mode, i.e., with static coil array. Identical coils are located around a cylinder frame with the symmetry to be measured. The number of coils is equal to the multipole order to be measured: three coils for a sextupole, etc.

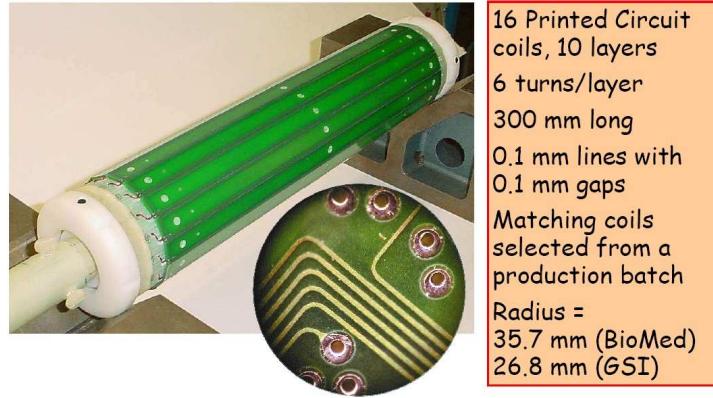


Fig. 5: Measuring shaft with 16 individual coils made with printed circuit technology. A large number of harmonics can be measured in pulsed mode by using various connection schemes (courtesy of A. Jain, BNL).



Fig. 6: (a) The CNAO fluxmeter with 11 curved coils, (b) inserted in one of the CNAO dipole magnets

Figure 5 details an array of coils able to measure a large number of harmonics by different connection schemes to put in series the individual coils. Small discrepancies between the individual coils can be eliminated by turning the coil frame at different angular positions. One should not forget that a coil scheme sensitive to a multipole of order n is as well sensitive to order $n \cdot (2m + 1)$. Reference [4] describes the full theory of this technique.

3.3 A dedicated coil array to measure curved dipoles in pulsed mode: the CNAO fluxmeter

A dedicated array of 11 curved coils has been assembled to measure the CNAO dipole magnets [5]. These 1.5 T dipoles have a bending of 22.5° and operate with a field rise time of 2 seconds. Eddy current effects present in particular in the ends of the magnet yoke have to be taken into account for both the field integral as a function of current and the field quality over the 130 mm useful aperture.

Figure 6(a) shows the 11 coils fixed on the frame that can be entered from a zero field region into one magnet end to measure the remanent field, or measure in static position during the field ramp [Fig. 6(b)]. A cross-calibration 12th coil can be mounted on top of each individual coil of the fluxmeter. This cross-calibration gave correction factors to equalize the effective width of the 11 coils within 10^{-4} relative accuracy.

Note that all correction factors described in Section 2.3 have to be applied to calculate the field errors in terms of multipoles. Figure 2 gives the vertical field component in the horizontal symmetry plane of 10 CNAO dipole magnets measured with this fluxmeter, taking into account these systematic errors.

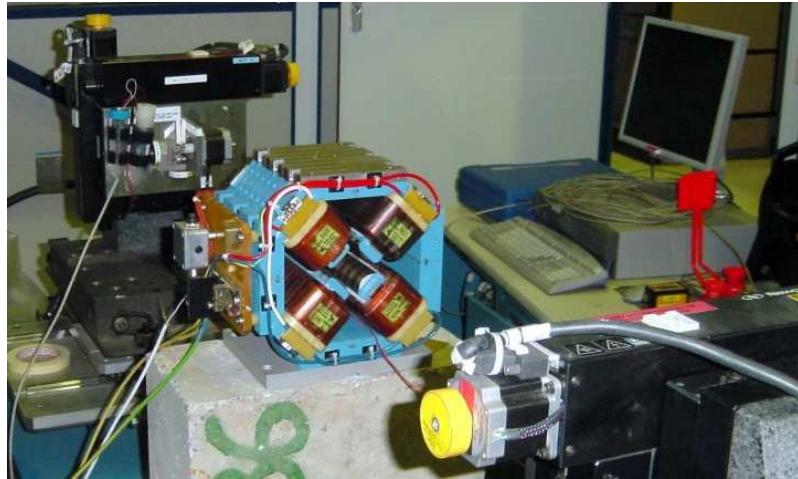


Fig. 7: The SSW equipment installed to measure a quadrupole magnet. The wire is displaced in the quadrupole aperture by 2D high-precision stages on both sides.

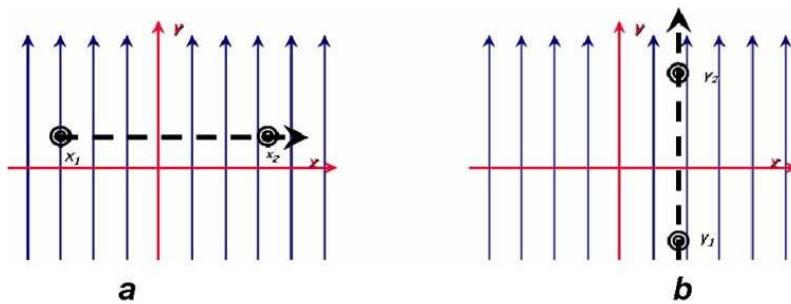


Fig. 8: (a) The single stretched wire displaced horizontally to measure the field integrated over the dipole length. (b) A vertical displacement gives zero signal if displacement and field direction are parallel.

4 The single stretched wire technique with a dipole magnet

Particle beams are sensitive to field integrated along their trajectory. The SSW (Single Stretched Wire) method consists of a high tensile conducting wire moved inside the magnet aperture by precision displacement tables. CuBe wires 0.1 mm thick are commonly used. The stages at both sides are assumed to move by precisely the same amount. The return wire is kept fixed, as much as possible in a field-free region. The flux lines crossed during this displacement, $\Psi(x_1, x_2)$, and measured by a voltage integrator give the field integrated over the displacement, $d = x_2 - x_1$, and over the SSW length, L_w . Figure 7 shows the system developed by FNAL [6] and mounted to measure a quadrupole magnet. When measuring a perfect dipole ($B_y = Cst$, $B_x = 0$) as in Fig. 8(a) the integrator gives

$$\Psi(x_1, x_2) = \int_0^L \int_{x_1}^{x_2} B_y(x, l) \cdot dx \cdot dl = d \cdot \int B \cdot dl . \quad (10)$$

The measurement accuracy of the dipole strength is linked to the calibration of the integrator gain ($10^{-4} - 10^{-5}$) and to the precision of the mechanical displacement. Commercially available stages reach an accuracy better than $10^{-4} = 1 \mu\text{m}/10 \text{ mm}$. This method was cross-checked 30 years ago against NMR mapping and gave agreement within few 10^{-5} .

Similarly, a vertical displacement in a dipole aligned vertically gives a zero flux variation [see Fig. 8(b)]. It is the simplest and most accurate method of finding the field direction of a dipole: accuracies of 0.1 mrad are commonly reached.

This method is simple and efficient to measure the first integral of wigglers and undulators. This first integral value, given by Eq. (10), corresponds to the angular deflection of the beam and is tuned to be zero in the relevant cross-section of the magnet. It has been complemented by the pulsed wire technique to measure and tune the second field integral value [7, 8]. Travelling Hall plate based measurements are nowadays preferred since in addition they give more details on the regularity of the undulator periods.

5 Align a quadrupole with the single stretched wire

The single stretched wire technique is relevant to finding the axis, main field direction, and longitudinal position of a quadrupole magnet [6]. These three alignment steps will be treated separately so as to be easily understood. An automated acquisition system is better for iterating changes in the reference position of the wire for both stages according to full measurement cycles until an accurate coincidence between wire and magnet axes is reached.

A pure quadrupolar field is defined by

$$B_y = G \cdot x \quad , \quad B_x = G \cdot y . \quad (11)$$

The magnetic axis is defined by the line where

$$B_x = B_y = 0 . \quad (12)$$

The main field direction is defined by the symmetry planes:

- $B_x = 0$ in the horizontal symmetry plane,
- $B_y = 0$ in the vertical symmetry plane.

Moving a SSW vertically from position y_1 to position y_2 [Fig. 9(a)] gives the measured flux of Eq. (13), i.e., a parabolic dependence. The effective length L_{eff} hides the integral over the magnet length.

$$\Psi(y_1, y_2) = \int_0^L \int_{y_1}^{y_2} G \cdot y \cdot dy \cdot dl = L_{\text{eff}} \cdot \frac{G}{2} \cdot (y_2^2 - y_1^2) . \quad (13)$$

A further correction must be applied to take into account the wire sagitta that could reach millimetres for a 10 to 15 m long distance between the stages. The system of Fig. 7 incorporates a driver to make the measurement with different wire tensions (Fig. 11). The measurement data are extrapolated to infinite tension. This will be detailed in Section 6.1 since this error source is more detrimental for the measurement of the strength of a quadrupole. To obtain accuracy in the result requires therefore time, even with fully automated equipment and procedures, since loops at different tensions are internal to iteration to align the wire coordinate system to the magnet axis. An accurate measurement of the gradient can only take place after a full alignment procedure.

5.1 Align the quadrupole axis and field direction

Equation (13) tells us that two measurements are needed to find y_c , minimum of the parabola giving the horizontal plane where the axis is located. The method in use is to measure in an iterative way the fluxes over two equal intervals then displace the central point of the measurement y_c until the following condition is reached:

$$\Psi(y_c, y_c + d) = \Psi(y_c, y_c - d) , \quad (14)$$

$$(y_c + d)^2 - y_c^2 = (y_c - d)^2 - y_c^2 . \quad (15)$$

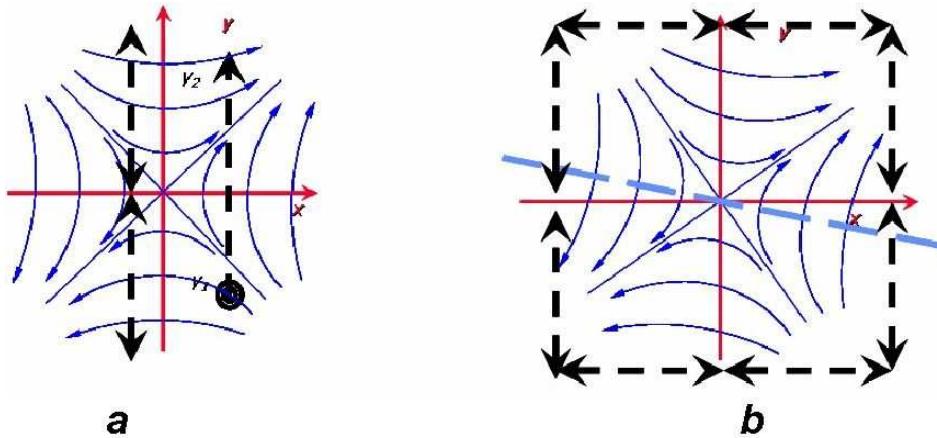


Fig. 9: (a) The single stretched wire displaced vertically in a quadrupole. Two measurements are needed to find the minimum of the parabolic dependency, i.e., the horizontal symmetry plane. (b) The axis and field direction (i.e., tilt of the quadrupole field) are found with eight measurements.

The point where $B_x = 0$ is found for a given vertical line. The main field direction is found by displacing this movement line along x , and the vertical symmetry plane is found by equivalent horizontal displacements. A full measurement is done by the eight displacements of the external square of Fig. 9(b) and repeated until the SSW coordinate system coincides with the quadrupole axis. It is obviously important that the two stages move parallel to each other.

5.2 Stretched wire non-parallel to the quadrupole axis

Figure 10(a) sketches, for the vertical direction, how to tune the parallelism between the SSW and magnet axes. A displacement of one stage going from an angle α_1 to α_2 with respect to the magnet axis gives a flux:

$$\Psi(\alpha_1, \alpha_2) = \int_{\alpha_1}^{\alpha_2} \int_{L_1}^{L_2} G \cdot \alpha \cdot l \cdot dl \cdot d\alpha = G \cdot \frac{\alpha_2^2 - \alpha_1^2}{2} \cdot \frac{L_2^2 - L_1^2}{2}. \quad (16)$$

The effective length of the magnet ($L_2 - L_1$) should be known, and the middle of the magnet is supposed to coincide with the middle of the wire. This is, however, of minor importance since this method, similarly to the one of Section 5.1, implies a series of two symmetric displacements iterated until the parallelism is found, i.e., until they give equal flux values.

5.3 Measure the longitudinal location of the magnet

Once the magnet is fully centred and the axis is parallel to the wire reference position, anti-parallel movements of the wire give the longitudinal position d of the magnet with respect to the middle of the wire length L_w [Fig. 10(b)]. The angular movement α has to be symmetric with respect to the magnet axis. The flux for such angular movement is given by

$$\Psi(\alpha_1, \alpha_2) = \int_{\alpha_1}^{\alpha_2} \int_{L_1}^{L_2} G \cdot x(\alpha \cdot l) \cdot dl \cdot d\alpha = G \cdot \frac{\alpha_2^2 - \alpha_1^2}{2} \cdot d \cdot L_{\text{eff}}, \quad (17)$$

with

$$x = \left(l - \frac{L_w}{2} \right) \cdot \alpha, \quad L_1 = \frac{L_w - L_{\text{eff}}}{2} + d, \quad L_2 = \frac{L_w + L_{\text{eff}}}{2} + d.$$

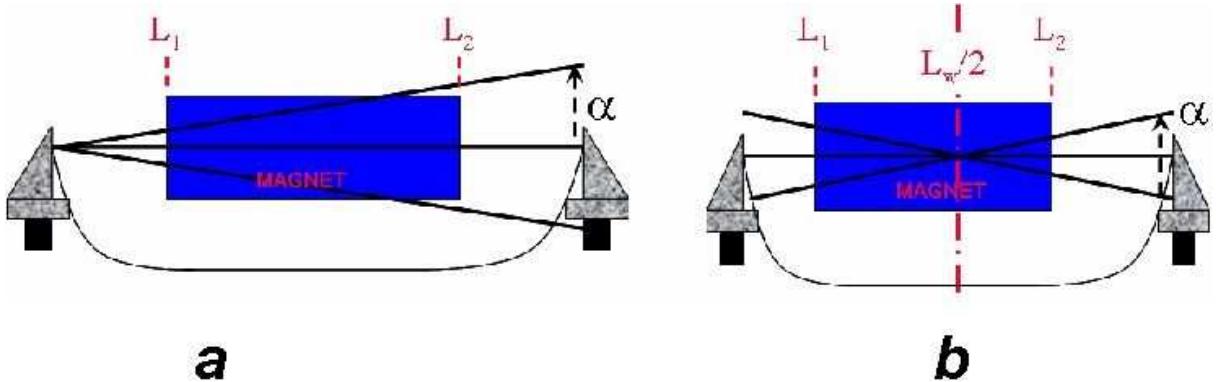


Fig. 10: (a) Two angular displacements give the same result if done symmetrically with respect to the quadrupole axis, i.e., when this axis and the reference line of the SSW coincide. (b) The same angular displacements with the fixed point at the middle of the SSW length give the distance between this SSW middle and the middle of the magnet length.

5.4 Automatic alignment of the SSW reference system to the quadrupole axis in all directions

The system has to be designed to automatically align the SSW with respect to a quadrupole according to these various principles put together. The SSW reference system is aligned iteratively, by one compound measurement with eight co-parallel and eight counter-parallel displacements according to Fig. 9(b), followed by modifications of the SSW reference system. This alignment campaign is tedious since each of these 16 data sets has to be done with different tensions of the wire to extrapolate the results to infinite tension, as detailed in Section 6.1. The longitudinal location is calculated with the last measurement data. The precision obtained with the FNAL equipment for wires up to 16 m long is remarkable: about 0.1 mm for the axis at both magnet ends and 0.03 mm for the integral, 5 mm for the longitudinal position.

6 Measure the integrated strength of a quadrupole with the SSW

The difficulties encountered when measuring a quadrupole strength are that the wire has a natural deflection in the millimetre range if the wire or magnet length reaches several metres. In addition, it is difficult to find CuBe wire that has zero magnetic susceptibility. Unfortunately the industrial standards describing this type of material rarely include the impurity content of non-zero susceptibility. Several batches purchased from the same manufacturer could have appreciably different magnetization and the only way to sort the best batch is to test with a permanent magnet. In conclusion the wire deflection depends on the position in the quadrupole cross-section, in both amplitude and direction. The accuracy is also limited by the high order multipoles present in the magnet that perturb the value of $G \cdot L_{\text{eff}}$ obtained from Eqs. (13)–(15). The only way to estimate accurately that error source is to perform a full measurement of the multipole content for instance with the rotating coil system (Section 8). These perturbations grow with the distance from the magnet axis, and a detailed estimation is needed to either limit the range of the displacement allowed or to include relevant correction factors in the data analysis. In practice, the estimation of the quadrupole strength is obtained from the last set of data measured once the SSW reference system is fully aligned with the quadrupole axis (Section 5).

6.1 Extrapolate the results to infinite tension of the wire

The sagitta due to gravity can easily amount to millimetres with a CuBe wire at maximum tension. The wire tension is measured by the vibration frequency f to get enough precision. This vibration can be induced by the stage motors giving a horizontal or vertical kick and is sensed via the voltage across the wire since it oscillates in a magnetic field. An extrapolation to infinite tension is usually done by looking

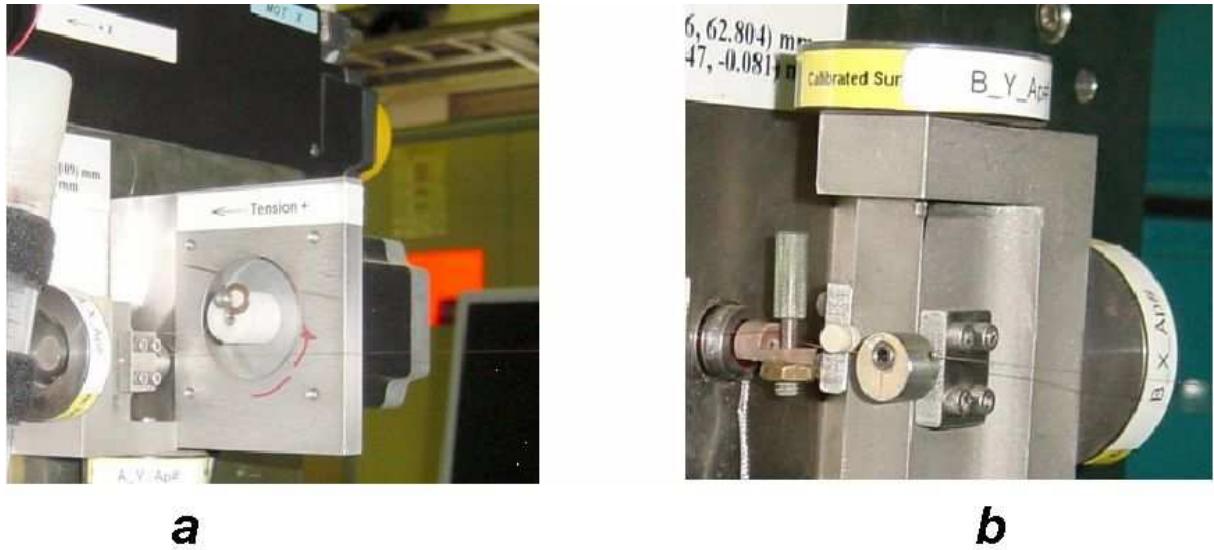


Fig. 11: (a) Stepping motor to adjust the wire tension in order to extrapolate the measurement to infinite tension. (b) Tension gauge used to drive the stepping motor at the expected tension.

at the curve of the integrated gradient measured as a function of the measured $1/f^2$, i.e., proportional to the wire tension according to the following relations:

$$\text{sag} = \frac{W \cdot g}{8T} L_w^2 \quad (18)$$

$$f = \frac{1}{2L_w} \sqrt{\frac{T}{W}} \quad (19)$$

with W the weight of the wire per unit length, g the gravity constant, and T the wire tension.

Figure 11(a) shows the remotely controlled actuator to change the wire tension. Figure 11(b) shows a tension gauge used to control the adjustment of the tension in particular to avoid breaking the wire.

6.2 Measure the gradient with horizontal displacements of the wire

Figure 12 shows the curves of the transfer function obtained at different currents into an LHC quadrupole magnet [9] for a given type of CuBe stretched wire. Three observations can be made about this measurement campaign summarized in Table 1:

- the slope of the measured gradient depends, as expected, on the frequency but also on the square of the gradient, i.e., the current in the quadrupole magnet;
- gradient values measured by vertical displacements were found to be systematically higher than with horizontal displacements;
- this dependence on the gradient is different according to the batch of CuBe wire used, and this finding correlates with the wire being paramagnetic or diamagnetic as sensed by a permanent magnet.

A detailed analysis proved that the $1/f^2$ dependence was linear for horizontal displacements but not for vertical displacements. This is explained by looking at the dependence of the magnetic forces in both cases. The force variation due to the field gradient is horizontal for an horizontal displacement:

$$F_x \propto G^2 \cdot (x_{\text{step}})^2, \quad (20)$$

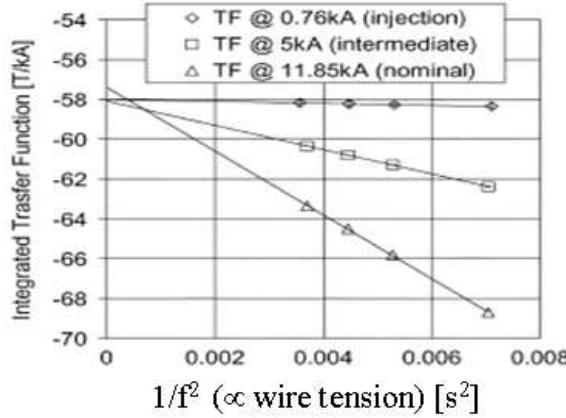


Fig. 12: Measured gradient, integrated over the quadrupole length, divided by the current as a function of the wire tension measured by the vibration frequency

Table 1: Slopes of measured quadrupole strength, as a function of inverse squared vibrating frequency, for different types of wire, in [T/s²].

Wire No.	0.76 kA	5 kA	11.5 kA	χ
1	30.4	2000	9500	paramagnetic
2	6.1	500	5000	paramagnetic
3	2.3	50	474	diamagnetic
4	—	—	380	diamagnetic

but has a parabolic dependence with vertical displacements

$$F_y \propto G^2 \cdot (y_{\text{step}} + \text{sag})^2. \quad (21)$$

The gradient obtained with vertical displacements should therefore be calculated with a fit to a parabolic law, so is less stable than for the horizontal displacements. It is therefore better to choose the result obtained with the horizontal displacements.

Relative discrepancies of $2 \cdot 10^{-4}$ for the measured gradient were obtained between measurement with vertical and horizontal displacements, giving an estimate of the accuracy that can be obtained with this method.

7 The vibrating wire technique

Mechanical oscillations of the stretched wire can be induced by the Lorentz force created by AC current flowing into the wire going through the static field of the measured magnet. This technique, proposed by A. Temnykh, has sub-micrometre sensitivity to sense a quadrupole axis. It can be used to measure separately the axis of several magnets aligned on a girder and has been extended to find sextupole axes.

Reference [10] details the resolution of the differential equation to solve for standing wave solutions:

$$W \frac{\partial x^2}{\partial t^2} = T \frac{\partial x^2}{\partial z^2} - \gamma \frac{\partial x}{\partial t} + I(t)B(z) \quad (22)$$

with $x(0, t) = x(L_w, t) = 0$ the boundary conditions, W the weight of the wire per unit length, T the wire tension, γ the damping coefficient, $I(t) = I_0 \cdot e^{i\omega t}$ the driving AC current, and $B(z)$ the transverse magnetic field.

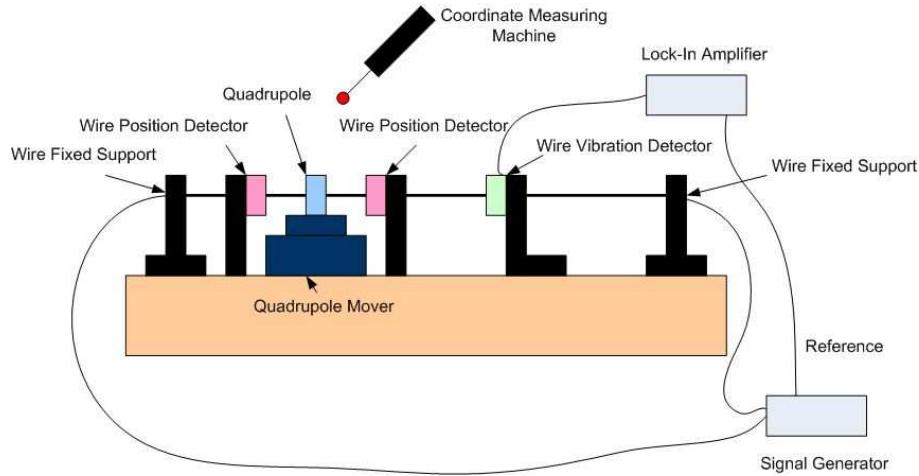


Fig. 13: Overview of the vibrating wire equipment measuring the magnetic axis in a quadrupole. The wire vibration detector and the quadrupole are longitudinally located to have maximum signal with the second harmonic of the natural oscillation frequency (courtesy of Z. Wolf, SLAC).

The solution is a sum of standing waves of amplitudes x_n that can be measured. From there the B_n coefficients of a sinus wave expansion of $B(z)$ are estimated

$$x(z, t) = \sum x_n \sin\left(\frac{\pi n z}{L_w}\right) e^{i\omega t}, \quad (23)$$

with

$$x_n = \frac{I_0}{W} \cdot \frac{1}{\omega^2 - \omega_n^2 - i\gamma\omega} \cdot B_n, \\ \omega_n = \frac{\pi n}{L_w} \sqrt{\frac{T}{W}}.$$

$B(z)$ can be reconstructed knowing the series of B_n :

$$B(z) = \sum B_n \sin\left(\frac{\pi n z}{L_w w}\right). \quad (24)$$

In practice, the AC current is tuned to the natural oscillation frequency f_{nat} of the wire or one of its harmonics $m \cdot f_{\text{nat}}$. The amplitudes of these oscillations are measured by inexpensive photo interrupters. The wire motion in the vertical, respectively horizontal, plane is caused by the Lorentz forces between the wire current and the horizontal, respectively vertical, magnetic field. Therefore two wire vibration detectors are mounted orthogonal to each other.

A procedure to centre one quadrupole is straightforward. The wire oscillation detectors can be located in a longitudinal position where the amplitude is large for the vibration harmonic considered. That is why it is separated from the position detectors in Fig. 13 that gives an overview of the equipment [11]. Since the Lorentz force is zero if the wire is aligned with the quadrupole axis, the magnet or the wire can be moved until vanishing oscillations are seen in both planes. Some coupling between the measured oscillations in both directions could exist due to imperfect orthogonality between the detectors. It has to be measured and taken into account to achieve high accuracy (Fig. 14). These measurements obtained by wire motion controlled by piezo-electric actuators demonstrate the sub-micron resolution [12].

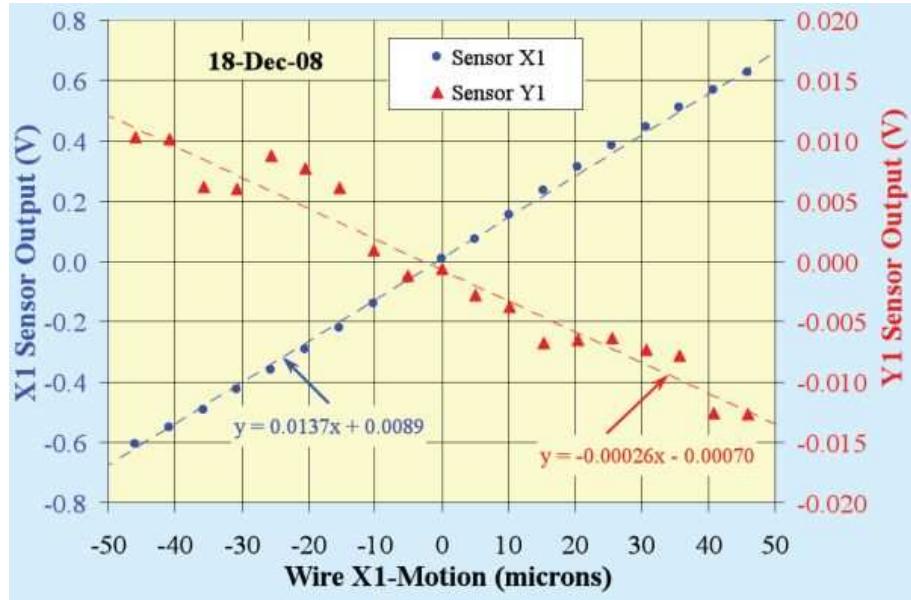


Fig. 14: Amplitudes measured by the x and y photocouplers for small displacements of the vibrating wire around the quadrupole axis in the x direction (courtesy of A. Jain, BNL)

8 Description of the harmonic coil method

The harmonic coil method was developed with early analog integrators forcing the measuring coil to rotate stepwise between consecutive angular positions [13]. Fast angular encoders and purpose-developed voltage integrators with zero dead time between the angular positions are the basis of today's systems acquiring several hundred points per turn with a rotation rate as high as 10 turns per second. It is the best method for measuring higher order multipoles within a well-established theoretical frame, in particular of superconducting and quadrupole magnets having circular apertures. Progress in data acquisition equipment and data analysis tools alleviates the complexity of the formalism applied to the amount of data to treat, so that fully automated instruments and data analysis processes have been developed for measurements of series of magnets with high confidence in the final results.

8.1 Flux enclosed by a simple rotating coil

A perfect dipole magnet gives a constant vertical field everywhere in the useful aperture. The flux enclosed by the simple coil described in Fig. 15 will be, considering an infinitely thin winding,

$$\Psi(\theta) = N_t \cdot L \cdot \int_{R_1}^{R_2} B_1 \cdot \cos \theta \cdot dR. \quad (25)$$

N_t and L are respectively the number of turns and length of the measuring coil. The coil is supposed to be shorter than the magnet. The coil's effective surface can be calibrated independently and is given by

$$\Sigma_{\text{coil}} = N_t \cdot L \cdot \int_{R_1}^{R_2} dR = N_t \cdot L \cdot (R_2 - R_1). \quad (26)$$

The use of a voltage integrator connected to the measuring coil makes it possible to eliminate the time coordinate in the induction law of Faraday. The voltage integrator read as a function of the angle gives the flux directly from the zero angle where it is reset. The constant of integration is irrelevant for this method.

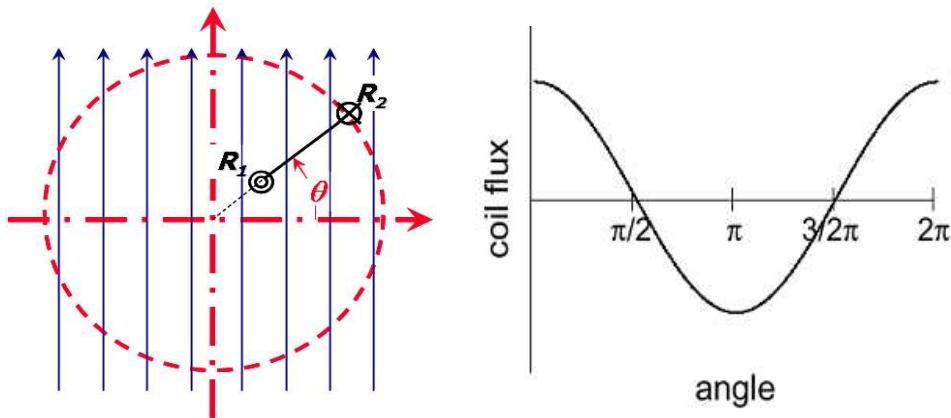


Fig. 15: 2D representation of the flux seen by a simple coil rotating in a dipole field

The units of Eq. (25) are

$$\Psi = \text{volt} \cdot \text{sec} = \text{tesla} \cdot \text{m}^2 = \text{weber}. \quad (27)$$

It is important to realize that the harmonic coil method does not make use of the voltage integrated over a given time, but rather over a given angular interval. The advantage of using a voltage integrator that can be externally triggered is that it eliminates to the first order the problem of a constant speed of rotation. A real system in fact measures differences of fluxes between two incremental angular positions. The angular encoder mounted on one coil end is a fundamental piece of equipment, as described in Section 9.2. The integrator is triggered by this encoder and collects incremental fluxes $\delta\Psi_k$, and the left part of Eq. (25) becomes

$$\Psi(\theta_i) - \Psi(\theta_0) = \sum_{k=1}^i \delta\Psi_k, \quad (28)$$

with

$$\delta\Psi_k = \Psi(\theta_k) - \Psi(\theta_{k-1}).$$

8.2 Formalism for multipoles measured by rotating coils

The power of the harmonic coil method is its ability to measure any type of 2D magnetic field. It can be demonstrated [14] that a rotating coil measures the 2D field integrated over its length as long as the field component parallel to the rotation axis is zero on the two coil ends. The complex equation to best describe this 2D field is

$$B(x + i \cdot y) = B_y(z) + i \cdot B_x(z) = \sum_{n=1}^{\infty} C_n \cdot \left(\frac{z}{R_r} \right)^{n-1}. \quad (29)$$

The components $C_n = B_n + iA_n$ are the normal and skew multipoles of the field. By definition for accelerator magnets, the normal components indicate a vertical field in the horizontal plane whilst the ‘skew’ terms apply for an horizontal field. The C_n are in tesla at the reference radius R_r . Figure 16 shows the field lines for normal and skew dipoles (C_1) and quadrupoles (C_2).

The field quality is usually described as errors relative to the main field component B_M ($M = 1$ for a dipole, $M = 2$ for a quadrupole) at the reference radius R_r . These errors are called ‘units’ and are given by

$$c_n = b_n + i \cdot a_n = 10^4 \frac{C_n}{B_M}. \quad (30)$$

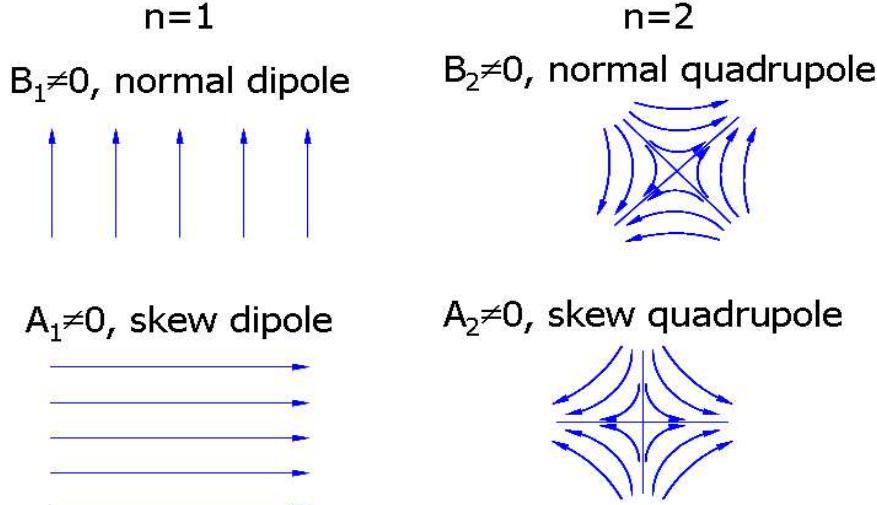


Fig. 16: Field lines of normal and skew dipole and quadrupole magnets

The reference radius R_r is an important concept for accelerator magnets having apertures much smaller than one metre. R_r corresponds in practice to

- the useful aperture for the beam,
- 2/3 of the yoke aperture in resistive magnets,
- 2/3 of the coil aperture in superconducting magnets,
- the radius where the multipoles relative to the main field, the c_n in Eq. (30), have the same order of magnitude in a real magnet.

It is important to carefully choose this reference radius at the beginning of a project. It will intervene in the discussions between all actors involved: beam optic physicists, magnet designers, measurement crew, and data analysis teams.

The voltage integrated over a simple rotating coil described in Fig. 15 and rotating in any 2D field will therefore be

$$\Psi(z) = N_t \cdot L \cdot \operatorname{Re} \int_{R_1}^{R_2} B(z) \cdot dz . \quad (31)$$

Since the coil rotates

$$z = x + i \cdot y = R \cdot e^{i\theta(t)} . \quad (32)$$

And by applying Eq. (29) and integrating it over dR

$$\Psi(\theta = \omega \cdot t) = \operatorname{Re} \left(\sum_{n=1}^{\infty} N_t \cdot L \cdot \frac{R_2^n - R_1^n}{n \cdot R_r^{n-1}} \cdot C_n \cdot e^{in\theta} \right) . \quad (33)$$

This allows a formal separation between what belongs to

- the measured field components C_n ,
- the time dependence of the signal $e^{in\theta(t)}$,
- the coil sensitivity factor K_n defined as

$$K_n = N_t \cdot L \cdot \frac{R_2^n - R_1^n}{n \cdot R_r^{n-1}} . \quad (34)$$

The K_n are calculated once for each measuring coil used. As will be detailed in Section 10 they can be complex numbers in the case of tangential coils, or coils not perfectly aligned radially. These calculations are substantial if the windings can no longer be considered point-like. Their values can be improved by individual calibrations [15].

The power of the harmonic method is that the multipoles of the field are directly given by the Fourier analysis coefficients Ψ_n of the integrated voltage over a coil turn $\Psi(\theta)$:

$$\Psi_n = K_n \cdot C_n = K_n \cdot (B_n + iA_n). \quad (35)$$

8.3 Centre the measurement results by measuring the field axis and direction

The harmonic method measures with high accuracy the direction of the main field component with respect to the measuring coil direction. It gives as well the axis of a quadrupole or sextupole magnet with respect to the rotation axis of the measuring coil. These measurements are fundamental for accelerator magnets since the beam closed orbit is defined by the residual tilt of the dipole magnets and the quadrupole misalignments. This property is also useful to align the set of measured multipoles in the symmetry axis of the magnet. This was required, for instance, for the system described in Fig. 17 where the 12 intermediate bearings of the 15 m long measuring shafts are poorly centred with respect to the superconducting magnet. Simple equations given below formally change the reference axis system of the multipole set.

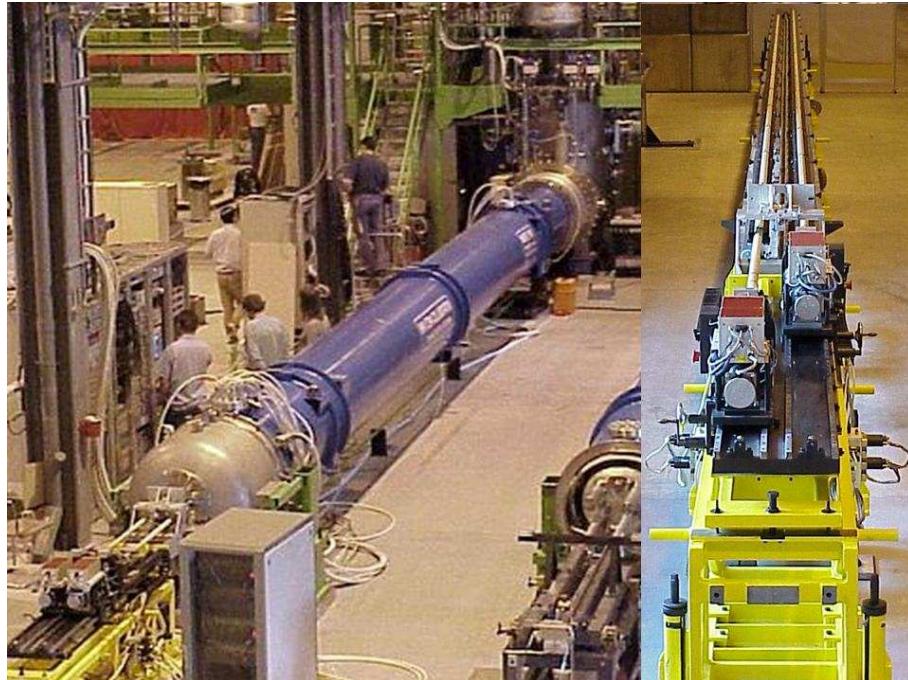


Fig. 17: LHC superconducting magnets measured with a 15 m long shaft. The ‘Twin Rotating Unit’ at the bottom of both pictures holds the motors and encoders.

8.3.1 Find the main field direction

The harmonic coil method gives the field description, the C_n in Eq. (35), in the reference frame defined by the measuring coil when at $\theta = 0$, usually given by the reference angle of the encoder. The main field direction corresponds to a zero main skew term ($A_1 = 0$ for a dipole magnet). This main field direction often does not correspond with enough accuracy to the magnet mechanical symmetry plane, hence the

usefulness of this measurement. In addition, both the magnet and measuring system may not be aligned with gravity, used as an external reference. Three frames are considered:

- θ_m : zero of the encoder, reference for the Fourier analysis;
- θ_g : gravity, magnet fiducials when aligned;
- θ_f : field direction of the magnet ($A_1 = 0$ for a dipole magnet).

To rotate the multipole coefficients from one frame to another is straightforward from Eq. (33). Assuming respectively θ_m and θ_f the angles of the measurement frame and the magnet symmetry plane with respect to gravity give:

$$z_f = z_m \cdot e^{i(\theta_m - \theta_f)}, \quad (36)$$

$$C_{m,n} = C_{f,n} \cdot e^{in(\theta_m - \theta_f)}. \quad (37)$$

The accuracy of finding the field direction in the coil reference frame is high: typically better than 0.1 mrad. The difficulty is to find the coil reference direction with respect to an external reference. The coil average winding position is commonly misaligned by a few mrad with respect to the coil frame. In addition, extreme care is needed to have a perfect alignment with respect to the encoder reference angle. An easy calibration is possible by turning end to end either the magnet to be measured or the encoder plus coil set. It was possible, on account of the size of the magnets to be measured, to design the bench shown in Fig. 18 for this easy calibration.

8.3.2 Coil rotation axis different from the magnet symmetry axis

Similarly, the measuring coil does not necessarily rotate about the magnet axis. The reference frame, where $z_c = 0$, of the magnet axis is usually defined by having both dipole components $B_1 = A_1 = 0$ in a quadrupole. With $d \cdot R_{\text{ref}}$ being the distance from the rotation axis to the measuring frame, the two sets of multipoles are related by

$$z_m = z_c - d \cdot R_{\text{ref}}, \quad (38)$$

$$C_{m,n} = \sum_{k=n}^{\infty} \frac{(k-1)!}{(n-1)!(k-n)!} C_{c,k} \cdot d^{k-n}. \quad (39)$$

This so-called ‘feed-down’ correction is used when the rotation axis cannot be accurately aligned in the magnet, for instance with a magnet aperture small compared to the length. These magnets are measured with ‘moles’ travelling along the aperture [16] or coil shafts with intermediate bearings resting in the magnet [17]. It must be stressed, however, that displacing the reference frame corresponds to describing the field outside the measurement circle, thus extrapolating the measurements outside their validity range. This feed-down correction loses validity for large values of d .

Equation (39) can be simplified, assuming that the coil rotation axis is near the magnet axis, i.e., $d \ll 1$:

$$C_{m,n} = C_{c,n} + n \cdot C_{c,n+1} \cdot d. \quad (40)$$

The position of the rotation axis of the measuring system with respect to the the axis of a quadrupole is therefore

$$(d_x + i \cdot d_y) \cdot R_{\text{ref}} = \frac{C_{m,1}}{C_{m,2}} \cdot R_{\text{ref}}. \quad (41)$$

Measuring this distance with an accuracy of 0.01 mm is easily achieved for short length measuring coils. As in Section 8.3.1 the issue is to refer the rotation axis to the magnet fiducials. An easy calibration is done when the magnet can be turned upside down.

8.4 Possible configurations according to the magnet size

Figure 18 shows a rotating coil bench measuring linac permanent quadrupoles that are short and small. The motor, the bearings of the rotating coil, and the angular encoder are accurately aligned in a straight line with reference to the base plate.



Fig. 18: Bench measuring a small permanent quadrupole magnet for the drift tube of a linac. The motor is on the left and the encoder on the right.

Figure 17 shows in comparison a 15 m long, 35 tonne LHC superconducting dipole magnet. Reference [18] describes the mathematical model established on the series measurement of the LHC superconducting magnets, giving an idea of the size of the measurement project. The measuring shaft [17] in the right part is composed of 12 segments 1.15 m long and separated by ceramic bearings resting on the anti-cryostat, therefore rotating off-axis by up to 3 mm with respect to the symmetry axis of the magnet. By construction, the measuring segments are not accurately aligned azimuthally with respect to each other. Equations (37)–(39) have therefore to be applied to express the results of the measurements in the reference frame of the magnet axis.

Similarly, these LHC magnets were measured by ‘moles’ [16] in the manufacturing industries. These moles group in one device the coil, motor, and encoder and travel along the magnet length pulled by a cable. Measuring the field direction with the help of an on board inclinometer requires dedicated calibration. Note that other techniques (Section 5, [19]) can measure the magnetic axis of quadrupole magnets.

Figure 19(a) shows a coil structure designed for easy alignment of the magnet with respect to the bearings of the measurement system [20]. This coil setup allows one to separately align the magnetic axis on both magnet ends, then to fine tune the alignment with the full length coil. The full length coil gives the magnet strength, or integrated field, value directly relevant for the accelerator beam. The magnet effective length is obtained by calculating the central field given by the difference between the full length coil and the sum of the end coils. The measurement from the end coils can be compared to 3D simulation of the magnet ends. Note that a measurement with a coil shorter than the magnet is valid only if the longitudinal component of the field is zero on both ends of these coils [14].

Small-aperture magnets render difficult the manufacture of rotating shafts with the coils on both sides of the axis. The solution of Fig. 19(b) is preferred even though the sum of three (or more) data sets must be done. The effective length of the blind parts between the segments must be accurately measured and located where the field does not vary along the axis.

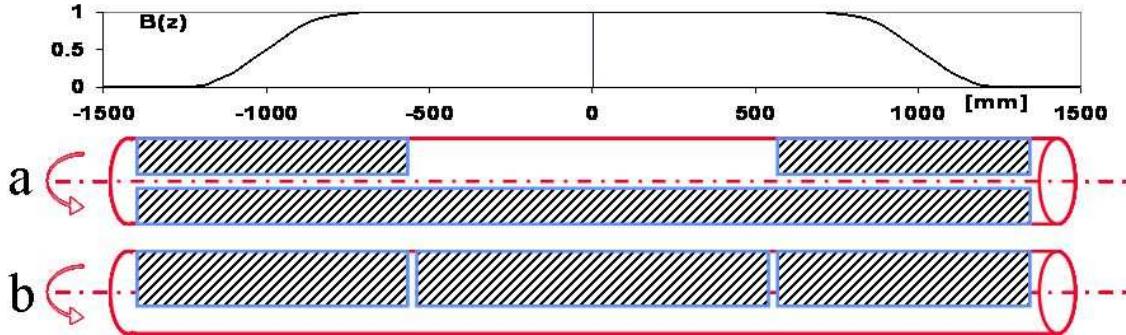


Fig. 19: (a) Coil shaft with a full length coil and two end coils. (b) Separated coil structure when solution (a) cannot be manufactured. It implies blind spaces between the coils.

9 Accuracy limitations of the harmonic coil method

Mechanical or electronic imperfections mainly degrade the measurement of the ‘higher order’ multipoles, i.e., those with harmonic numbers higher than the magnet multipole order. The three main error sources will be studied in detail:

- voltage integrator offset coupled with irregular rotation rate of the coil,
- error in the coil angle measurement due either to the angular encoder or to torsions of the coil shaft during rotation,
- instability or movement of the rotation axis of the coil shaft due to gravity, bearings quality, or vibrations.

Schemes of compensation coil arrays, connected in opposition, have been developed [21] to remove the signal coming from the magnet main multipole thus allowing the increase of the amplification factor at the input of the integrator. More importantly, these compensation coil assemblies remove non-linear coupling coming from the main harmonic and degrading the high-order harmonic measurement.

9.1 Offset of the integrator coupled with varying rotation rate

Stability of the rotation rate of the harmonic coil does not appear at first order in the method. The angular fluxes, given in Eqs. (8.1)–(33) are explicitly independent of time. However, the voltage integrator has an offset that gives a flux error inversely proportional to the rotation rate. The average offset over a turn can be eliminated afterward by one of the following methods in the case where the field is static during the coil rotation. Obviously, a careful adjustment to zero the voltage offset before the measurement is the most reliable method.

- The coil shaft can continuously rotate if the signal goes via slip rings from the coil to the integrator. The flux integrated over a full turn must be zero if the field is static. This formula is also valid if the current variation is constant and the angle zero corresponds to a zero flux in the coil, but this correction is sensitive to noise in the signal and could be detrimental. The incremental offset can be eliminated at first order by dividing the offset integrated over one turn by the number of angular intervals:

$$\oint \text{Offset} = [\Psi(2\pi) - \Psi(0)_{\text{measured}}] . \quad (42)$$

- Without the use of slip rings, the rotating coil has to come back in reverse rotation due to the instrumentation cable. Averaging the incremental fluxes measured during the forward and backward turns removes at first order the offset. Note that it implies as well either a constant or constantly changing current.

Variation of the rotation rate can further be eliminated if the time duration of the individual angular steps are accumulated together with the incremental fluxes and the corresponding correction is applied before the Fourier analysis takes place. Experience has shown that the resulting signal-to-noise enhancement is limited by imperfect removal of this effect, even for rotation rates measured constant within a few per cent.

We will see in Sections 9.3 and 9.5 that most of these degradations of the high order multipole measurement due to electronics can be reduced by the use of compensation coil schemes.

9.2 Measurement errors related to the angle and encoder

The accuracy in measuring the harmonic coefficients of the field depends directly on the angular precision of the integrator triggers. The encoder quality, the torsion of the frame holding the coil or linking it to the encoder must all be considered. Reference [14] gives the formalism to calculate these effects taking into account all kinds of torsions and vibrations. The case of a periodic error when measuring a pure dipole magnet will be detailed here. General equations can be found in Ref. [22].

An angular misalignment between the axis of the encoder and the axis of the coil can lead to a difference between the measured and real angular position of the coil. This error is best approximated by a sine function. A torque on the bearing of the encoder due to a parallel misalignment between these axes gives a similar error of amplitude defined as ϵ :

$$\theta_{\text{meas}} = \theta + \epsilon \cdot \sin \theta . \quad (43)$$

A perfect dipole aligned with respect to the measuring equipment is defined by $B_1 = 1$, all other coefficients $B_n, A_n = 0$. As the angular encoder errors are supposed to be small, a first-order estimation gives

$$\Psi(\theta) \propto \cos(\theta_{\text{meas}}) = \cos(\theta + \epsilon \cdot \sin \theta) = \cos(\theta) - \frac{\epsilon}{2} \cdot (1 - \cos 2\theta) . \quad (44)$$

An erroneous quadrupole term, not present in the magnet, will show itself in the Fourier analysis of the integrated voltage. Figure 20 represents it. Its value is related to the amplitude of the error of the encoder by

$$\delta B_2/B_1 = \delta b_2 = \frac{\epsilon}{2} \cdot \frac{K_1}{K_2} . \quad (45)$$

In order to give orders of magnitude, Eq. (45) can be further simplified with the following hypothesis for the radii sketched in Fig. 15:

$$R_1 = 0 , \quad R_2 = R_{\text{ref}} .$$

The erroneous quadrupole $\delta b_2 = 10^{-3}$, i.e., 10 units for an angular error $\epsilon = 1$ mrad since the ratio of the sensitivity factors is

$$\frac{K_n}{K_1} = \frac{1}{n} \cdot \left(\frac{R_2}{R_{\text{ref}}} \right)^{n-1} = \frac{1}{n} . \quad (46)$$

More generally, any systematic angular error can be decomposed in a Fourier series:

$$\theta_{\text{meas}} = \theta + \sum_k \gamma_k \cos k\theta + \epsilon_k \sin k\theta . \quad (47)$$

These γ_k and ϵ_k will generate erroneous normal and skew multipoles when measuring a pure dipole magnet:

$$\begin{aligned} \delta b_n &= \frac{n K_n}{2 K_1} (\epsilon_{n-1} + \epsilon_{n+1}) , \\ \delta a_n &= \frac{n K_n}{2 K_1} (\gamma_{n-1} + \gamma_{n+1}) . \end{aligned} \quad (48)$$

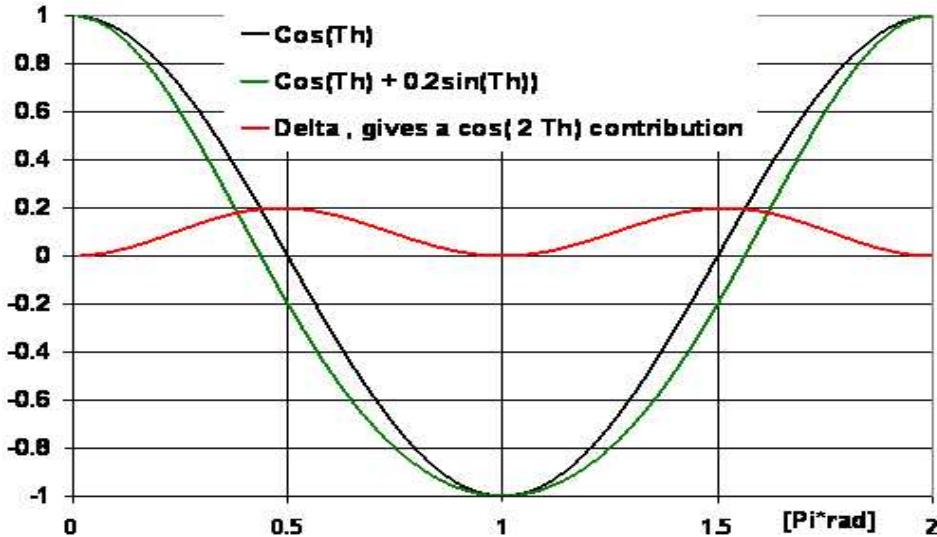


Fig. 20: Flux measured in a dipole with a perfect angle measurement and a first-order error. Their difference gives an erroneous quadrupole term.

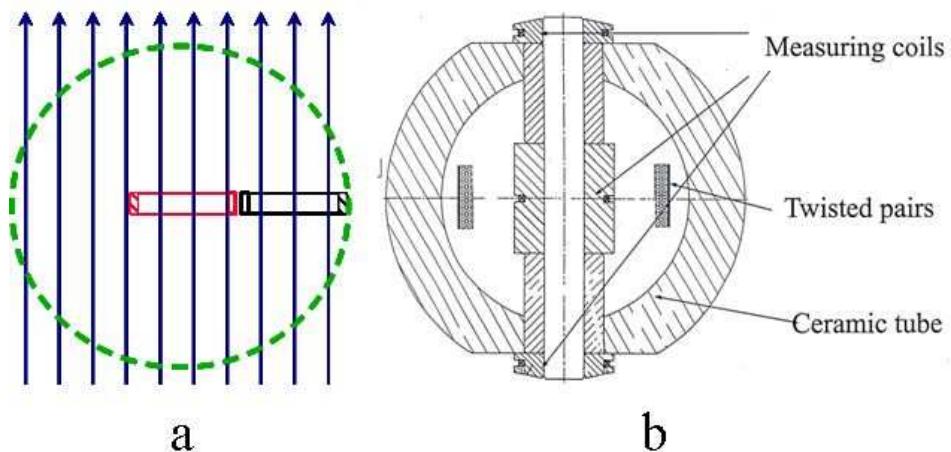


Fig. 21: (a) Assembly of two radial coils with same effective surface connected in opposition to remove the main dipole component from the signal. (b) Cross-section of the three tangential coils as used in Ref. [18]. One of the external coils is kept as a spare.

9.3 Removal of the main harmonic in a dipole magnet by using compensation coil scheme

The previous section shows that high precision in the measurement of the multipoles requires high quality encoders. In addition, the geometry of the equipment to be built, namely the ratio between magnet aperture and length, can lead to severe difficulties in achieving a torsion-free frame holding the coils.

A suitable geometry for the measuring-coil system may be chosen that reduces the angular precision needed. Two coils having the same sensitivity with respect to the dipole term (the same surface) are electrically connected in opposition at the input of the integrator. The resulting sensitivity factor as calculated in Eq. (34) becomes $K_1 = 0$. Their radial positions are sketched in Fig. 21. Consequently, the dipole harmonic is rejected in the flux measured. Rejection ratios of 300 to 2000 can be achieved by sorting the coils according to their effective area and by adjusting them to be parallel. This assembly of coils is sometimes called ‘bucking coils’ in the literature.

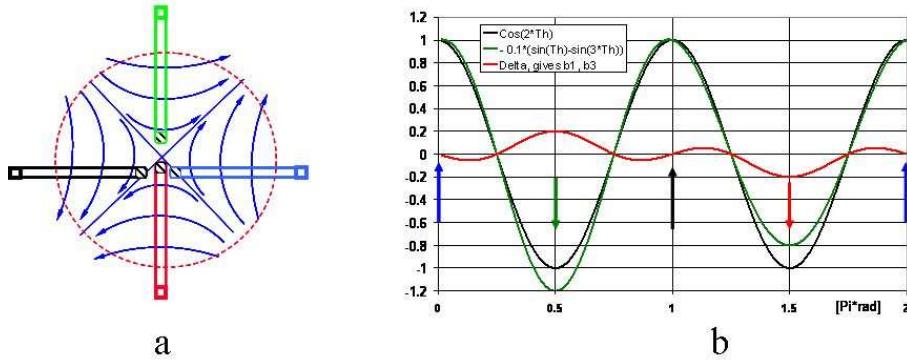


Fig. 22: (a) A non-rigid coil shaft moving up and down, due to gravity, during rotation in a quadrupole field. (b) The flux measured with a perfect rotation and a first-order error. Their difference gives erroneous dipole and sextupole terms.

9.4 Measurement errors related to coil transverse movement

In a dipole magnet, the measuring coil assemblies of Fig. 21, displaced laterally or vertically, will induce no voltage on the integrator connected to it. This is not true if higher harmonics are present or when a quadrupole or sextupole magnet has to be measured. A spurious signal therefore shows up if the mechanics is such that the harmonic coil does not describe a perfect circle when rotating. Calculations show that without proper rejection of the main harmonic, one cannot measure with high accuracy the field quality of a quadrupole.

The deflection of the frame holding the coil gives a simple measure of this effect. The hypothesis is made that the coil is wound on a flat plate that bends due to its own weight when horizontal, and is straight when vertical [Fig. 22(a)]. The centre of the coil consequently moves up and down twice per turn. This vertical displacement of amplitude $d \cdot R_{\text{ref}}$ is approximated by

$$\text{Displ.} = i \cdot d \cdot R_{\text{ref}} \cdot \cos(2\theta). \quad (49)$$

The bounds of the integral of Eq. (31) become with the hypothesis of a simple coil [given by Eq. (46)]

$$\begin{aligned} R_2(\theta) &= R_{\text{ref}} \cdot (i \cdot d \cdot \cos(2\theta) + e^{i\theta}), \\ R_1(\theta) &= R_{\text{ref}} \cdot i \cdot d \cdot \cos(2\theta). \end{aligned} \quad (50)$$

These lateral movements during coil rotation induce non-linear coupling with the quadrupolar field and generate erroneous multipoles of lower and higher orders. Figure 22(b) sketches the curves giving the angular flux with and without this vertical movement. A first-order calculation, i.e., neglecting terms in d^2 , into a pure quadrupolar field (only $B_2 \neq 0$) gives the following erroneous skew components:

$$\delta B_1/B_2 = \delta b_1 = d \quad (51)$$

$$\delta b_3 = -3d. \quad (52)$$

A deflection due to gravity of 0.02 mm with $R_{\text{ref}} = 20$ mm (i.e., $d = 0.001$) generates an erroneous dipole corresponding to 0.02 mm of axis displacement and a relative sextupole $b_3 = 0.003$ (or 30 units). This simple calculation therefore stresses the importance of designing a measuring shaft to hold the rotating coils as rigidly as possible, and having as smooth as possible rotation with the help of high-quality bearings and motorisation.

The formal development of the erroneous components created by any transverse movement in any real magnet can be found in Refs. [14]–[22].

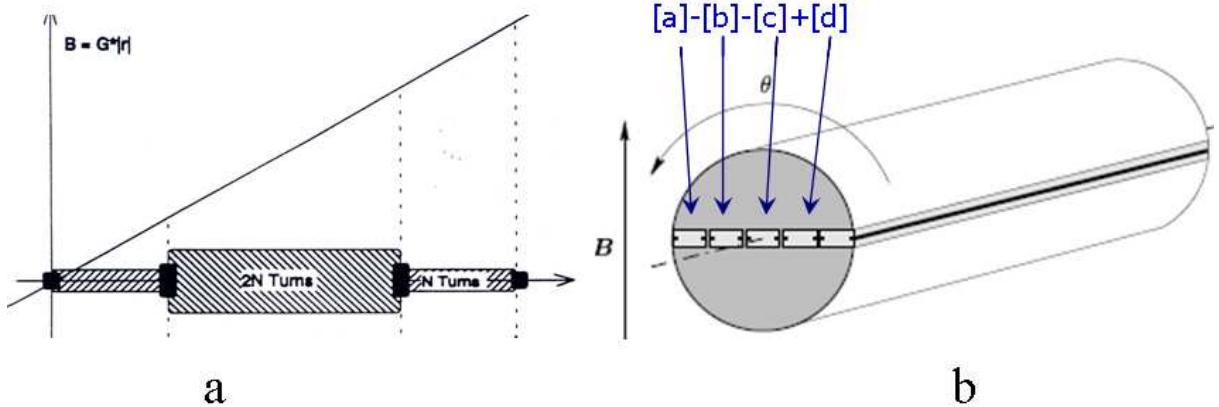


Fig. 23: Two geometries to compensate the dipole and quadrupole components: (a) two coils having the same effective area, (b) four equal radial coils give the same compensation, one of the external coils is kept as a spare.

9.5 Removal of the main harmonic in a quadrupole magnet by using compensation coil scheme

A coil scheme for an effective removal of the main quadrupole harmonic has to be insensitive to both angular and transverse movements in a pure quadrupole, i.e., to both the quadrupole and dipole sensitivity factors as defined in Eq. (34):

$$K_1 = K_2 = 0 . \quad (53)$$

The two schemes of Fig. 23 fulfil these two conditions.

It is difficult to have exactly the same effective area for the two coils of the scheme of Fig. 23(a), i.e., the inside coil having half the width and twice the number of turns of the outside one. This scheme induces in addition a high loss of sensitivity to measure the low order multipoles, in particular the sextupole. Although such a coil assembly rejects correctly the quadrupole harmonic, a factor of more than five is lost in sensitivity to measure the sextupole term compared to the sensitivity given by the single external coil. Since this coil assembly is insensitive to any lateral displacement in a quadrupole, turning about another axis will maintain the rejection. A maximum value for the sextupole sensitivity factor K_3 is obtained with the radii of the external coil being [23]:

$$\begin{aligned} R_2 &= R_{\text{ref}}, \\ R_1 &= -1/2 \cdot R_{\text{ref}} . \end{aligned} \quad (54)$$

The other scheme is based on four equal coils giving the same geometrical properties [Fig. 23(b) and Fig. 24]. A rejection ratio of more than 100 can be achieved by sorting and careful positioning of the four coil assemblies. Note that these two types of coil assemblies can be used to measure dipole magnets according to Section 9.3 by using only the central and external coils.

10 Calculation of the sensitivity factors of the harmonic coils

10.1 Tangential coil sensitivity

Tangential measuring coils are preferred for easier manufacture and better stability of the rotating coil frame (Section 9.5).

In order to calculate the sensitivity of tangential coils, Eqs. (31)–(34) have to be generalized by putting R_1 and R_2 as 2D variables in the complex plane:

$$K_n = N_t \cdot L \cdot \frac{Z_2^n - Z_1^n}{n \cdot R_r^{n-1}} . \quad (55)$$

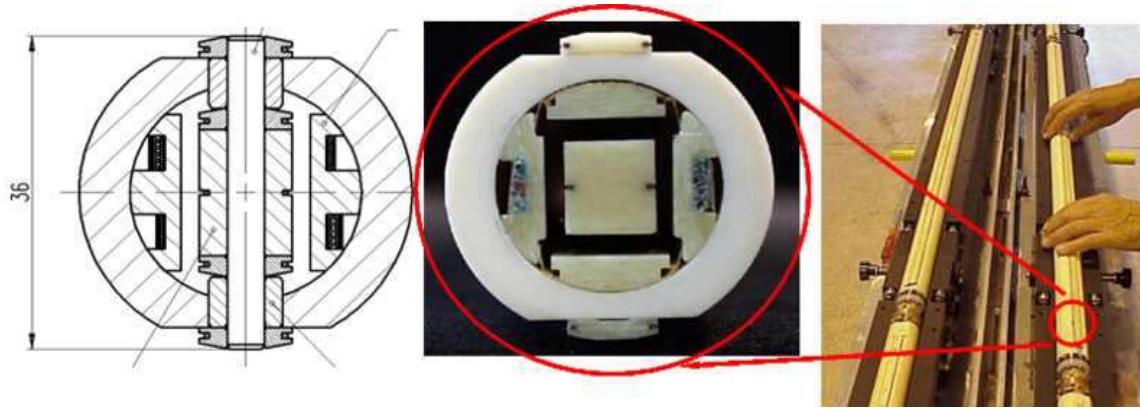


Fig. 24: A compensation coil scheme to measure quadrupoles. Four tangential coils are used to remove the dipole and quadrupole components, one of the external coils is kept as a spare.

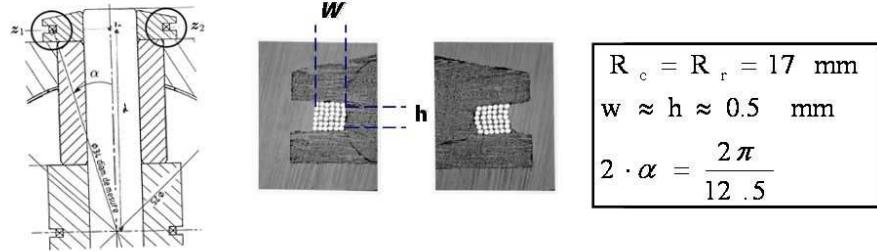


Fig. 25: Parameters of the tangential coil cross-section used for the series measurements of the LHC magnets

The tangential coil of Fig. 25 is horizontal on the y_+ for the first point of rotation and gives real expressions for the K_{2n+1} . The normal components of the multipoles belonging to the dipole symmetry have indeed a vertical field on the y axis. The normal components of the quadrupole, octupole, etc. have an horizontal field on the y axis and the K_{2n} factors have imaginary values. Equation (55) is simplified to

$$K_n = -2 \cdot i^{n+1} \cdot N_t \cdot L \cdot \frac{R_c^n \sin(n\alpha)}{n \cdot R_r^{n-1}}, \quad (56)$$

with the two winding positions

$$\begin{aligned} Z_1 &= R_c \cdot e^{i(\pi/2+\alpha)}, \\ Z_2 &= R_c \cdot e^{i(\pi/2-\alpha)}. \end{aligned}$$

Equation (56) shows that tangential coils are insensitive to multipole order having an angular period, or an integer number of periods, corresponding to the opening angle. This drawback is easily overcome by having the opening angle either small enough or corresponding to

$$2\alpha = \frac{2\pi}{n + 1/2}. \quad (57)$$

Note that the K_n would all be imaginary if the tangential coil starts to rotate in vertical position centred on the x_+ axis since the field of all normal components (the B_n) is vertical on the x axis. Equation (55) gives with this hypothesis

$$K_n = -2 \cdot i \cdot N_t \cdot L \cdot \frac{R_c^n \sin(n\alpha)}{n \cdot R_r^{n-1}}. \quad (58)$$

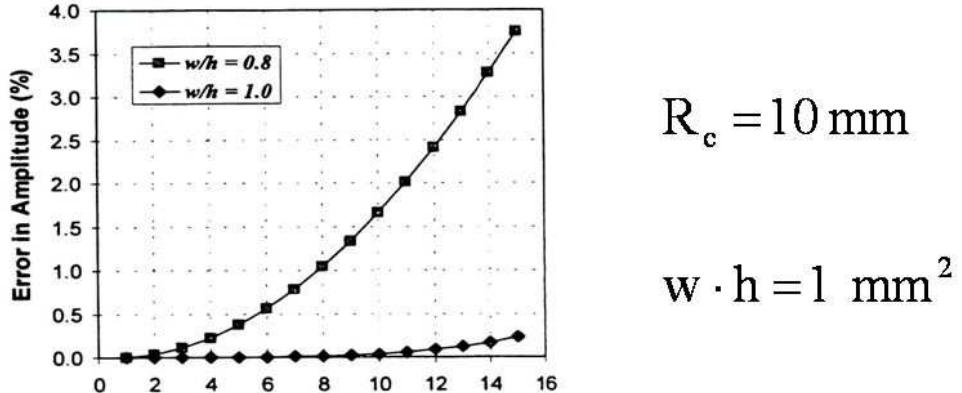


Fig. 26: Error on the sensitivity factor K_n as a function of the multipole order if the winding is considered point-like for a square and rectangular cross-section (courtesy of A. Jain)

Equation (55) is valid for any coil and should be used in the case of compensation coil schemes (Sections 9.3–9.5) where the relative angle between the different coils should be taken into account. General expressions for any ill-positioned coil can be found in Ref. [24].

10.2 The coil winding is not point-like

Equations (34) and (55) are valid for coils having point-like windings. The windings are usually multi-turn and their dimensions, width and height in Fig. 25, should be taken into account by summing the contribution of the position of each individual turn. This summation can be simplified by a normal integral over the section of the winding with the assumption that the density of the turns is homogeneous over this section. The point-like radius of Eq. (55) has to be replaced by the n -order average over the winding section S [24]:

$$\langle Z^n \rangle = \frac{\int_S z^n \cdot dz}{S}. \quad (59)$$

The importance of the corresponding correction is, however, small. Their difference with the point-like case grows with the multipole order whilst the multipole perturbations they measure generally decrease with the order. In addition, the corrections to apply become negligible for square-shaped windings even for high harmonic numbers. This is demonstrated by Fig. 26 extracted from Ref. [22]. A coil winding of 1 mm^2 located at 10 mm radius and calculated with the point-like approximation would give an error of two per mille for the 30-pole ($n = 15$). These effects should therefore be estimated only for measuring coils with aspect ratios far from a square shape like coils based on printed circuits.

11 Experience with the harmonic coil method

11.1 Precision of the measurement of the main field component

The measurement of the field strength integrated over the magnet length, or of the transfer function defined as this integral divided by the current, is usually insensitive to electronic or mechanical noise sources like the shaft rigidity, smoothness of the rotation, vibrations. However, it depends strongly on the calibration of the coil surface for the dipole measurement and the positioning of the coils in the coil frame for quadrupoles and higher order magnets. Techniques for calibrating the coils are described elsewhere in these proceedings. They can be based on measurements with the single stretched wire technique described in Section 4. Accurate measurements of the gains of the electronics and of the excitation current are obviously key elements. Relative precisions of 10^{-4} are obtained to measure dipole magnet strengths and 3 to 10 times worse for quadrupoles with careful calibrations of all elements [9].

11.2 Measuring the higher-order multipoles

The current stability must be extremely high over the turn duration, in the range of ppm to tens of ppm, in order not to create spurious high-order multipoles if measured with a single coil. It is possible to acquire the current value for each incremental flux and apply a first-order correction but this requires a high bandwidth and high-quality measurement.

Experience indicates that the low-order multipoles are more sensitive to the equipment imperfections analysed: voltage integrator offset, rigidity of the frame, irregularities in the rotation giving both torsional and lateral movements. In other words, the sextupole is the difficult multipole to measure in a quadrupole, and that justifies a careful design of the compensation coil scheme to increase K_3 as explained in Section 9.5. Repeatability of 0.01 unit (or 1 ppm relative to the main component) has been achieved in the best cases [25].

The use of compensation schemes improves the measurement of higher-order multipoles for the following reasons.

- The measured signal has a smaller amplitude and can therefore be amplified, reducing electronic noise. It also reduces the errors coming from the offset and the non-linear coupling between offset, rotation rate variations, and signal of the main order harmonic.
- Excitation current variations, including the ripple from the magnet power supply, also couple with the main order signal which is removed with compensation coils. For the same reason, measurements with slowly ramping currents are improved.
- An improvement proportional to the rejection ratio is obtained for all noises generated by mechanical instabilities: displacement of the rotation axis of the coil shaft due to gravity, quality of the bearings, or vibrations.

11.3 Measurement of pulsed magnets with harmonic coils

Measurement of pulsed magnets, i.e., excitation current changing with time, is hardly covered by the rotating coil method. The measurement of the magnet strength, i.e., the main multipole, is spoiled by the field change over the duration of the coil revolution. Accurate synchronization for the measurement of the excitation current at each angular position could be mandatory to get enough precision. Obviously, increasing the rotation rate of the coil shaft and the bandwidth of the electronics greatly helps and must be taken into account at the design phase of the rotating coils, i.e., the highest number of turns for the coil windings is not necessarily the optimum. As an example, the 15 m long coil shaft of Fig. 17 currently measures, at 8 turns per second, the field quality of the LHC magnets with high temporal accuracy.

The measurement of the field quality, i.e., the higher-order multipoles, is greatly improved [26] in this case by the help of a coil scheme to compensate the main multipole (Sections 9.3 and 9.5).

For fast ramping magnets, it is possible to measure flux variation between zero and nominal current with the harmonic coil fixed at a sufficiently large number of angular positions to perform the Fourier analysis. Experimental work is going on at GSI [27] for the FAIR project and at CERN for the new Linac quadrupole magnets having a duty cycle of a few milliseconds. This method requires an accurate angle positioning, obtained for instance by a high-resolution stepping motor. It would bring the advantages of the Morgan coil [3] technique with a simplified coil frame to be manufactured.

12 Conclusion: compare the different measurement techniques

Coils and stretched wires are tools particularly adapted to measurement of accelerator magnets. They give reduced values for parameters controlling particle beams: cross-section components of the field integrated along the magnet length.

The SSW based system can be considered reference equipment to measure in a static field dipole and quadrupole strength, i.e., main field components, where precisions of 10^{-3} to 10^{-4} are needed.

These instruments measure with high accuracy the main field direction and magnetic axis of quadrupole and sextupole magnets. They are being used more and more for these measurement goals in particular with the requirement to better align magnets for synchrotron light sources and with the need for smaller and smaller apertures for high-energy physics accelerators.

Flip coils are a reduced technique of the harmonic rotating coil method. The theory associated with their use is, however, mandatory to correctly measure fast pulsed and curved magnets.

The rotating coil method is a general and accurate method to measure the field quality of magnets: integrated field value, higher order multipoles, and magnetic axis. Recent instrumentation and acquisition systems allow high bandwidth and fully automated measurements. This method is the obvious choice for normal quadrupole magnets, and for superconducting magnets having circular apertures and where beam optics considerations require unprecedented precision in the field quality.

These various methods complement each other. They are complemented by the use of Hall plates for local measurements and of NMR based instruments for high absolute accuracy and calibration. A cross-check between these various methods should be used whenever possible to ascertain precision in magnet measurements.

Acknowledgements

The sections related to the single stretched wire are mostly extracted from the experimental and theoretical work done in different laboratories, in particular by J. DiMarco, A. Jain, A. Temnykh, and Z. Wolf. The list of names of my numerous CERN colleagues who are at the basis of my competence in the field of magnetic measurements is too long to mention here.

References

- [1] W. F. Brown and J. H. Suer, Rev. Sci. Instrum. **16** (1992) 276.
- [2] M. I. Green, Search coils, CAS – Measurement and Alignment of Accelerator and Detector Magnets, Anacapri, Italy, 1997, CERN 98-05, p. 143.
- [3] G. H. Morgan, Proc. 4th Int. Conf. on Magnet Technology, Upton, NY, 1972, USAEC CONF-720908 (Atomic Energy Commission, Washington, DC, 1973?), p. 787.
- [4] A. Jain et al., Field quality measurements at high ramp rates in a prototype dipole for the Fair project, IEEE Trans. Appl. Supercond., vol 18-2, p. 1629, June 2008.
- [5] S. Rossi, Developments in proton and light-ion therapy, Proc. EPAC 2006, Edinburgh, 2006 (Jacow, Geneva, 2006), p. 3631.
<http://accelconf.web.cern.ch/AccelConf/e06/PAPERS/FRYAPA01.PDF>.
- [6] J. DiMarco et al., Field alignment of quadrupole magnets for the LHC interaction regions, IEEE Trans. Appl. Supercond., vol. 10-1, p. 127, March 2000.
- [7] G. Rakowsky et al., Measurement and optimisation of the Visa undulator, Proc. PAC 1999, New York, 1999, (IEEE, Piscataway, 1999),
<http://accelconf.web.cern.ch/AccelConf/p99/PAPERS/THA35.PDF>.
- [8] T. C. Fan, C. S. Hwang, and C. H. Chang, A systematic study on the pulsed wire system for magnetic field measurements on the long undulator with high field, Rev. Sci. Instrum. **73** (2002) 1430.
- [9] N. Smirnov et al., Focusing strength measurements of the main quadrupoles for the LHC, IEEE Trans. Appl. Supercond., vol. 16-2, p. 261, June 2006.
- [10] A. Temnykh, Vibrating wire field-measuring technique, Nucl. Instrum. Methods Phys. Res., A **399** (1999) 185.

- [11] Z. Wolf, Vibrating wire technique for quadrupole fiducialization, Proc. of 14th International Magnet Measurement Workshop, Geneva, September 2005,
<http://indico.cern.ch/contributionListDisplay.py?confId=0517>
- [12] A. Jain, Results from vibrating wire R.D. for alignment of multipoles in NSLS-II, Proc. of 16th International Magnet Measurement Workshop, Bad Zurzach, October 2009,
<http://immmw16.web.psi.ch/Presentations>
- [13] J. Cobb and D. Horelick, Proc. 3rd Int. Conf. on Magnet Technology, Hambourg, 1970 (DESY, Hambourg, 1972), p. 1439.
- [14] W. G. Davies, Nucl. Instrum. Methods, **A311** (1992) 399.
- [15] M. Buzio, Fabrication and calibration of search coils, these proceedings.
- [16] J. García Pérez et al., Performance of the room temperature systems for magnetic field measurements of the LHC superconducting magnets, IEEE Trans. Appl. Supercond., vol. 16-2, p. 269, 2006.
- [17] J. Billan et al., Twin rotating coils for cold magnetic measurements of 15 m long LHC dipoles, IEEE Trans. Appl. Supercond., vol. 10-1, p. 1422, March 2000.
- [18] S. Sanfilippo et al., Magnetic performance of the main superconducting magnets for the LHC, IEEE Trans. Appl. Supercond., vol. 18-2, p. 132, 2008.
- [19] M. Buzio et al., A device to measure magnetic and mechanical axis of superconducting magnets for the Large Hadron Collider, IEEE Instrumentation and Measurement Technology Conf., Sorrento, Italy, 2006 (IEEE, New York, 2006), p. 759.
- [20] O. Pagano, P. Rohmig, L. Walckiers, and C. Wyss, Proc. 8th Int. Conf. on Magnet Technology, Journal de Physique, vol. 45, Colloque C1 (1984), p. 949.
- [21] R. M. Main, J. T. Tanabe, and K. Halbach, Measurements and correction of the PEP interaction region quadrupole magnets, IEEE Trans. Nucl. Sci., vol. NS26-3, p. 4030, June 1979.
- [22] A. Jain, Harmonic coils, CAS – Measurement and Alignment of Accelerator and Detector Magnets, Anacapri, Italy, 1997, CERN 98-05, p. 175.
- [23] L. Walckiers, The harmonic coil method, CAS – Magnetic Measurement and Alignment, Montreux, Switzerland, 1992, CERN 92-05, p. 138.
- [24] L. Deniau, Coils calibration, correction factors for rectangular windings,
<https://edms.cern.ch/file/356512/1/mta-in-98-026.pdf>
- [25] L. Walckiers et al., Sensitivity and accuracy of the systems for the magnetic measurements of the LHC magnets at CERN, Proc. EPAC 2000, Vienna, Austria (EPS, Geneva, 2000),
<http://accelconf.web.cern.ch/AccelConf/e00/PAPERS/THP2A17.pdf>
- [26] J. Buckley, D. Richter, L. Walckiers, and R. Wolf, Dynamic magnetic measurements of superconducting magnets for the LHC, IEEE Trans. Appl. Supercond., vol. 5-2, p. 1024, 1995.
- [27] P. Schnizer et al., A mole for measuring pulsed superconducting magnets, IEEE Trans. Appl. Supercond., vol 18-2, p. 1648, June 2008.

Fabrication and calibration of search coils

M. Buzio
CERN, Geneva, Switzerland

Abstract

In this paper the techniques available to make and calibrate magnetic search coils are reviewed, with emphasis on harmonic coil systems as the commonly-used optimal choice for integral measurements of accelerator magnets in terms of measuring range, accuracy, and cost. The topics treated, drawing extensively on half a century of experience at CERN, include mechanical and electrical design criteria, practical fabrication techniques, metrological considerations, and various calibration methods for coil parameters such as surface area, rotation radius, tilt angle etc. in both static or time-varying magnetic fields.

1 Introduction

1.1 Scope

This lecture is concerned with the design, construction, and calibration of search coils for the measurement of accelerator magnets. This well-established type of sensor represents in many practical cases the best choice in terms of range of application, accuracy and cost-effectiveness, especially where the field map over large volumes is sought. Sensors typically used for point-like measurements, such as Hall-effect plates or NMR probes are described in other lectures of this school (see Refs. [1] and [2] respectively).

The material presented here is drawn mostly from the author's experience within CERN's magnetic measurement team, which has been accumulating, over more than five decades, the instruments and know-how needed to characterize an immense variety of magnets for the many accelerators of the complex. By sharing this experience, this lecture aims at outlining the most important techniques that have to be mastered by anyone who wants to build or simply use correctly a search coil, with special emphasis on the practical aspects of the work. As the variety of requirements and parameters is wide, coils of the type described here can hardly be found off-the-shelf, and it is hoped that the topics discussed will help interested readers to choose and implement the technique that works best for them. More information can be found in earlier CERN Accelerator Schools almost exclusively devoted to the subject of magnetic measurements [3, 4].

1.2 Working principle

Search coil is the generic name of a widespread class of sensors consisting of one or more loops of conducting wire, exposed to a magnetic field \mathbf{B} and generating an output voltage V_C via Faraday's induction law (see Ref. [5] for a comprehensive general review). Let us consider a coil made with N_T closed turns spanning an area A with normal unit vector \mathbf{n} and boundary ∂A ¹, as shown in Fig. 1. The output voltage is given by the total rate of change of the linked magnetic flux Φ as in Eq. (1), where the negative sign (a.k.a. Lenz's law) means that the induced e.m.f. tends to generate a current which, by the right-hand rule, gives rise to a field opposing the variation of the flux.

¹ Practical coils are usually planar; however, the shape can in principle be arbitrary. Also, note that the loop will always be closed as the voltage terminals need to be connected to a voltmeter.

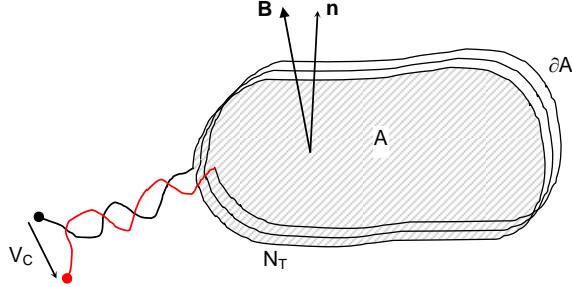


Fig. 1: Schematic representation of a search coil

$$V_C = -\frac{d\Phi}{dt} = -\frac{d}{dt} \iint_A \mathbf{B} \cdot \mathbf{n} dA = -\iint_A \frac{\partial \mathbf{B}}{\partial t} \cdot \mathbf{n} dA - \oint_{\partial A} \mathbf{v} \times \mathbf{B} d\ell. \quad (1)$$

The r.h.s. of Eq. (1) implies that an output voltage can be generated in two ways:

- by a time-varying field, typically measured with a fixed coil (known in this case as a *flux loop*);
- by moving and/or deforming the coil with local velocity \mathbf{v} . This is usually achieved by rigid translation (in non-uniform fields only), rigid rotation (in which case we speak more properly of *harmonic coil*), or by forming a variable-geometry loop using for instance a stretched wire suspended between two translation stages [6].

If the geometry and position of the coil are known, the average value of the field over the coil's area can be inferred from either the instantaneous or the time-integrated voltage:

$$\Phi - \Phi_0 = - \int_0^t V_C dt. \quad (2)$$

The advantage of using integration lies in the fact that the measured volt-seconds correspond directly to magnetic flux variations between the start and end configurations, irrespective of the irregularities of the path followed; moreover, the integration automatically filters out unwanted high-frequency noise components. On the other hand, any systematic error at the integrator's input, such as a voltage offset, gives rise to a drift which increases inexorably with time and may affect the accuracy of the result.

1.3 Requirements

In general, magnetic measurements are an essential step at various stages in the lifetime of accelerator magnets. In the prototyping phase, measurement results are used to verify design calculations, material properties, and fabrication methods; during series production the main aim is to monitor the quality of the manufacturing process; finally, throughout the operation of the machine, spare, reference, or refurbished magnets often have to be re-measured to assess their response to new operating conditions (current cycling, environmental conditions) or time-dependent effects such as those due to eddy currents or ageing.

In the common case where beam optics can be approached via the thin-lens approximation (magnet length \ll betatron oscillation wavelength [7]) the main target of a magnetic measurement is the integral field quality. This may include the strength of the main component, harmonic content (or, equivalently, uniformity of the main component), field direction, and location of the magnetic centre. All these quantities, in the most general case, have to be provided as a function of current excitation and history, ramp rate, and time. Local field quality, restricted by definition to a small fraction of the magnet's length, may also be of interest for several reasons, e.g., for spotting manufacturing or material defects [8], or for mapping the field as needed, for instance, in the case of spectrometers. High-resolution dynamic field measurements are also often used indirectly to localize and follow the

propagation of quenches in superconducting magnets, as the redistribution of current in the magnet coil's cross-section causes detectable field harmonic changes [9].

These measurements have typically to conform to exacting specifications: long and narrow magnet gaps; main fields ranging from $\sim 10^{-4}$ T (typical residual field level) to about ~ 10 T (in superconducting magnets); relative accuracy of the order of 10^{-4} for the main field and 10^{-6} for the field errors. Given these constraints, slender search coils are often the natural and most cost-effective choice, especially so for integral measurements where harmonic (rotating) coils provide directly a field description in the format required by beam optics.

2 Circuit model

2.1 General case

In general, a search coil can be represented as the secondary winding of a transformer circuit where the primary is the magnet generating the flux, as the circuit scheme in Fig. 2 shows. The coil is therefore characterized by the following parameters:

- L_c = self-inductance, depending on the coil geometry only
- L^* = mutual inductance, depending on the coil geometry and the distribution of field (hence on position and orientation of the coil with respect to the magnet, field level as the magnet approaches saturation, ramp rate in case of eddy currents that modify the field pattern, etc.)
- R_c = electrical resistance, depending on the coil geometry and material (and, to a small extent, on field level via magneto-resistance)
- C_c = self-capacitance, depending on the geometry and the dielectric properties of the insulation
- C^* = parasitic capacitance, depending on the geometry and the relative position w.r.t. nearby conductors such as the magnet coil and yoke, external structures, ground planes, etc.

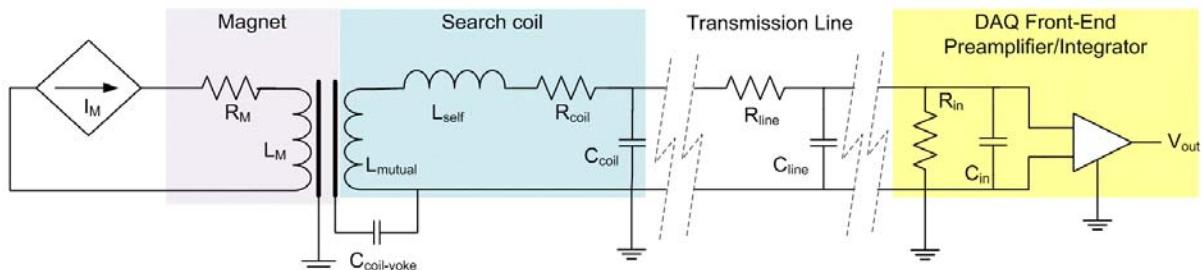


Fig. 2: Circuit diagram of a typical search coil measurement setup.
The magnet is supplied by the current source I_M .

The output voltage of the coil is fed to a data acquisition system (DAQ) via a transmission line, typically a twisted-pair or a coaxial cable. The acquisition, usually a digital integrator or an ADC/voltmeter, is mainly characterized by its own input impedance which should be as high as possible to minimize the current circulating in the coil.

2.2 Low-frequency approximation

A complete circuit description like the one seen above is fortunately unnecessary in the majority of practical cases as long as the spectral content of the signal remains sufficiently below the resonant frequencies (this condition is typically satisfied below 10–100 kHz, which excludes for example very fast pulsed linac or kicker magnets). In this case, one may use a low-frequency approximation (see

Fig. 3) where the magnet coupling is replaced by a simple voltage source and the only meaningful parameters are the resistance of the coil and of the input stage of the DAQ. The resulting output tension V_{in} and parasitic current I_{coil} are given by

$$V_{in} = \underbrace{\frac{1}{1 + \frac{R_{coil}}{R_{in}}}}_{k_R} V_{coil}, \quad I_{coil} = \frac{V_{coil}}{R_{coil} + R_{in}}. \quad (3)$$

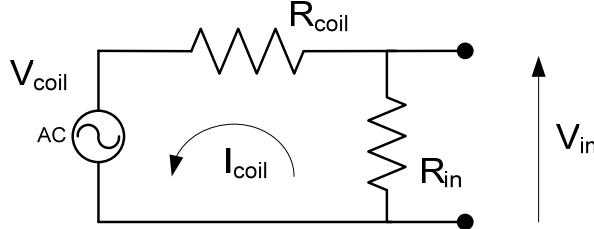


Fig. 3: Simplified search coil circuit in the low-frequency approximation. The inductive coupling to the magnet is represented by an AC voltage source and transmission line effects are neglected.

For typical input resistances in the $10^6 \Omega$ range and coil resistances of the order of a few ohm, parasitic currents I_C of the order of a few $10^{-6} \Omega$ are expected. These may have the following detrimental effects on the accuracy of the measurements:

- the measured voltage is smaller than the real coil voltage by the factor $k_R = 1 + R_c/R_{in}$, which must be taken into account if high accuracy is desired;
- especially at high frequencies, the measurement might be perturbed by additional voltages due to the coil self-inductance and capacitance.

The factor k_R can be used to apply a correction to the results, however, care must be taken to consider the possible external influences on the values of resistance such as operating temperature (for copper, $\partial R/R/\partial T \approx 0.004/\text{°C}$) or magneto-resistive effects. The major factor affecting R_{in} is normally the gain of the internal preamplifier stage, although one has also to expect a degree of dependency upon the frequency content of the signal. Joule heating effects in the coil due to I_C , on the other hand, are normally completely negligible (for example, an air-cooled 30 μm wire can carry indefinitely 5 mA without appreciable temperature rise).

2.3 Dipole-compensating coils

As explained in detail in another lecture of this School [10], accurate higher harmonic measurements require the use of an array of coils connected in such a fashion as to suppress the signal of the main field component. This ‘bucking’ or ‘compensated’ configuration not only improves the S/N ratio of the field error signal; crucially, it also abates the sensitivity to mechanical and geometrical imperfections of the rotating system. As an example, Fig. 4 shows a dipole compensation scheme in which two equal coils are mounted on the same rotating shaft parallel to each other. Two acquisition channels are run in parallel: coil 1 provides the so-called absolute signal, containing essentially the main dipole harmonic, while coil 2 is connected in series opposition to obtain the compensated signal which, ideally, contains only contributions from the error field. Assuming equal parameters for the two channels, absolute and compensated voltages can be calculated from:

$$V_{in}^{ABS} \cong \frac{1}{1 + 2 \frac{R_C}{R_{in}}} V_{C1} + \frac{R_C}{R_{in}} V_{C2}, \quad V_{in}^{CMP} \cong \frac{1}{1 + 3 \frac{R_C}{R_{in}}} (V_{C1} - V_{C2}) - \frac{R_C}{R_{in}} V. \quad (4)$$

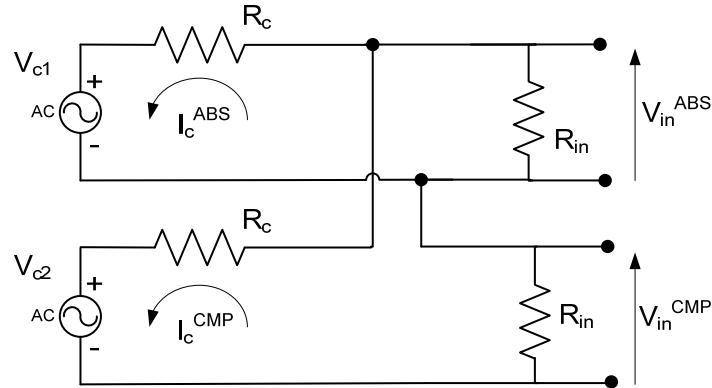


Fig. 4: Circuit diagram of two identical coils in a dipole-compensating configuration

Assuming for simplicity that $R_c/R_{in} \ll 1$ and that higher harmonics are small compared to the dipole (so that $V_{c1} \approx V_{c2}$) we find that also in this case the measured absolute signal is reduced by a factor $(1+R_c/R_{in})$. The compensated signal, ideally given by the difference $(V_{c1}-V_{c2})$, contains in fact an additional contribution from V_{c2} that practically reintroduces the suppressed main harmonic even in case of geometrically perfect compensation. Analogous compensation schemes with two, three, or four coils are commonly used for quadrupole measurements, where the objective is to cancel out the dipole and quadrupole components using an appropriate combination of coil areas and rotation radii. Compensation of higher-order harmonics is in theory possible but seldom implemented, owing to the extreme mechanical accuracies needed. In all cases, the quality of the compensation can often be improved by adding appropriately dimensioned resistors in series to the coils, and modifying Eq. (4) accordingly.

3 Coil design criteria

3.1 Main technological options

In the majority of cases a search coil for accelerator magnets has a rectangular, elongated shape which, among other things, helps to achieve a precise geometry and simplifies the calculation of calibration coefficients. The circular shape, by contrast, is far easier to make and is very common in the wider context of inductors for electronics. Search coils of widely ranging dimensions are referenced in the literature, from sub-millimetre sizes for point-like measurements to several metres for geomagnetic and RF applications. The length range of interest for our application normally lies from a few centimetres to one metre or so, longer magnets often being measured by a sequence of smaller coils.

The main technologies available for winding cores are listed in Table 1 along with their main advantages and drawbacks. In the size range of interest the approach we have found most convenient consists in winding a mono- or multi-filament wire around a thin rigid rectangular form, which is in turn fastened to a suitable support, usually a cylindrical shaft, permitting rotation or translation. We predominantly use individual G10 forms which can be produced relatively cheaply in large quantities (see Fig. 5). As an alternative, the winding core can be obtained directly on the rotating shaft by carving out a series of parallel grooves, as shown for example in Ref. [11]. In this case the mechanical accuracy and stability of the measurement head is improved, at the cost of greater difficulty in the process of winding or repairing the coils. This technology may be the only practical choice, for instance, in the case of small rotating shafts, where there is no room for a traditional type of assembly.

Table 1: Summary of main coil fabrication techniques

Coil Type	Pros	Cons
Separate winding form	relatively easy to machine, handle, and wind accurately large series production feasible allows sorting for bucking easy replacement if needed	need rigid support long-term stability problems
Monolithic support	accurate and stable	winding many turns more difficult potentially costlier
Printed Circuit Board (PCB) coils	extremely precise geometry high compensation order possible cheap in large runs	size limitations calibration coefficients need geometrical corrections (end effects, large aspect ratio conductor)
Air-core	cheap and easy to make	small sizes only poor geometrical accuracy

**Fig. 5:** A sample of CERN-made search coils of various shapes and sizes

3.2 Conducting wire

In this section we shall discuss the general criteria guiding the choice of a particular kind of wire, considering especially standard coils made by winding a single-filament wire in a spiral pattern. In the following sections we shall discuss the two major available alternatives, i.e., multi-filament and Litz wires.

Search coils are commonly wound from standard insulated copper wire, widely available on the market in gauges down to few hundredths of a millimetre. The next best metal in terms of conductivity, ductility and cost, i.e., aluminium, becomes competitive only for those current-carrying coils where the conductivity-to-weight ratio is important (aluminium's is about twice that of copper's).

The first real choice to be made concerns the shape of the wire cross-section, which could be either round or square. While a square shape could in principle allow more densely packed turns, it has two serious drawbacks: first, gauges smaller than 0.1~0.2 mm are difficult to find; second, and more important, with a square shape the unavoidable twist introduced during winding may result in a highly irregular geometry, detrimental to the accuracy of the measurement [12]. For these reasons, in our own work a round wire section has always been preferred.

The next major design choice is the diameter of the wire, which is the result of a compromise between two contrasting requirements: 1) a small cross-section to pack turns densely, thus saving space and approximating the geometrical ideal of a point-like cross section; 2) a large cross-section to increase mechanical strength and decrease electrical resistance (see Section 3.4). During winding, the wire has to be pulled taut to better adhere to the support; the ideal tension is just below yield, at which point the wire would neck and thus spoil the geometrical quality of the coil. Commercial wires are available in sizes normalized according to different standards, some of the most popular being the American Wire Gauge (AWG)[13] and the IEC 60228 [14].

The effective wire size must also include the necessary insulation, taking into account that the dielectric strength is approx. $20 \text{ V}/\mu\text{m}$ for PVC and $8 \text{ V}/\mu\text{m}$ for polyurethane. While in normal operating conditions coil voltages are by design of the order of only a few volts, off-normal events such as superconducting magnet quenches or power supply trips may give rise to dB/dt rates up to $10\text{--}100 \text{ T/s}$ and hence coil voltages up to several hundred volts. To give an idea, common polyurethane-coated wire has an insulation thickness between $3.5 \mu\text{m}$ and $9 \mu\text{m}$ according to standard IEC 60317 [15], corresponding to the capacity to withstand a (rather conservative) inter-turn voltage drop of at least 30 V.

3.3 Materials

The material used to make the winding core and the coil support must have the following characteristics:

- high stiffness: to minimize measurement errors due to deformation and vibrations
- low thermal expansion coefficient
- mechanically stable, with a hard, non-porous surface to minimize moisture absorption
- non-magnetic: to avoid a) perturbations of the field to be measured and b) magnetic forces
- non-conducting: to avoid the occurrence of eddy currents and the consequent field perturbations and magnetic forces (NB: all measurement methods, whether the coil is moving or stationary, imply a flux change and therefore an e.m.f.)

The materials most commonly used are listed in Table 2 along with their main properties. It should be noted that good mechanical properties are generally in conflict with machinability, with the notable exception of Macor. This is a boro-silicate glass ceramic, loaded with mica to inhibit crack propagation, which associates high machinability with standard tools (achievable tolerances down to one μm) to high stiffness and low thermal expansion. Its main disadvantage is the high price, plus the difficulty of finding monolithic pieces longer than a few hundred millimetres on the market.

In most practical cases, the default choice is glass-reinforced epoxy laminates which are cheap, reasonably rigid and stable and allow tolerances down to $\sim 0.01 \text{ mm}$, provided coated tools are used and frequently replaced. Among the standard grades classified by the National Electrical Manufacturer's Association (NEMA), G10 is the one with the highest hardness and the lowest thermal expansion. It is sensible to discuss the appropriate quality grade with the supplier, as the type of fibres, the chemical characteristics of the impregnation, etc. can make a great difference to the finished result. One must take into account the possible effects of the anisotropy of the composite, which, for example, should never be machined in the direction parallel to the fibres, lest the subsequent relaxation of in-built stresses warp irremediably the part. As another example, tubes fabricated with the fibres helically wound in one direction only can have nasty side-effects, such as unexpected torsion induced by temperature changes.

Table 2: Physical properties of most common materials for coil forms and supports

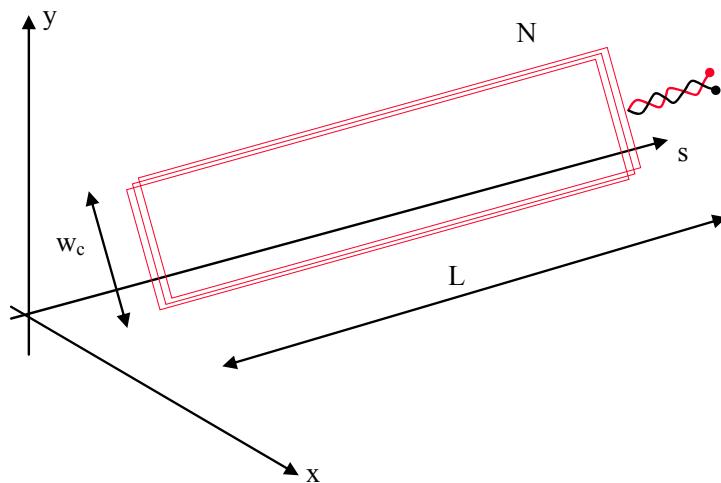
	Density	Young's modulus	Thermal expansion	Resistivity ^a	Dielectric constant	Magnetic ^b susceptibility
	ρ [kg/m ³]	E [GPa]	α @ 300 K [ppm/K]	σ [Ω m]	ϵ_r [-]	χ_m [-]
MacorTM	2520	64	0.9	$>10^{14}$	6	$<10^{-5}$
VycorTM (96% Si)	2180	66	0.8	$>10^{14}$	3.8	$<10^{-5}$
Quartz (fused Si)	2200	72	0.6	$>10^{14}$	3.8	$<2 \cdot 10^{-7}$
Carbon fibre	1600	250	6.5	10^{-5}	n.a.	$-1.6 \cdot 10^{-5}$
UltemTM 2300	1510	5.5	20	10^{17}	30	$<10^{-5}$
G10	1820	25	10	$>10^{14}$	5.2	$<10^{-5}$
Al₂O₃	3980	380	6.5	$>10^{14}$	9.1	$<10^{-5}$

^a All materials are practically perfect insulators, except for carbon fibre. ^b All are also non-magnetic.

At the two extremes of the performance scale we find glass-reinforced plastics like Ultem, cheap and easy to machine but rather poor mechanically, and sintered ceramics such as alumina (Al_2O_3), with outstanding rigidity and stability but exceedingly difficult to handle, hence expensive. While raw alumina powder is very cheap, being the basis for a number of large-volume industrial applications (production of aluminium, refractory materials, abrasives, electrical insulators etc.), the sintered material is very hard and fragile and so it has to be ground with special techniques using diamond-coated tools (tolerances can be a few micrometres). Only a few firms in Europe are able to produce large parts; for instance, the 1.3 m long tubes used for the LHC rotating coil system (see Section 5) had to be built in three parts and then glued together. Other materials are normally reserved for very specific cases: quartz is also difficult to machine, but useful for cryogenic measurements due to its thermal stability; or carbon fibre which is slightly conductive, but has the highest stiffness-to-weight ratio.

3.4 Design of rectangular coils

We shall now consider the design criteria for a rectangular-shaped coil, by far the most common case. We will assume an ideal geometry as shown in Fig. 6, where the transverse size of the windings is negligible w.r.t. the size of the coil.

**Fig. 6:** Ideal geometry of a rectangular search coil

The main design parameters are listed below:

- 1) R_c = coil resistance. As seen above, a low resistance minimizes the error due to a finite input stage impedance as well as the thermal noise $\sqrt{4k_b RT \Delta f}$.
 - 2) L_c = coil self-impedance. This parameter is important only in the high-frequency regime.
 - 3) L = coil length. This parameter is normally dictated by the size of the magnet being measured and by the nature of the information sought, e.g., integral or local field.
 - 4) w_c = coil width. This parameter too is often fixed by geometrical constraints (e.g., a coil of width w_c rotating at a radius $R \gg w_c$ will be blind to the harmonic of order $2\pi R/w_c$).
 - 5) N_T = number of turns.
 - 6) \emptyset_w = wire diameter.
 - 7) A_c = coil area exposed to flux change. The area should be dimensioned to obtain an output voltage as large as possible, although not exceeding the input range of the acquisition electronics (typically, 5 or 10 V peak-to-peak) to avoid saturation of the input stage (this can have nasty side-effects such as a slow recovery of the nominal performance of the electronics).
- The output voltage can be derived from Eq. (1) for three canonical cases as follows:

$$-V_c = \frac{\partial \Phi}{\partial t} = \begin{cases} A_c \dot{B} & \text{Fixed coil in a time-varying field} \\ A_c B \omega & \text{Coil rotating at angular speed } \omega \text{ in a constant field} \\ A_c \nabla B v & \text{Coil translating at speed } v \text{ in a constant field with a gradient} \end{cases} \quad (5)$$

Note that the calculation of peak voltage must include a reasonable safety factor, at least 2~3, to takes into account the unavoidable voltage fluctuations due to irregularities of motion, power supply ripple, mechanical or electrical noise etc.

These seven parameters are linked by three equations:

$$\begin{cases} A_c = N_T \ell_c w_c \\ R_c = \frac{8}{\pi} N_T \rho \frac{\ell_c + w_c}{\emptyset_w^2} \\ L_c = \frac{\mu_0}{\pi} N_T^2 \left(\ell_c \ln \frac{\ell_c}{\emptyset_w} + w_c \ln \frac{w_c}{\emptyset_w} + 2\sqrt{\ell_c^2 + w_c^2} - \ell_c \sinh^{-1} \frac{\ell_c}{w_c} - w_c \sinh^{-1} \frac{w_c}{\ell_c} - \frac{7}{4}(\ell_c + w_c) \right) \end{cases} \quad (6)$$

In other words, one may fix four parameters and the other three will follow. As an example of parametrization, Fig. 7 shows what one obtains by fixing coil width and wire diameter, and considering area and length as independent.

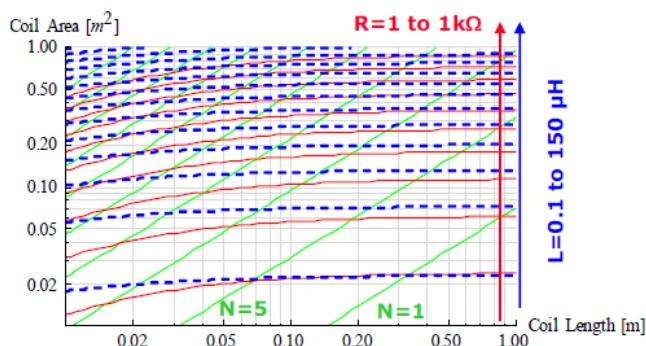


Fig. 7: Chart showing coil parameters L_c (in blue), R_c (red) and N_T (green) as a function of coil length and total area, assuming $w_c = 10$ mm and $\emptyset_w = 0.06$ mm

4 Coil winding methods

Once the general parameters of a coil such as geometry, number of turns etc. have been chosen, then different manufacturing methods can be employed, each one with specific strengths and weaknesses. The main possibilities are treated in the following sections.

4.1 Single-strand coils

In the simplest case, a single strand of wire can be wound around the form in a helical pattern to make a coil. This basic technique is widely adopted in the larger context of discrete electronic inductors and is implemented in a number of commercial winding machines, all based on a simple combination of rotational plus cyclic translational motion of the winding form w.r.t. the wire bobbin. This method, clearly well-suited to the semi-automatic production of large numbers of identical parts, has two main disadvantages: 1) the maximum size allowed for the coil is limited by the size of the winding machine; and 2) the geometrical quality of the resulting coil can be quite bad, especially for high turn counts as shown in Fig. 8. In our own experience, coils up to 3000 turns fabricated with this method are used almost exclusively for applications such as quench localization in superconducting magnets, where the accuracy of the geometrical calibration factors (see Section 7.1) is of secondary importance. Wire in gauges from 32 to 200 micrometres is commonly used. Wire tension should be monitored accurately during winding, because the wire (especially the smaller one) breaks very easily.

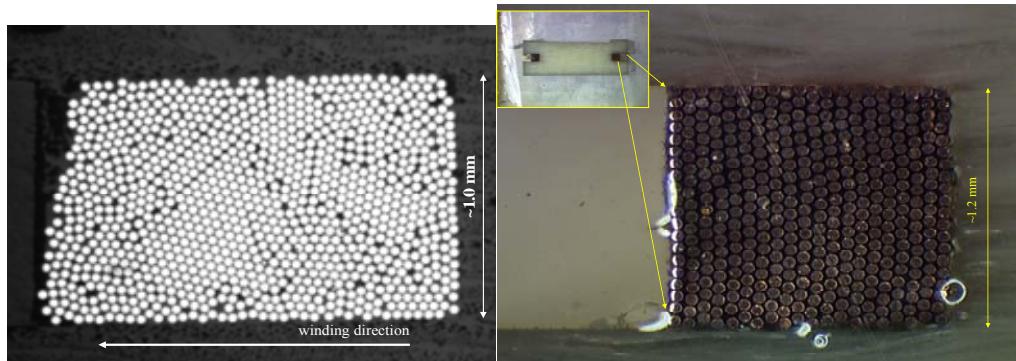


Fig. 8: Comparison between the cross-sections of a 900-turn single-filament, machine-wound (left) and a 400-turn, 20-filament manually wound coil (right)

Figure 5 shows a number of samples of single-strand coils taken from CERN stock. To enhance long-term stability after winding we brush on an outer layer of methyl acetate glue (Cementit™ Universal) diluted with acetone. In case of short circuits or other damage, it is possible to strip off the wire and recover the form by dissolving the glue with more acetone or by heating it up to around 100°C.

4.2 Litz-wire coils

The winding process can be greatly accelerated by using pre-assembled cables made with a plurality of wires. The traditional solution consists in the so-called ‘Litz’ wire (from *litzendraht*, ‘woven wire’), a kind of cable including up to several thousand extremely fine insulated filaments (see Fig. 9). Litz wire has been used extensively since the 1940s in the radio frequency domain, where the subdivision of the conductor substantially reduces the AC losses due to skin effect. By using a wire with the appropriate number of conductors a whole coil can be made with just one winding turn, which is a big advantage in case of awkward geometries (e.g., very long coils, or thin grooves carved in a shaft). Unfortunately, this fabrication method carries some heavy penalties:

- there must be as many solderings as there are turns, and the wires to be connected must be painstakingly singled out one at a time;

- more importantly, each conductor follows a spiraling path along the coil and thus its transverse position undulates with a period equal to the twist pitch of the cable, and an amplitude up to its diameter. This implies that the local coil coefficients may vary in the range of several per cent, so that accurate coefficients can be obtained only by averaging over a given longitudinal tract.

At CERN, the most prominent use of Litz wire can be found in the coils for integral measurement of main Super Proton Synchrotron dipoles, which date back to the 1980s and have unique length requirements (more than 7 m).



Fig. 9: An example of multi-filament Litz wire [from www.litz-wire.com].

4.3 Multi-strand coils

A better performing alternative to Litz wire is the multi-strand wire, a pre-formed ribbon which allows precise layering over multiple turns leading to very uniform geometries and highly accurate measurements. Examples of application can be found, e.g., at SLAC [16] and at CERN. The cable used at CERN is manufactured practically on-demand by MWS Wire Industries, Westlake Village, CA and consists of from 3 up to 20 polyurethane-insulated filaments with diameter from 170 μm down to 60 μm respectively. The filaments are bonded with polyvinyl butyral (PVB), a compound easily soluble in alcohol which facilitates soldering of the connections.

Winding a multi-strand coil, as shown in Fig. 10, is a fully manual procedure that requires skill and care. The winding form, which for practical reasons has in our case a maximum length of about two metres, pivots around a horizontal axle while the wire is pulled taut and guided to slide into the groove. As the wire rolls off the spool it dips in a bath of AralditeTM AY103 with HY991 hardener (a transparent, low-viscosity, general-purpose epoxy adhesive). After winding, the coil is inserted into a polymerization clamp to squeeze air bubbles out and to avoid the possible bulging of the wire at the corners (note that the wire tension tends to drop naturally along the short edge). A release agent (e.g., Teflon spray) should be added at this stage to prevent the coil from sticking to the clamp. The curing is carried out under IR lighting for a few hours at a temperature generally below 100°C, a precise thermal cycling being not strictly necessary. As a certain amount of springback upon removal of the clamp is to be expected, the geometry of the finished coil must be checked accurately.

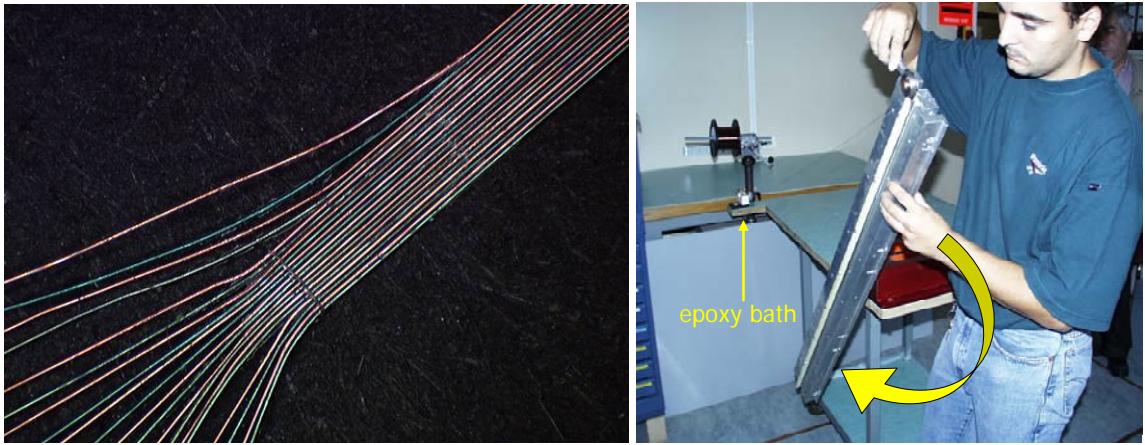


Fig. 10: An example of 20-filament flat wire (left) and a coil being manually wound (right)

The benefits deriving from this winding method are evident in Fig. 8, which shows the beautifully regular cross-section of a 20×20 turns coil developed to measure superconducting LHC magnets at room temperature (hence with currents of a few ampere and fields of a few millitesla). In our experience, such a high turn-count is unfortunately rather hard to obtain as a large proportion of the units (up to 2/3) is wasted, mostly due to the thin wire breaking during winding or the insulation tearing under the pressure of the clamps.

The electrical connections, very delicate due to the small size of the wires, must present the smallest possible area to the changing flux so as not to affect accuracy (this is important if the coil is designed to be calibrated and/or used when fully immersed in a magnetic field, see Section 7.2.1). First of all, the two ends of the ribbon have to be folded and overlapped as shown in Fig. 11. Then, all necessary solderings are made under a binocular microscope with the help of a tiny PCB connector, made with a complex procedure that involves an initial machining phase followed by aluminization and then deburring with a paste made by mixing micro glass balls, cementite and acetone. The connector should be glued or built into the coil to be easily removable seeing that, as one might expect, short circuits or broken wires are quite common. A thermo-retractable sleeve wrapped around the connector and the output cable is mandatory to relieve stresses and improve reliability. To conclude, it should be noted that this type of connector represents a constraint w.r.t. the minimum transversal size of a coil, and this may limit the applicability of this technology to very-small-aperture magnets.

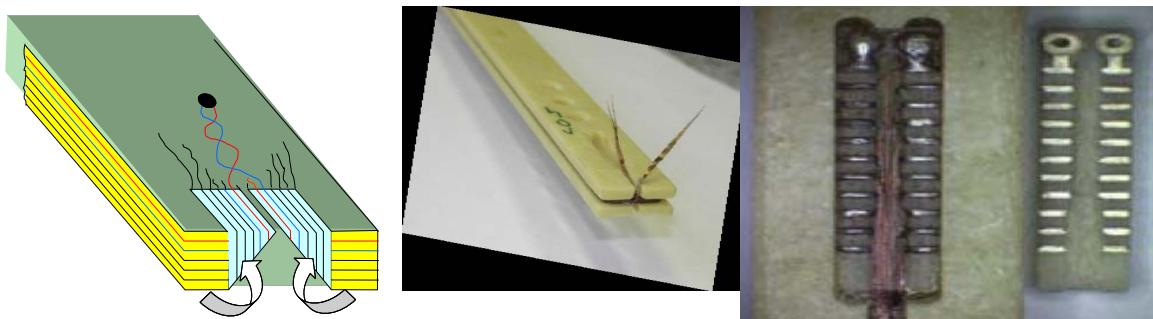


Fig. 11: Detail of the electrical connections of a multi-filament ribbon. The drawing on the left shows schematically the folding given to the ends of the wire (for clarity only one turn is represented). In reality, the folds are overlaid to minimize flux change pick-up (centre). The photo on the right shows the 10-mm PCB used to make the connections.

4.4 Printed Circuit Board coils

A rather different technological option for building coils consists in the possibility of laying spiral traces on a PCB, as shown in Fig. 12 [17]. A coil built with this technique is characterized by very precise (few micrometres) but relatively sparse traces, with a width of the order of 100 µm, thickness about 30 µm, and horizontal gap between traces also about 30 µm. The figure shows a board designed to be a part of a multi-layer assembly, as one can see from the many small metalized holes included to connect the ‘turns’ between layers (assemblies with up to 32 layers have been realized, although the cost tends to escalate rapidly).

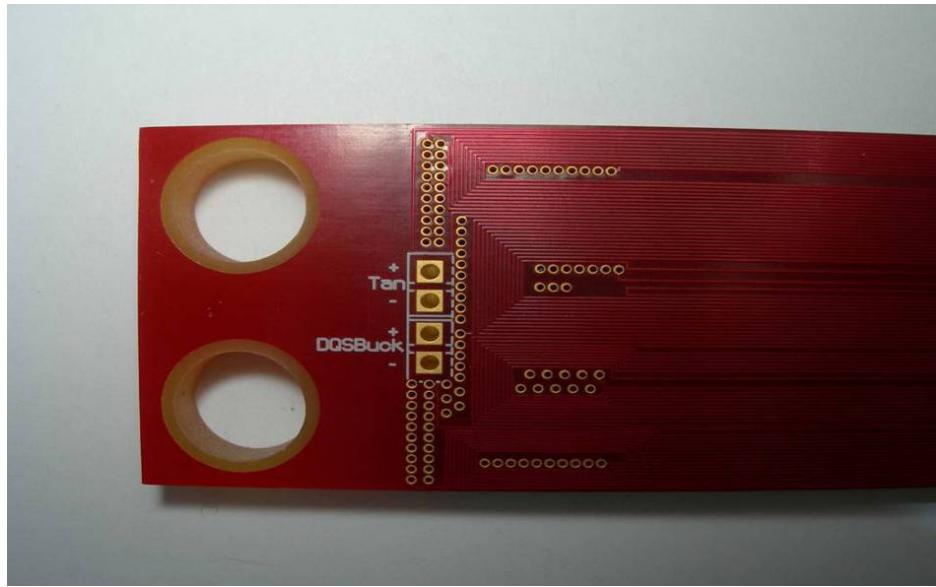


Fig. 12: Quadrupole-bucked, multi-layer PCB coil (courtesy of J. Di Marco, Fermilab).
The large holes on the left are used for alignment.

This kind of coil has some key advantages such as the possibility of industrial (read cheap and reproducible) manufacturing of large runs and the potential to scale down to coil sizes of a few millimetres, very difficult to make otherwise. Moreover, one may take advantage from the very precise positioning of the tracks to implement advanced multipole compensation schemes, e.g., dipole or quadrupole compensation with very large bucking factors (in the 1000–10 000 range) or more exotic higher-order schemes. On the other hand, there are a number of serious issues:

- owing to the low conductor density, the sensitivity is comparable with traditional designs only if many layers are stacked
- higher conductor density implies small cross-section and thus higher electrical resistance
- optimizing the trace layout is not a trivial task and requires close interaction with experts
- cost is high for short runs and multiple layers
- PCB lengths above ~1 m are not easily feasible
- the high aspect ratio of the coil must be taken into account in the calculation of coil sensitivity factors for accurate measurements [12]
- the end regions are comparatively large, so they too must be taken properly into account if the PCB is designed to be fully immersed in a magnetic field

The coil shown in Fig. 13 represents an interesting variation on this theme, where the rigid PCB is replaced by a flexible Kapton foil which is bent and glued directly onto the surface of a quartz rod [18]. While such a solution is undoubtedly elegant and precise, provided the foil is aligned correctly to the rod, the very low number of coil turns forces the instrument to turn rather fast.



Fig. 13: A flexible coil printed on a Kapton foil and rolled around a Ø20 mm rod (from Ref. [18])

4.5 Coils for special applications

Let us mention briefly a few special cases of application.

4.5.1 Curved coils

For strongly curved magnets, accurate beam dynamics calculations may require the field integral to be measured along the actual path followed by the particles, which will be in general an arc of circle between two straight entry and exit segments. In such a case, the coil winding methods discussed above can be adapted at least in two ways:

- the coil is fabricated straight, then it is bent to the desired shape and fastened to a rigid support. The major drawback in this case is that bending stresses stretch the winding at the outboard and crush it at the inboard, causing the total surface to change and risking, at worst, detachment from the support at the inboard. The robustness and long-term stability of the coil are also impaired.
- The coil is wound directly on a curved form. While at the outboard the wire can be pulled taut as needed, at the inboard obviously this is not possible. The wire has to be glued to the form in a gradual and extremely careful way, possibly with the help of small clamps. Accurate geometry can be very difficult to achieve.

4.5.2 Coils for cryogenic tests

Magnetic measurements and quench localization in superconducting magnets often require coils to operate at liquid helium temperature. Mainstream materials such as copper and glass-reinforced epoxy in general cope well with cryogenic conditions. Grade 3 wire should be specified to guard against thermal shocks, and differential thermal contractions (see Fig. 14) should be taken into account to estimate stresses in the windings:

$$\sigma_{Cu} = \frac{\int_{4.2}^{300} (\alpha_{Cu} - \alpha_{Support}) dt}{\pi \frac{\emptyset_{Cu}^2}{2 A_{Support}} N_T \frac{E_{Cu}}{E_{Support}} - 1} E_{Cu} \quad (7)$$

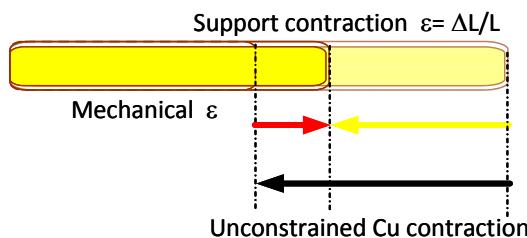


Fig. 14: Differential thermal contractions in a coil

Where necessary, specific adhesives such as Araldite™ GY285 (very low viscosity) + Jeffamine™ hardener should be used. This one, in particular, confers elasticity to the epoxy and the specific capacity to withstand differential expansion and thermal shocks without delamination.

4.5.3 *Coils for radiation environment*

In certain applications (e.g., ‘B-train’ field measurement system for real-time accelerator control) a coil has to be installed inside an operating magnet, with the consequent ionizing radiation risk due to synchrotron radiation and beam losses. Because of radiation, a search coil would suffer the same kind of effects which are well known for magnet coils, i.e.,

- copper can be activated if trace impurities are present (small diameter wire grades are normally 99.90% pure, with up to 0.04% oxygen to improve ductility);
- ceramic materials and glasses have very low activation;
- organic components like adhesives and cable insulation can be moderately activated. The major problem is related to certain polymers (e.g., polyethylene) which, in time, disintegrate completely.

In CERN’s experience, over several decades no adverse effects have ever been observed in the few coils that have been exposed to radiation.

5 Coil head assemblies

Once the coil has been made, one has normally to fix it to an appropriate support to make a usable measurement head. While clearly there are as many possible ways to do this as there are different applications, we shall discuss here the rotating coil shafts made for LHC cryodipoles [19] as an example on which to base a few general considerations.

The coil shaft is shown in Fig. 15. The main points to be noted are the segmented structure, which allows the coil to follow the curved cold bore; the outer ceramic tubes, which guarantee the required torsional rigidity over the whole length of the shaft; and the vertical stack of three parallel, 1.1 m long rectangular coils within each segment, designed to provide dipole compensation with the third coil used for rotating mass symmetry and as a spare.

The major factor affecting the accuracy of multipole measurements is the alignment of the coil w.r.t. the rotational axis of the assembly, as even small errors in either radial position or angular orientation dramatically affect the harmonics of interest. In our case, the alignment is provided by a number of ceramic pins, while the fastening function is performed with nylon screws and secured by applying a line of glue along the edges of the top and bottom coils. The adhesive used is UV Loctite™ 322, which makes a flexible bond that can be scratched off if the segment has to be disassembled for repairs. On the downside, this glue has a very low viscosity and tends to penetrate between the coil and the support, so UV light must be applied immediately after deposition to accelerate the polymerization and prevent any possible swelling beneath the coil. Another caveat concerns poor long-term behaviour, especially when repeated thermal cycles or shocks are applied, as the glue tends to crumble and the residue can pollute the ball bearings.

An important component in this shaft is represented by the flexible bellows connecting adjacent segments, which allow quasi-homokinetic rotation in the bent configuration. The bellows are made of titanium, which has both the large elastic range and high electrical resistivity required by its function. It should be noted that the bellows are fastened to the ceramic shaft by means of Araldite. As the bond between the titanium and the ceramic was initially found not to be good, the surface was first roughened with acid (unsuccessfully) and then etched mechanically.

The last components to be highlighted are the ball bearings, which must be rigorously non-magnetic and non-conducting in order to work inside a 9 T field. The parts chosen are made of Si_3N_4 and are characterised by extremely low friction and smooth movement. They are mounted permanently at the extremity of each shaft segment, so as to provide a stable axis reference for calibration. While many suppliers of standard sizes exist, unfortunately our application required a custom design that could be obtained only from one manufacturer (Cerobear GMBH, Herzogenrath, DE) which, predictably, meant high costs.

A very interesting alternative is represented by the so-called Olive Hole Ring Jewel bearings, as used for instance at SLAC [11]. These parts are made of exceptionally hard artificial sapphire (monocrystalline Al_2O_3) and offer a very small static friction coefficient of 0.15 associated to an extremely low price. Their major drawback lies in the fact that they cannot bear any axial load, so existing mechanical designs have to be modified accordingly.

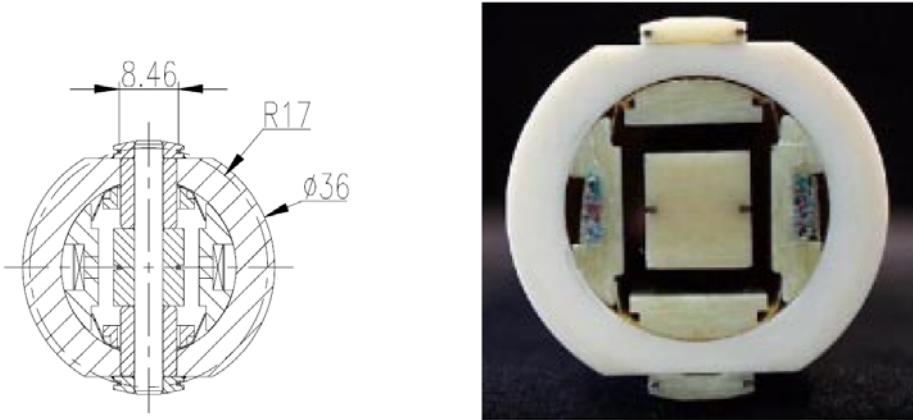


Fig. 15: Cross-section of a 15 m long, dipole compensated coil shaft for LHC main cryomagnets

6 Quality controls

At various stages during the fabrication of a search coil it is possible to carry out a number of quality-control checks meant to verify conformity with specifications. Simple go/no go tests and more precise measurements have to be done between manufacturing steps, prior to delivery, and also during operation (routine or on-call maintenance). In the following sections we shall discuss different kinds of checks as they are carried out at CERN.

6.1 Visual and geometrical inspection

To start with, careful visual inspection can easily reveal irregularities in the geometry as well as air bubbles, cracks or swellings in the epoxy (a binocular microscope can be of help). Next, accurate mechanical measurements of the length and width of both winding form and finished coil are mandatory to assess the quality of the winding and curing process, as well as to gather essential parameters for the calculation of calibration coefficients (see Section 7.1). While sub-micrometre 3D coordinate measuring machines are nowadays readily available and provided with knife-edge feelers that can penetrate easily into the winding groove, reasonably good results can be obtained inexpensively by means of simple rigs such as that shown in Fig. 16. In this case, a standard mechanical micrometer is made to slide along all four edges of the coil, which is then reversed to repeat the length and width measurements and allow cancellation of systematic offsets by averaging.

The data obtained from tests done on a batch of 15 1150 mm \times 8.5 mm, 6 \times 6 turn coils show that longitudinal variations of coil width amount to about 1% and are in good measure due to the irregularity of the winding support. Local variations of the winding width amount to about 10% and

add considerably to the total spread, despite the comparatively high quality obtained with 6-strand flat cable. The main source of this variation seems to be the thickness of the polymerized epoxy layers, while wire diameter variations play a minor role. In addition, comparison between the mechanical coil area thus obtained and the magnetic surface (Section 8.1) for a more recent series of about 150 units shows that the spread of the correlation is about 1%, i.e., adequate to spot and reject major faults but not good enough to allow accurate prediction of magnetic surface from mechanical measurements.

For the sake of completeness it should be remembered that accurate geometrical measurements are in principle possible with microphotogrammetric techniques, both optical and radiographic (in this case, non-destructive).

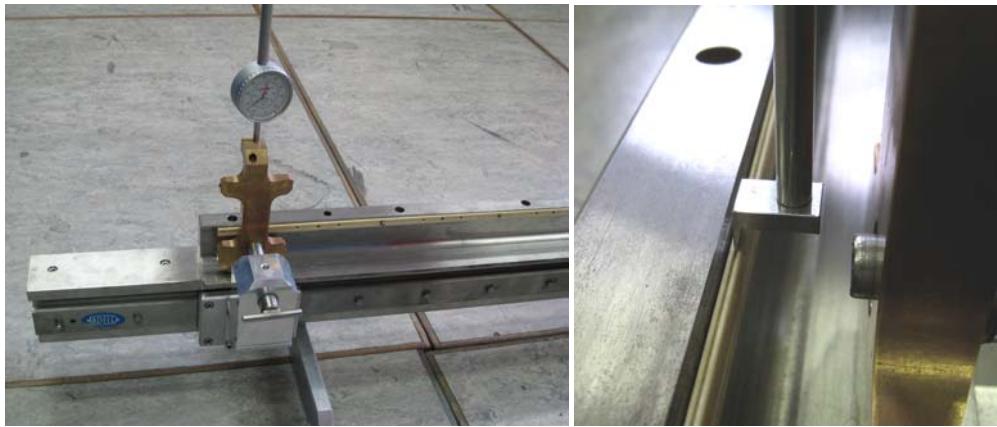


Fig. 16: A mechanical calibration rig for the geometry of rectangular coils. The micrometer on the left can be turned by 180° to cancel out offsets and by $\pm 90^\circ$ to measure the overall length. The detail on the right shows the knife-edge feeler used to get into the winding groove (typically 0.4 to 1.2 mm wide).

6.2 Magnetic measurement of coil width

As the width of a coil is of paramount importance for accurate measurements (see Section 7.2) we shall now describe a measurement method which is based on a strictly localized magnetic field and is therefore able to provide information on its magnetic equivalent directly. The principle is illustrated in Fig. 17, which shows the field profile we would ideally like to obtain along a given coil, corresponding to the total flux:

$$\Phi = N_T \int_0^L w(s)B(s)ds \approx N_T \int_{\bar{s}-\Delta/2}^{\bar{s}+\Delta/2} w(s)B(s)ds \approx N_T w(\bar{s}) \underbrace{\int_{\bar{s}-\Delta/2}^{\bar{s}+\Delta/2} B(s)ds}_{Bd\ell_{ref}} . \quad (8)$$

Once the flux is obtained with any one of the standard methods (i.e., by flipping or translating the coil, by pulsing or by AC modulating the field) the average equivalent magnetic length can be easily calculated from

$$w(\bar{s}) = \frac{\Phi}{N_T Bd\ell_{ref}} . \quad (9)$$

A possible way of obtaining the desired flux distribution based on permanent magnets and a suitably shaped iron yoke is also shown in Fig. 17. Alternative implementations based, for example, on AC coils may provide higher sensitivities, although permanent magnets ensure the stability required for accurate results. In all cases, the iron yoke has the important role of carrying the flux

return away from the coil, lest the net flux measured drop dramatically (see the dotted curve in the figure).

It should be noted that, unless one knows with extreme precision the field map of the magnetic source and thus the absolute value of $Bd\ell_{ref}$ [see Eq. (16)], the width measurement obtained has mainly a relative character. The absolute value can be recovered by comparing the profile $w(s)$ to an integral measurement providing the missing average value (see Section 7.2).

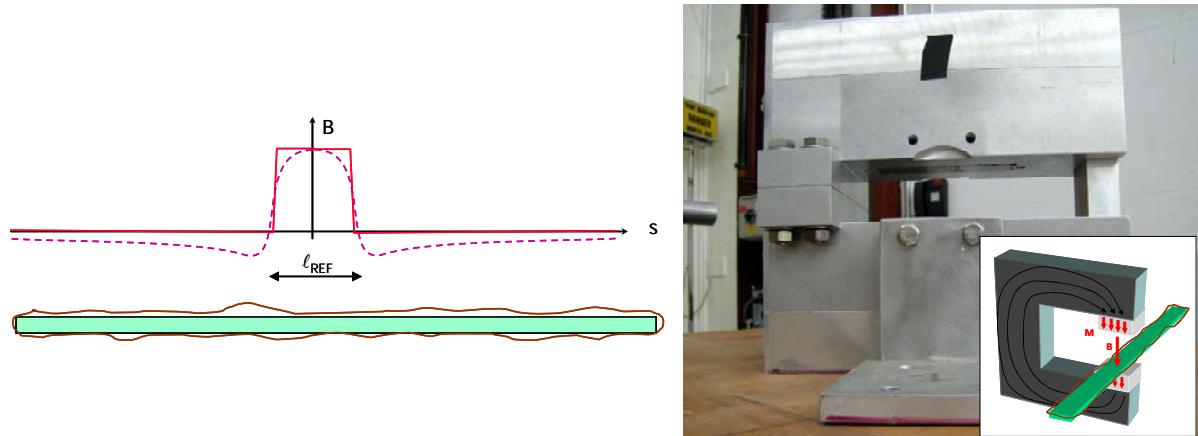


Fig. 17: Magnetic measurement of local coil width. The ideal and actual profile of the test field are shown on the left. The photo of an existing field source and its schematics (inset) are shown on the right.

6.3 Electrical quality controls

Basic electrical continuity tests should be repeated frequently throughout all phases of manufacturing, so as to detect as early as possible units with a broken wire. Once the coil is finished, its resistance should be measured as precisely as possible (i.e., with the four-wire method) in order to check for possible short circuits and to allow for later correction of the errors caused by finite input impedances (see Section 2.2). Referring back to Eq. (6-2), since the resistance is proportional to the number of turns we obtain

$$\frac{1}{R_c} \frac{\partial R_c}{\partial N_T} = \frac{1}{N_T} . \quad (10)$$

One can see that in case of high turn count the sensitivity of the measurement drops dramatically: e.g., for a 400 turn coil the short of a single turn lowers the resistance by only 0.25%, which is quite difficult to measure correctly (consider that such a variation for copper is equivalent to a temperature uncertainty of just 0.5°C). The coil self-inductance given by Eq. (6-3), on the other hand, is purely a function of geometry and therefore is less prone to ambient perturbations. As the dependence upon the number of turns is squared, the sensitivity to shorts will be inherently higher²:

$$\frac{1}{L_c} \frac{\partial L_c}{\partial N_T} = \frac{2}{N_T} . \quad (11)$$

² In fact, the shorted turn may add a mutual inductance that further subtracts from the initial self-inductance.

6.4 Coil polarity check

The polarity of a coil, i.e., the left- or right-handedness of the winding, is a seemingly simple issue that causes many headaches in practical situations, especially in complex acquisition systems where the signal path includes perhaps dozens of connections and therefore possibilities for inversions. The device shown in Fig. 18, based on a pair of AC-driven coils (of known polarity) represents an inexpensive way of resolving the ambiguity by using an oscilloscope to compare the phase of the source to that of the coil output. The optimal frequency range is typically around a few kHz, i.e., high enough to provide a strong output signal, yet lower than possible resonances.

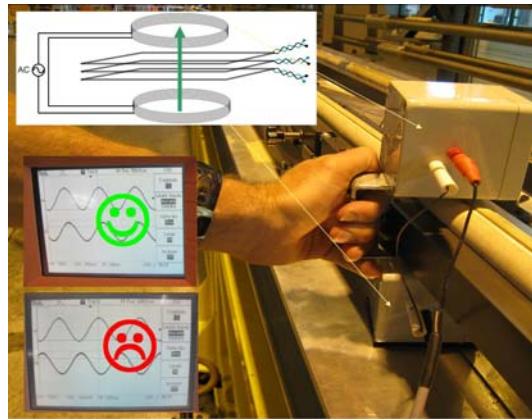


Fig. 18: AC coil polarity checker, testing one of the long LHC coil shafts

6.5 Magnetic quality controls

As one might expect, the surest way to prove that a coil is fit for magnetic measurements is indeed to go and actually make some, possibly in a well-known reference magnet which can be used to compare measured field and harmonics with expected values. Moreover, visualization and analysis of coil voltages, fluxes and other signals in the data acquisition and processing chain can often provide clues suggesting the source of observed malfunctions. In Fig. 19 we can see an example, taken from the test campaign of LHC dipoles, in which measurement results were intermittently grossly inaccurate while static resistance and polarity checks showed no apparent fault in the coils. In fact, the yellow absolute voltage trace shows clearly a disturbance occurring systematically at a certain angle during the forward rotation. This might be attributed to a failing contact, which was indeed identified and repaired.

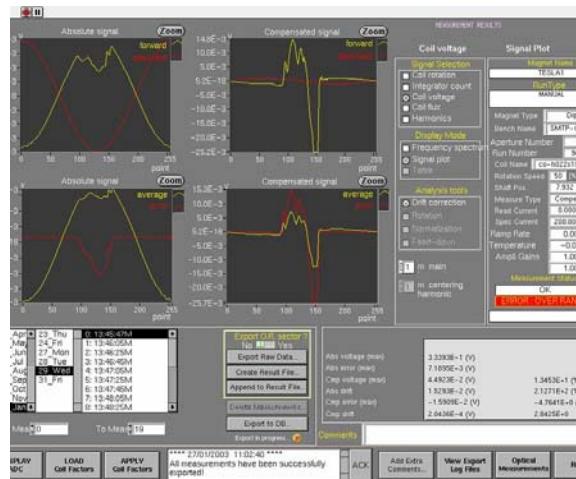


Fig. 19: Example of on-line detection of a failing electrical connection from the voltage signals of a rotating coil

7 Calibration of search coils

7.1 Basic theory of calibration

We shall henceforth deal with the principle and practice of coil calibration, referring essentially to the theory of harmonic coil measurements as developed in other lectures of this course [10], [20]. To recapitulate on the essentials: in the so-called thin-lens approximation, valid when the magnet length is much smaller than the wavelength of betatron oscillations, the field can be adequately described by a 2D power series expansion with complex coefficients $C_n = B_n + iA_n$, i.e., the harmonics. The flux Φ seen by a rotating search coil as a function of the azimuth ϑ is given by

$$\Phi(\vartheta) = \Re \left(\sum_{n=1}^{\infty} \frac{\kappa_n}{r_{ref}^{n-1}} C_n e^{in\vartheta} \right), \quad (12)$$

where the κ_n are complex coefficients known as coil sensitivity factors depending upon the geometry of the coil. In the case of an ideal rectangular shape³ they can be defined as

$$\kappa_n = \frac{N_T L}{n} (z_2^n - z_1^n) = \frac{N_T L}{n} \left(R_2^n e^{in\frac{\alpha}{2}} - R_1^n e^{-in\frac{\alpha}{2}} \right) e^{in\varphi_0}, \quad (13)$$

where $z_1 = R_1 e^{i(\varphi_0 - \alpha/2)}$ and $z_2 = R_2 e^{i(\varphi_0 + \alpha/2)}$ represent the position of the coil windings in the complex plane (x,y) at the start of the voltage integration, φ_0 is the initial phase, and α the coil aperture. By measuring flux increments over a discrete number N of angular steps and computing their Fourier transform coefficients Ψ_n , the field harmonics can be reconstructed from

$$C_n = \frac{2r_{ref}^{n-1}}{N} \frac{\Psi_{n+1}}{\kappa_n}, n = 1.. \frac{N}{2}. \quad (14)$$

On the basis of this relationship between coil and field coefficients, the process of coil calibration can be defined as **finding the set of κ_n needed to infer field harmonics C_n from the measurement of $\Phi(\vartheta)$** .

The expressions for the lowest-order κ_n in three common cases are written out in full in Table 3. All coefficients can be calculated from the coil length L , width w , and average radius R_o ; they all are proportional to total coil area $A_c = N_T L w$; they all increase like R_o^{n-1} . It is worth while pointing out that the κ_n are independent from the field being measured, i.e., search coils are inherently linear sensors. Indeed, contrary to a bafflingly popular misconception, the measurement gets easier as the field gets stronger as the S/N ratio of the output voltage improves.

³ The correction to be applied to take into account the finite size of the winding, as described in [12], is completely negligible for dipole, $<10^4$ for quadrupole, and $<10^3$ up to dodecapole if the cross-section is square and smaller than 1 mm², and the rotation radius is larger than 10 mm.

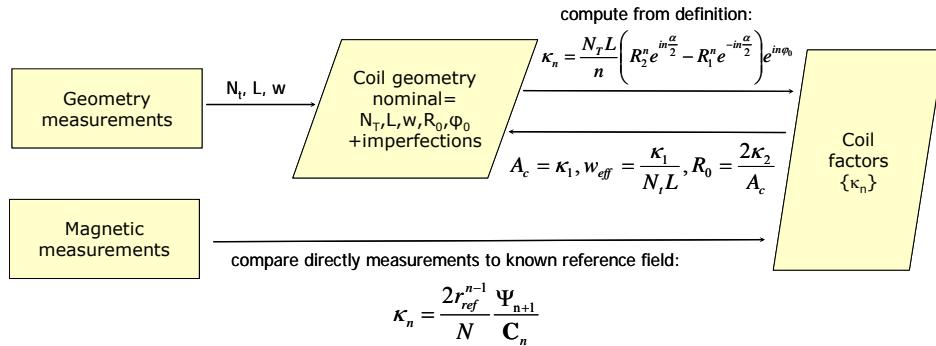
Table 3: Lowest-order coil sensitivity factors in three common cases

n	Radial coil $\phi_0 = 0$	Tangential coil $\phi_0 = \pi/2, w = 2R_0 \sin\alpha/2$	Tangential coil $\phi_0 = \pi/2, \alpha \approx 0$
κ_1	$N_T L w$	$N_T L w$	$N_T L w$
κ_2	$\frac{1}{2} N_T L w R_0$	$i \cos \frac{\alpha}{2} N_T L w R_0$	$i N_T L w R_0$
κ_3	$N_T L w \left(\frac{w^2}{12} + R_0^2 \right)$	$-\frac{1}{3} (1 + 2 \cos \alpha) N_T L w R_0^2$	$-N_T L w R_0^2$
κ_4	$N_T L w R_0 \left(\frac{w^2}{4} + R_0^2 \right)$	$-i \cos \alpha \cos \frac{\alpha}{2} N_T L w R_0^3$	$-i N_T L w R_0^3$
κ_5	$N_T L w \left(\frac{w^4}{80} + \frac{w^2 R_0^2}{2} + R_0^4 \right)$	$\frac{1}{5} (1 + 2 \cos \alpha + 2 \cos 2\alpha) N_T L w R_0^4$	$N_T L w R_0^4$
κ_6	$N_T L w R_0 \left(\frac{w^4}{16} + \frac{5}{6} w^2 R_0^2 + R_0^4 \right)$	$\frac{i}{3} N_T \cos \frac{\alpha}{2} (4 \cos^2 \alpha - 1) L w R_0^5$	$i N_T L w R_0^5$

When we are faced with the practical problem of finding the κ_n of a coil, different alternative routes can be followed. The dataflow diagram in Fig. 20 represents the relationships between the different variables and procedures entering into play. We see that

- The geometrical parameters needed to compute the κ_n derive from a combination of nominal values, geometrical measurements and magnetic measurements.
- Purely geometrical measurements, such as those described in Section 6.1, can in principle provide all basic length and width parameters. In practical situations, however, only the length can be measured to the level of accuracy required for high-precision end results (typically $10^{-3} \sim 10^{-4}$). (NB: if the coil is meant to measure shorter magnets, including fringe fields, then total length and surface are irrelevant).
- Measurements in a suitable reference magnet, i.e., one whose coefficients C_n are known, can be used to derive some of the κ_n and then work backwards to infer surface, width, and radius, which are then used to compute the missing coil coefficients.

The magnetic calibration option is the one that gives best results, since all sources of error are included in the same way they will be in real use. Failing that, due for example to the lack of a suitable reference magnet, geometrical measurements remain the second-best alternative. Calibration measurements in a dipole (to obtain coil surface) and in a quadrupole (to obtain rotation radius) as described in Section 8 are the norm; using higher-order multipole magnets is far less common.


Fig. 20: Schematic representation of the dataflow for search coil calibration

7.2 Calibration of finite-length coils

In practical use, the theoretical considerations discussed above in the ideal 2D case must be extended to the third dimension, taking into consideration end effects linked to both the coil and the magnet. Let us take as an example the configuration shown schematically in Fig. 21, where we shall assume that the coil length L is known accurately from geometrical measurements. We shall focus our attention on the case of a dipole magnet, the extension to other cases being trivial. The magnetic flux through the coil, which can be obtained, for example, by flipping the coil, fully rotating it, or keeping it fixed while the field is ramped up, is given by the following expression:

$$\Phi = N_T \int_0^L w(s)B(s)ds = \overline{B} \underbrace{\overline{w}_{eff}}_{A_{eff}} L = \overline{B} L \overline{w}_{eff} . \quad (15)$$

The main point here is that the quantity which is really of interest vis-à-vis accelerator physics is emphatically not this flux, but rather the integrated field (or $Bd\ell$, as it is commonly called), defined by (assuming the coil longer than the magnet):

$$Bd\ell = \int_0^L B(s)ds . \quad (16)$$

The following averaged quantities can also be defined:

$$\begin{cases} \overline{B} = \frac{1}{L} \int_0^L B(s)ds & \text{Average field} \\ \overline{w}_{eff} = \frac{\int_0^L N_T w(s)B(s)ds}{\int_0^L B(s)ds} = \frac{\Phi}{Bd\ell} & \text{Average effective width} \\ A_{eff} = L \overline{w}_{eff} & \text{Effective coil surface} \end{cases} \quad (17)$$

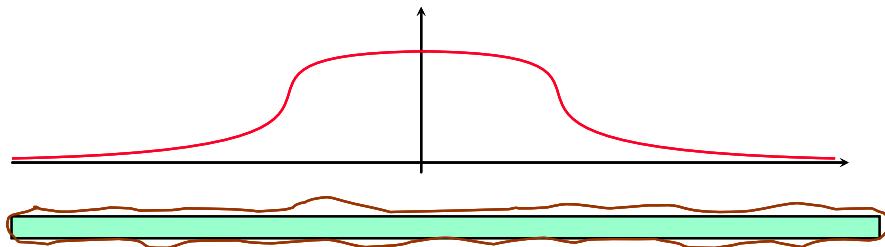


Fig. 21: Schematic representation of longitudinal field profile along a coil

In the general case, it is important to note that unless the width is strictly constant along the coil, the magnetic flux cannot be obtained from the mere knowledge of average field and average width alone, since

$$\frac{1}{L} \int_0^L w(s)B(s)ds \neq \frac{1}{L} \int_0^L B(s)ds \frac{1}{L} \int_0^L w(s)ds . \quad (18)$$

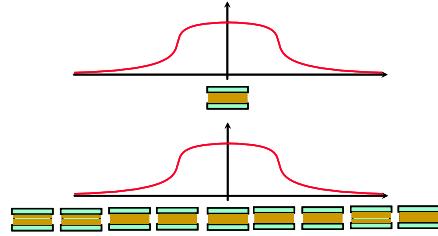
The effective width over the profile of interest [Eq. (17-1)], in fact an average width weighted with the field and lumped together with the number of turns for convenience, turns out to be the key parameter. Knowing that we can easily derive the wanted field properties from the flux measurement

$$\begin{cases} \bar{B} = \frac{\Phi}{A_{eff}} = \frac{\Phi}{L\bar{w}_{eff}} = \frac{Bd\ell}{L} \\ Bd\ell = \frac{\Phi}{\bar{w}_{eff}} \end{cases} . \quad (19)$$

The central problem of calibration lies therefore in the correct evaluation of the effective width, which requires the same $B(s)$ profile during calibration and normal measurements. Failing that, even in the most favourable case, i.e., $B = \text{const.}$, the error committed on the field integral will be of order $\delta w/w$, easily in the per cent range for narrow coils. Two extreme scenarios can be discussed:

a) local measurement: $L_c \ll L_M$

assuming that longitudinal and transversal variations across the coil of B are negligible, the simple arithmetic average of the width is sufficient. The calibrated value A_{eff} can be used to obtain the average field from Eq. (19).

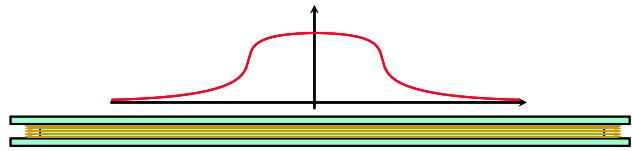


Utilization:

- local field quality (e.g. diagnostic tool for long superconducting magnets),
- integral measurement by scanning (errors tend to vanish as the coil gets shorter).

b) integral measurement: $L_c > L_M$

even perfect knowledge of $w(s)$ does not permit accurate absolute measurement of $Bd\ell$



Utilisation:

- relative measurements: e.g., tracking a reference magnet, or measurement of normalized harmonic errors.

Concerning the role of absolute vs. relative measurements, it should be emphasized that the characterization of main synchrotron magnets requires simply that all magnets be equal to each other within a given tolerance, seeing that magnets are usually powered in series and no individual corrections are possible. Absolute field maps are mandatory, on the other hand, in the case of spectrometers, or for instance if precise beam energy matching of cascading accelerators is required. Where absolute calibration is really necessary, possible alternatives include:

- Establish the shape of the $B(s)$ profile with independent methods (e.g., low absolute accuracy Hall probe measurements, or scaling from known similar magnets) and work out the effective width from Eq. (17-2)
- Use an independent measurement system, e.g., a combination of NMR/Hall probes or, more effectively, a stretched wire system.

7.2.1 End effects

Another factor to consider in order to obtain high accuracy measurements is the shape of the winding at the ends of the coil, which as we have seen in Section 4.3 must be somewhat rounded to ensure good bonding to the support. In case the coil ends stay outside of the magnetic field, including during

calibration, the issue becomes obviously irrelevant. For simplicity, we shall consider advantages and drawbacks of two extreme alternatives:

1) Straight edges (rectangular coil)

In this case, the coil length can be defined and measured precisely. This shape allows most easily the juxtaposition of several measurements to reconstruct the integral in long magnets. On the downside, very sharp corners are easily going to cut the wire during winding.

2) Semi-circular edges (racetrack coil)

An accurate measurement of the average field becomes possible, in this case, only if the field distribution across the whole coil is very nearly uniform. Since the total flux can be expressed as follows (for a coil with total length $L+w$ and end radius $w/2$):

$$\Phi = N_T \int_{-w/2}^0 2\sqrt{\frac{w^2}{4} - s^2} B(s) ds + N_T w \int_0^L B(s) ds + N_T \int_L^{w/2} 2\sqrt{\frac{w^2}{4} - (s-L)^2} B(s) ds , \quad (20)$$

easy calculation of $Bd\ell$ is possible only if the profile of $B(s)$ over the rounded regions is constant or at most linear.

7.3 Metrological considerations

We have seen that the calibration of a search coil from first principles (i.e., from the geometry alone) is not practical if high accuracy is sought, and the use of adequate reference magnets is therefore a necessity. This raises, inevitably, the question of how can we know the reference field in the first place. To frame the problem in the correct metrological context we have to consider first the relationships between the magnetic units for field and flux (T and Wb, or Tm² respectively) and the four base units for electromechanical problems (kg, m, s and A) in the International System (SI), as expressed by the diagram in Fig. 22 [21]. We find that:

1) **Magnetic flux [Φ] = Wb**

The weber is defined in terms of all four base SI units by means of a rather long chain of intermediate units:

$$Wb = V \times s = \frac{W}{A} s = \frac{J}{A} = \frac{N \times m}{A} = \frac{kg \times m^2}{A \times s^2} .$$

On account of this complexity, no primary standard exists, although the possibility of adopting as such the so-called flux quantum⁴ $\Phi_0 = \frac{h}{2e}$ is being discussed and appears very likely in the long run [22]. While historically a number of secondary flux standards have been proposed and used (see for instance Hibbert's device [23]), nowadays the availability of ppm-level voltage and time references makes it far easier to calibrate the whole integration chain in terms of purely electrical quantities.

⁴ The flux quantum is a universal constant, appealing because independent from any material properties, which covers an important role in the study of flux dynamics in superconductors. Since the flux quantum appears in the expressions for both the Josephson effect ($V = nf\Phi_0$) and the quantum Hall effect ($R = n\Phi_0/e$), it is being proposed as the basis for a future quantum redefinition of both the volt and the ohm, hence of all other SI units.

2) **Magnetic field [B] = T**

The field B (more accurately known as either magnetic induction or flux density, in order to distinguish it from the magnetic field H) can also be defined in terms of base SI units:

$$T = \frac{Wb}{m^2} = \frac{\text{kg}}{A \times s^2}.$$

Primary standards do not exist for the tesla, either. The accepted secondary standard is the NMR teslameter, an instrument based on the RF detection of nuclear resonance in a suitable sample that can routinely achieve absolute accuracies of a few ppm and resolutions of a fraction of a ppm over a wide range of measurement conditions [24]. This excellent performance is achieved by the transduction between field and frequency, which nowadays can be measured with extreme precision thanks to widely available high-precision digital oscillators. The main limitations of commercial units lie in the requirement of high field homogeneities (~ 10 ppm/mm, unless the gradient is somehow compensated) and low field rates of change ($10^{-2}/\text{s}$).

In everyday use, the availability of secondary standards of magnetic field and flux allows the practitioner to calibrate different types of field sources and instruments, which should be whenever possible cross-checked in redundant configurations to identify systematic errors and to increase the statistical confidence in the results.

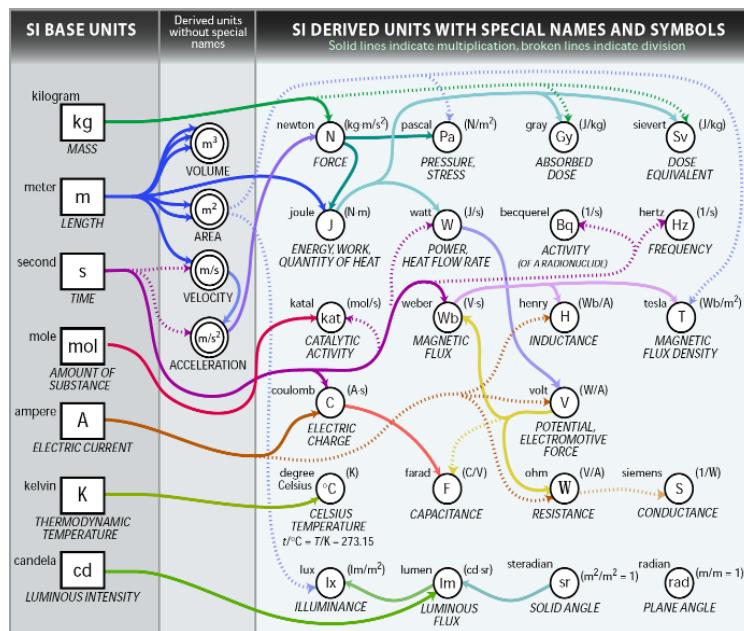


Fig. 22: Relationship diagram between base and derived SI units [from www.nist.gov]

7.3.1 Calculable field sources

While standard iron-dominated magnets are the most common instance of calibrated field sources, many issues linked to non-linear effects, ageing of the materials etc. add complications to their practical use. The so-called ‘calculable’ sources, on the other hand, attempt to realize secondary standards based as directly as possible on first principles, e.g., on well-known geometry, current and stable mechanics. Among the better-known sources of this kind we find the classical Helmholtz coils, shown in Fig. 23 [25].

In the simplest configuration, this source is made by a pair of identical and parallel circular coils separated by a distance equal to their radius R . As it can easily be proven, if both coils are fed the same current the generated field is very uniform since the lowest-order harmonic permitted by the symmetry (the sextupole) vanishes. Note that all even harmonics are also null, barring construction errors. The central field can be expressed as:

$$B_0 = \sqrt[3]{\frac{4}{5}} \frac{\mu_0 N_T I}{R} = 0.715 \frac{\mu_0 N_T I}{R} \quad (21)$$

and the relative field change $\Delta B/B_0$ is of the order of 6% at the centre of the two coils.

The main advantage of the Helmholtz arrangement is that the uniform field region is easily accessible. The concept can moreover be generalized to different configurations: for example, reversing the current in one of the coils creates a good quality quadrupole field, while a 2D version of the device, made with rectangular coils, can be built by letting the coil width be $\sqrt{3}$ times the coil spacing. More complex arrangements with three or more coils are also possible to obtain better uniformity of the desired field, or possibly arbitrary orientations of the resulting field vector.

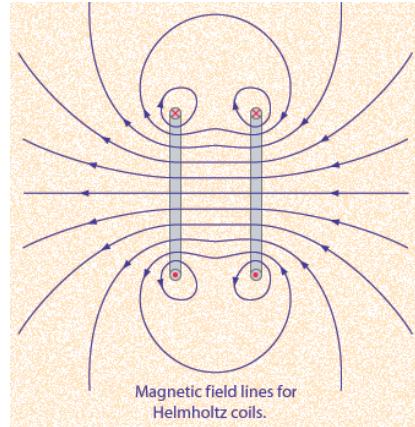


Fig. 23: Magnetic field produced by a standard Helmholtz coil configuration

8 Calibration of harmonic coils

In this section the specific techniques used to calibrate harmonic coils will be described in some detail. Referring back to Section 7.2 we shall deal, for the sake of simplicity, with the case of a short coil fully immersed in a magnetic field, i.e. we will consider all parameters averaged along the coil and ignore local fluctuations. All the methods illustrated here can of course be applied in the opposite limit, i.e., when the coil is shorter than the magnet; however, in this case the results must be understood to apply only to magnets having closely matching magnetic length and field profile w.r.t. the one used for calibration.

8.1 Surface

The total surface of a coil can be readily obtained by flipping it upside down in a uniform field of known average B over the coil area. Integration of the coil voltage gives

$$-\int_0^t V_c dt = \Phi - (-\Phi) = 2A_C \bar{B}. \quad (22)$$

This technique is fundamentally equivalent to a standard harmonic (rotating) coil measurement where only two integration points, 180° apart, are taken. The main advantage of the method lies in the

fact that an integral w.r.t. time is transformed into an integral w.r.t. angular position, thus making the end result insensitive to speed and trajectory fluctuations. A second flip back to the original position is always advisable in order to estimate and correct the error due to the integrator drift, which will respectively add and subtract to the two results. Errors due to imperfect 180° rotation grow as $(1 - \cos(\text{angular error}))$ and are therefore often negligible, typically around 10^{-4} per 1° .

Figure 24 shows the reference dipole magnet installed in the CERN coil calibration laboratory, including the apparatus for precisely controlled coil flipping. The central part of the dipole is mapped to a high precision with an NMR probe, so that the average field across coils of different length can be easily calculated. Coils thus calibrated can be used to measure the integral of long magnets by means of longitudinal scanning. In such a case, there will be an inevitable error due to the fringe field having a gradient, hence the approximation $w_{\text{eff}} = w_{\text{avg}}$ does not hold anymore. This error can be estimated (and, being systematic, eliminated) by cross-checking the integral field with an independent method, for example a stretched wire. Typical accuracies of the order of a few 10^{-4} can be routinely achieved in this case.



Fig. 24: CERN dipole apparatus for the calibration of coil surface and parallelism

8.2 Array parallelism

In many practical cases the measurement head includes an array of coils which are meant to have identical surface and orientation (see, for example, the compensated configurations described in Section 2.3, or the curved fluxmeter in Section 9). Any parallelism error between these coils is detrimental to accuracy as it may introduce spurious quadrature harmonics (e.g., a fictitious skew harmonic where the physical one is normal, and vice versa). The calibration of this error can also be carried out in a known dipole field, however, this time the coil must be flipped while *parallel* to the field.

With reference to Fig. 25, let us consider first the pair of coils 1 and 2, each having a tilt angle ε w.r.t. the support and a surface A . The respective dipole sensitivity coefficients are

$$\kappa_1^1 = A_1 (\cos \varepsilon_1 + i \sin \varepsilon_2) \approx A_1 (1 + i \varepsilon_2), \quad \kappa_1^2 \approx A_2 (1 + i \varepsilon_1) . \quad (23)$$

And the dipole coefficient of the difference signal is

$$\kappa_1^{diff} = \kappa_1^2 - \kappa_1^1 \approx \Delta A + i A \Delta \varepsilon , \quad (24)$$

so the flux difference measured by flipping $\Delta\Phi$ can be used to derive the parallelism error $\Delta\varepsilon$:

$$\Delta\Phi = 2\Re(\kappa_1^{diff} B_1 e^{in\varphi_0}) = 2B_1(\Delta A \cos\varphi_0 - A \Delta \varepsilon \sin\varphi_0) \Rightarrow \Delta \varepsilon = \frac{\Delta\Phi}{2AB_1} . \quad (25)$$

If this angular difference exceeds the prescribed tolerance (usually a few milliradian), one should shim or file the support to compensate, and then iterate. By repeating the procedure it is clear that all coils can be made parallel to one selected as the master; however, the tilt of the master itself remains unknown. The additional procedure for the calibration of the absolute field direction is described in the next section.

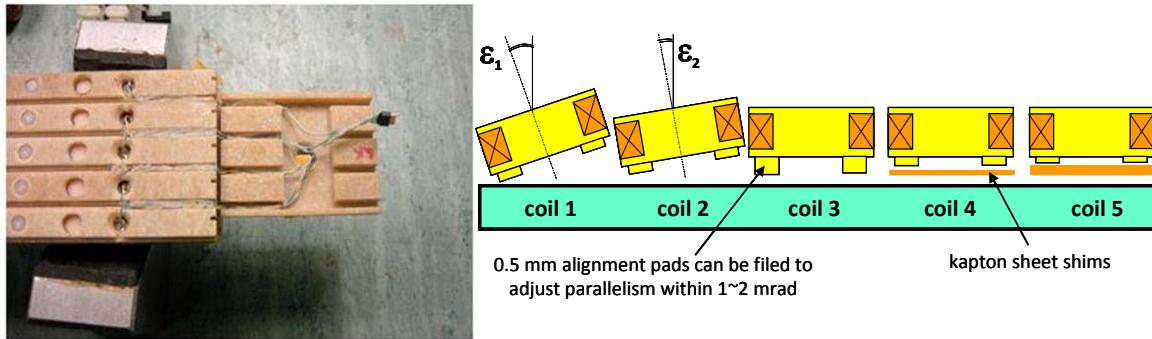


Fig. 25: A 5-coil array fluxmeter for quadrupole-compensated harmonic measurements, including a spare coil (left). Schematic representation of parallelism calibration procedure (right).

8.3 Field direction

The calibration of field direction measurement for a harmonic coil has the purpose of measuring the total angular offset $\Delta\alpha$ between the measuring coil and the angle corresponding to the start of coil voltage integration, usually marked by an index pulse from an angular encoder. The total offset shall include all contributions from different error sources, e.g., the angle between a coil and its support, the angle between support and encoder etc.

The simplest and most accurate method involves repeating a full harmonic measurement in a dipole magnet while reversing the relative orientation of coil and field, i.e., flipping either one by 180° around an axis normal to the coil rotation axis (see Fig. 26). Since the unknown offset respectively adds and subtracts to the measurement in the two different configurations, both the offset and the actual field direction α can be computed:

$$\begin{cases} \alpha_{meas}^1 = \alpha + \Delta\alpha \\ \alpha_{meas}^2 = \pi - \alpha + \Delta\alpha \end{cases} \Rightarrow \begin{cases} \alpha = \frac{\pi}{2} + \frac{\alpha_{meas}^1 - \alpha_{meas}^2}{2} \\ \Delta\alpha = -\frac{\pi}{2} + \frac{\alpha_{meas}^1 + \alpha_{meas}^2}{2} \end{cases} \quad (26)$$

When the flipping is not possible *in situ*, $\Delta\alpha$ may of course be measured in a convenient reference magnet and stored for later use. While the case of a dipole is somewhat more intuitive, the method can be extended to multipoles of every order. However, as the correct handling of the related inverse trigonometry in the analysis software is potentially messy, it is strongly suggested to double-check procedures and algorithms by applying known rotations to the magnet (if possible over 360°) in order to verify absolute values and signs of the measured rotations. Accuracies in the range of a fraction of a milliradian are routinely achieved with this method.

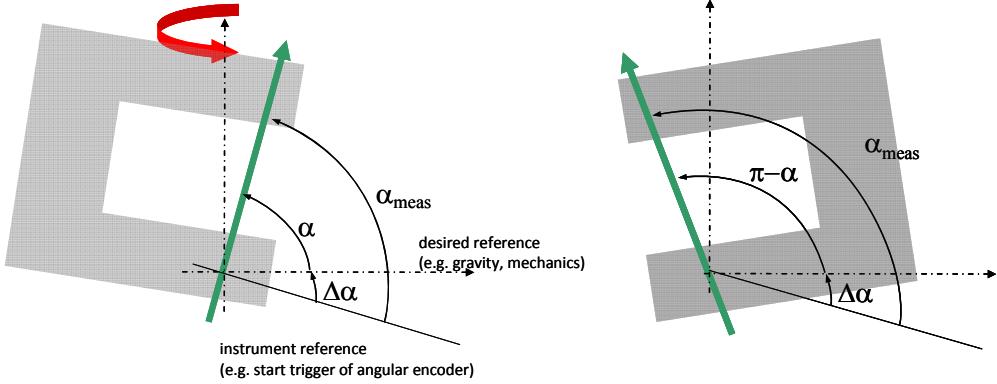


Fig. 26: Calibration of field direction in a reference dipole. The second configuration (right) is obtained by flipping either the instrument or the magnet by 180° around a vertical axis.

8.4 Rotation radius

As seen in Section 7.1, the measurement of all harmonics above dipole relies on accurate knowledge of the average coil rotation radius R_0 . The calibration of this quantity is particularly difficult, not least because the mechanical position of the rotation axis must not change between calibration and regular use, as might well be the case if, for instance, different ball bearings are used.

We shall outline briefly a calibration procedure based on precisely controlled coil displacements inside a quadrupole magnet, as shown in Fig. 27. We will assume the total coil surface A_c to be known and, without loss of generality, an horizontal initial offset between rotation axis and magnetic centre. First, the quadrupole strength can be inferred from the flux change associated to a translation Δx :

$$B_2 = \frac{r_{ref}}{\Delta x} \frac{\Phi(0, x_0 + \Delta x) - \Phi(0, x_0)}{A_c} . \quad (27)$$

Note that, even in the case of a dedicated reference magnet, it is advisable to re-measure the strength upon every calibration campaign in order to avoid errors due to drifts (magnet and electronics), temperature, power supply instability, etc.

Assuming the quadrupole to be pure, the flux seen by the coil as a function of azimuth ϑ can be expressed as

$$\Phi(\vartheta) = \Re(K_1 B_1 e^{in\vartheta} + K_2 B_2 e^{2in\vartheta}) = \frac{A_c B_2}{r_{ref}} \left(x_0 \cos \vartheta + \frac{1}{2} R_0 \cos 2\vartheta \right) , \quad (28)$$

where the B_1 term is generated by feed-down of the quadrupole. Among the many possible and essentially equivalent choices, we shall consider two coil rotations between $\vartheta = 0, \pi/2$ and π to obtain

$$\begin{cases} \Phi(0) = \frac{A_c B_2}{r_{ref}} \left(x_0 + \frac{1}{2} R_0 \right) \\ \Phi(\frac{\pi}{2}) = \frac{A_c B_2}{r_{ref}} \left(-\frac{1}{2} R_0 \right) \Rightarrow x_0 = r_{ref} \frac{\Phi(0) - \Phi(\pi)}{A_c B_2}, R_0 = r_{ref} \frac{\Phi(0) - \Phi(\frac{\pi}{2})}{A_c B_2} - x_0 . \\ \Phi(\pi) = \frac{A_c B_2}{r_{ref}} \left(-x_0 + \frac{1}{2} R_0 \right) \end{cases} \quad (29)$$

Analogous expressions can easily be derived for the y coordinate in the general case. For better results, it is suggested to centre the coil in an iterative fashion making use of the measured value of the

offset. Once centred, the coil can be made to rotate over a full turn and four 90° -spaced flux increments can then be acquired. Since these should all be equal in absolute value, the average can be taken as the wanted result and the differences will give an estimate of systematic and random errors (as always, redundant measurements are the key to improved accuracy).

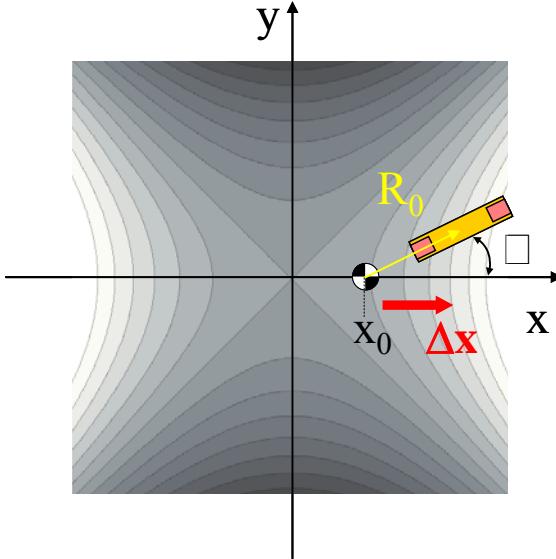


Fig. 27: Schematic representation of the calibration procedure for the average rotation radius of a harmonic coil (R_0) in a quadrupole magnet

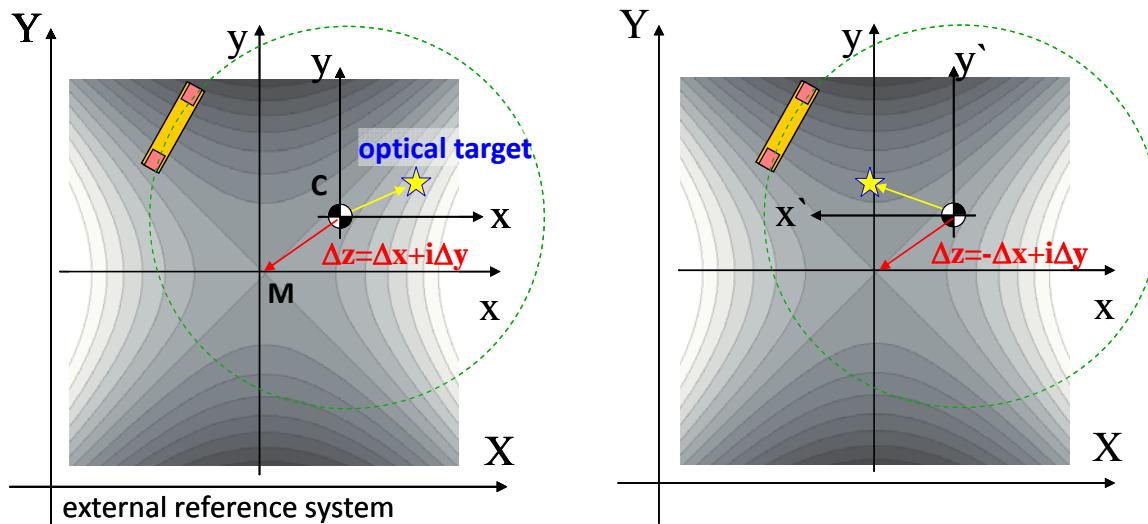


Fig. 28: Schematic representation of the calibration procedure for the magnetic centre using a harmonic coil in a quadrupole magnet. The second configuration (right) is obtained by flipping the coil by 180° around the vertical axis y . The optical target marked by the star is rigidly connected to the rotating coil.

8.5 Magnetic centre

Among the various existing methods for measuring the magnetic centre (or axis) of a multipole, the harmonic coil can provide excellent results as long as the position of the rotation axis is measured w.r.t. an external coordinate system, from which the centre can be then transferred to the magnet fiducial reference system. One of the possibilities consists in fixing an optical target on the rotation axis itself, so that it can be imaged by an external instrument such as a laser tracker or a theodolite.

The calibration of such a target can easily be accomplished in a quadrupole magnet, as shown in Fig. 28. We shall consider two configurations (1,2) in which the rotating coil probe is flipped by 180° around the vertical (y) axis. Let (x,y) be the reference axis of the coil, centred on its rotation axis C; (X,Y) the external coordinate system in which the optical target T is measured; $\Delta z = \Delta x + i\Delta y$ the complex offset of the magnetic centre M w.r.t. C, as obtained by feed-down from each measurement. The horizontal components of the centre position (X_M) and optical target offset (x_T) can be obtained as follows:

$$\begin{cases} X_{T1} = X_M - \Delta x_1 + x_T \\ X_{T2} = X_M + \Delta x_2 - x_T \end{cases} \Rightarrow \begin{cases} X_M = \frac{X_{T1} + X_{T2}}{2} + \frac{\Delta x_1 - \Delta x_2}{2} \\ x_T = \frac{X_{T1} - X_{T2}}{2} + \frac{\Delta x_1 + \Delta x_2}{2} \end{cases} . \quad (30)$$

The vertical component can be worked out in a similar manner, e.g., by means of a rotation of the coil head around the x axis. Analogous schemes can be devised for more complex cases, where a higher number of unknowns (including for example various mechanical imperfections) of the magnet and of the probe can be found by measuring more flipped configurations.

9 Calibration of fixed coils

In this final section we shall discuss the case of fixed coils, where the flux change is determined by the time variation of the external field. Since coil sensitivity factors depend only on the geometry and not on the mechanism of flux change, all results valid for harmonic coils still apply. Considering for simplicity the case of a dipole, integration of the output voltage of a fixed coil while the field is pulsed will give

$$-\frac{1}{W_{eff}} \int_0^t V_c dt = \int_0^L B(s, I(t)) ds - \int_0^L B(s, I_0) ds = Bd\ell - Bd\ell_0 . \quad (31)$$

The dynamic nature of the problem entails, however, some important issues related to the reproducibility of the magnetic field in case of cyclical powering. This is affected by a number of factors:

- stability of the power supply, including the accuracy of current measurements;
- hysteresis effects in the iron or in the superconductor, as the case may be. These depend in general upon powering history, temperature or even current ripple (this may shift the magnetic state back and forth along minor hysteresis loops);
- eddy currents and other dynamic effects, which introduce additional time and ramp-rate dependencies (e.g., magnetic aftereffect in the iron [26], decay and snapback in superconducting magnets [27]).

Repeated pre-cycling up to the maximum working currents can help stabilize the magnet onto a reproducible state by way of letting sufficient time for all transient phenomena to die out (a few cycles are often found to be sufficient). Further improvement of stability can be achieved by minimizing other sources of uncertainty, e.g., by starting and stopping the integration in a current range where the power supply is more stable (i.e., far from zero for two-quadrant converters); or choosing an

integration time which is an integer multiple of the dominant perturbation, e.g., the 50 Hz from the mains or the ripple of the power supply.

From Eq. (31), we see that a fixed-coil measurement provides only relative flux changes and not absolute values. Nevertheless, such a method is adequate to obtain information on integral multipole field errors, which are by definition a relative quantity (provided that the field distribution at $I = I_0$ is either known or can be safely ignored). We shall now consider the calibration procedure that applies in such a case, taking as an example the curved-coil fluxmeter built for the main dipoles of a hadron therapy synchrotron [28].

The fluxmeter in question, shown in Fig 29, is a flat array of 13 nearly identical and parallel coils lying on the plane where field uniformity has to be measured. Despite the effort expended in winding and sorting, the total surfaces of these coils differ from each other by a few 10^{-3} , which is larger than the field errors to be measured. This random effect is thought to be mainly due to the deformation of the coils, initially straight and then bent to conform to the curved beam paths through the dipole, and is exacerbated by mechanical handling and temperature changes.

The most efficient way to calibrate out these differences consists in taking one additional coil as a common reference, *provided that* this reference is used to measure exactly the same field as the coil it is being compared to. An elegant way to implement this solution is illustrated in Fig. 30: the reference coil and the array are first placed symmetrically w.r.t. to the mid-plane of the dipole, then the reference is moved sequentially on top of all coils in the array and a measurement in bucked configuration (e.g., series opposition) is taken to obtain

$$\frac{\Phi^j}{\bar{w}_{eff}^j} = \frac{\Phi^{ref}}{\bar{w}_{eff}^{ref}} = -Bd\ell^j + Bd\ell_0^j \quad \Rightarrow \quad \bar{w}_{eff}^j = \bar{w}_{eff}^{ref} \underbrace{\left(1 + \frac{\Delta\Phi^j}{\Phi^{ref}}\right)}_{k_j}, \quad (32)$$

where Φ^j and Φ^{ref} are the fluxes measured by the reference and j -th array coils respectively during calibration, and $\Delta\Phi^j$ is their difference. When measuring the field uniformity, the relative field differences w.r.t. the central coil (index 0) can therefore be expressed compactly by means of the calibration coefficients k_j as

$$\frac{\Delta Bd\ell^j}{Bd\ell^0} = \frac{\Phi^j}{\Phi^0} \frac{\bar{w}_{eff}^0}{\bar{w}_{eff}^j} - 1 = \frac{\Phi^j}{\Phi^0} \frac{k_0}{k_j} - 1. \quad (33)$$

In practical use, during series tests this calibration procedure was found to be necessary several times for each magnet and at least two to three times each month, in order to keep the errors due to field profile differences, temperature drifts, and mechanical instabilities down to an acceptable level.

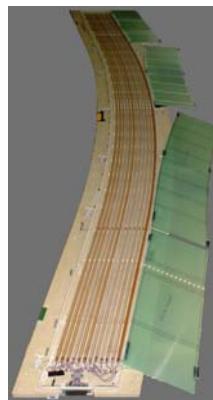


Fig. 29: 13-coil array fluxmeter for pulsed-mode measurements of main CNAO synchrotron dipoles

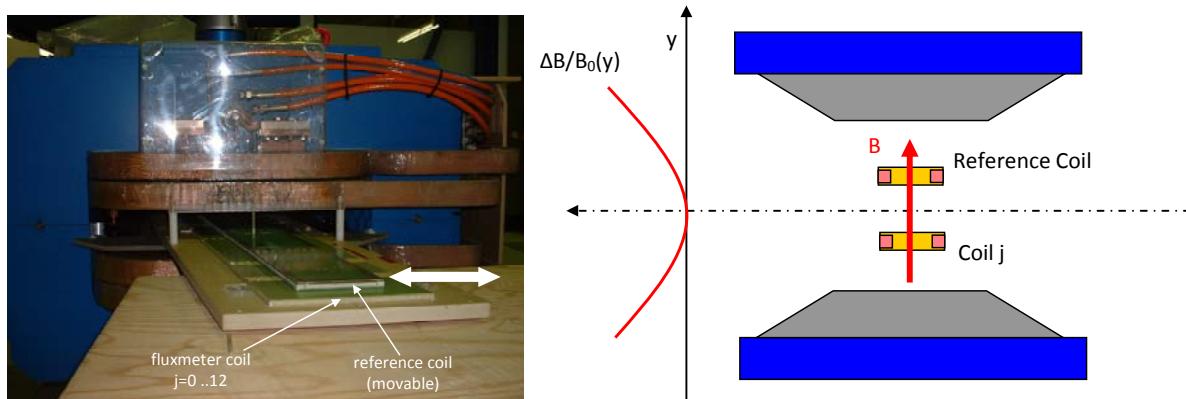


Fig. 30: Relative surface calibration procedure for the pulsed-mode fluxmeter. The fluxmeter is inserted in the magnet to be measured alongside an additional, movable reference coil (left). The reference and the calibrated coil are positioned symmetrically in the dipole field (right).

This method can obviously work only if the magnetic field is top–bottom symmetric, i.e., if only the allowed harmonics $B_1, B_3, B_5 \dots$ are present. Failing that, one could in principle repeat each measurement twice by swapping reference and calibrated coil in the same physical position and then subtracting the results, however, in this way the cycle reproducibility error would have to be taken into account explicitly instead of being automatically cancelled out.

It is important to stress that this calibration procedure cannot be carried out merely by taking one of the coils in the array as the reference, because in this case the measured flux differences would depend upon the unknown differences of field profile, hence the equation system would be underdetermined. Another conceivable alternative consists in the absolute calibration of each coil of the array, in which case the calculation of the desired field errors becomes trivial. Unfortunately, this approach is hardly practical due to the difficulties inherent in the absolute calibration which, as we have seen in Section 7.2, must be carried out in a magnet with the same longitudinal field profile that has been measured with an independent, absolute method. In the author's experience all ordinary options⁵ in this respect are either not accurate enough or not cost-effective.

10 Conclusions

It is hoped that this lecture has provided a comprehensive (but not exhaustive) overview of the different techniques that can be used to make and calibrate search coils, especially harmonic coils, in order to carry out high-accuracy measurements. It would be tempting to draw up a long final list of recommendations, such as to remember the fundamental importance of a sound mechanical design, or take the utmost care to control the effects of temperature, which is well known as the biggest enemy of precision; however, it is felt that good practice really boils down to common sense and a sound knowledge of physical factors affecting the materials and the electronic equipment. The final message is that, as experimentalists, we should never forget to put beautiful ideas and equations to the test in a variety of environments and conditions, because real-world tests often reserve unexpected surprises.

⁵ These might include stretched wire techniques (hardly sensitive enough in pulsed mode, and out of the question in the case of a curved magnet); a combination of maps obtained with NMR (magnet centre) and Hall (magnet ends) probes (this would suffer from both the inherent lower accuracy of Hall plates and the error due to the extrapolation from steady-state to dynamic conditions); a high-resolution scan with small coils (probably accurate enough but very time-consuming).

Acknowledgements

First and foremost I wish to dedicate this lecture to the late Jacques Billan, who was for three decades the person in charge of CERN’s coil manufacturing and calibration laboratory and to whom many of the solutions described in this paper, not to mention much of my own education, are due.

My gratitude goes to present and past team leaders L. Bottura, D. Cornuet, P. Sievers, and L. Walckiers for their guidance and their insight.

I would like to acknowledge gratefully the hard and professional work of present and past magnetic measurement team members, including: R. Beltron Mercadillo, G. Busetta, R. Camus, R. Chritin, D. Cote, G. Deferne, O. Dunkel, J. Dutour, F. Fischer, L. Gaborit, P. Galbraith, J. Garcia Perez, D. Giloteaux, P. Leclere, A. Musso, A. Ozturk, S. Pauletta, S. Sanfilippo, N. Smirnov, S. Stokic, L. Vuffray.

Many thanks also to J. Di Marco (FNAL) and A. Jain (BNL) for valuable discussions and suggestions.

References

- [1] S. Sanfilippo, Hall probes: physics and applications to magnetometry, *this School*.
- [2] L. Bottura, Overview of measurement methods, *this School*.
- [3] CERN Accelerator School: Magnetic Measurement and Alignment, Montreux, Switzerland, 1992, S. Turner (Ed.), CERN 92-05.
- [4] CERN Accelerator School: Measurement and Alignment of Accelerator and Detector Magnets, Anacapri, Italy, 1997, S. Turner (Ed.), CERN 98-05.
- [5] S. Tumanski, Induction coil sensors—a review, *Meas. Sci. Technol.* **18** (2007) R31–R46.
- [6] J. Di Marco *et al.*, Field alignment of quadrupole magnets for the LHC interaction regions, *IEEE Trans. Appl. Supercond.* **10** (2000) 127–130.
- [7] S. Humphries, Principles of Charged Particle Acceleration (Wiley, New York, 1986) [Digital edition 1999].
- [8] W. Scandale *et al.*, Production follow-up of the LHC main dipoles through magnetic measurements at room temperature, *IEEE Trans. Appl. Supercond.* **14** (2004) 173–176.
- [9] A. Siemko *et al.*, Quench localization and current redistribution after quench in superconducting dipole magnets wound with Rutherford-type cables, Applied Superconductivity Conference: ASC 96, pt.1-3, Pittsburgh, PA, USA, 1996, pp.179–182.
- [10] L. Walckiers, Magnetic measurement with coils and wires, *this School*.
- [11] Z. Wolf *et al.*, Short description of the SLAC rotating coil system used to measure the CIEMAT-made prototype ILC linac quadrupole, International Linear Collider Workshop, Chicago, 16–20 November 2008.
- [12] L. Deniau, Coil calibration correction factors for rectangular windings, CERN MTA-IN-98-026.
- [13] ASTM Standard B 258-02, Standard specification for standard nominal diameters and cross-sectional areas of AWG sizes of solid round wires used as electrical conductors, ASTM International (2002).
- [14] International Standard IEC 60228, Conductors of insulated cables (2004).
- [15] International Standard CEI/IEC 60317, Specifications for particular types of winding (2010).

- [16] C. M. Spencer *et al.*, A rotating coil apparatus with sub-micrometer magnetic center measurement stability, IEEE Trans. Appl. Supercond. **16** (2006) 1334–1337.
- [17] J. Di Marco, Harmonics measurements with radially bucked tangential probes, FNAL TD-03-044 (2003).
- [18] Y. Iwashita, Modification and measurement of the adjustable permanent magnet quadrupole for the final focus in a linear collider, Particle Accelerator Conference, PAC2007, Albuquerque, NM, USA, 2007, pp. 2719–2721.
- [19] P. Sievers *et al.*, Twin rotating coils for cold magnetic measurements of 15 m long LHC dipoles, IEEE Trans. Appl. Supercond. **10** (2000) 1422–1426.
- [20] A. Wolski, Maxwell's equations for magnets, *this School*.
- [21] B. N. Taylor *et al.* (editors), The International System of Units (SI), NIST Special Publication 330 (2008).
- [22] J. Flowers *et al.*, The route to atomic and quantum standards, Science **306** (2004) 1324.
- [23] H. E. J. G. Du Bois, The Magnetic Circuit in Theory and Practice (Longmans, London, 1896) [translation of Magnetische Kreise, deren Theorie und Anwendung].
- [24] K. Borer, The nuclear magnetic resonance system for the CERN muon storage ring, Nucl. Instrum. Methods **143** (1977) 203–218.
- [25] S. R. Trout, Use of Helmholtz coils for magnetic measurements, IEEE Trans. Magn. **24** (1988) 2108–2111.
- [26] R. Grössinger *et al.*, The influence of the magnetic viscosity on pulsed field measurements, J. Optoelectron. Adv. Mat. **6** (2004) 557–563.
- [27] G. Ambrosio *et al.*, A scaling law for the snapback in superconducting accelerator magnets, IEEE Trans. Appl. Supercond. **15** (2005) 1217–1220.
- [28] D. Cornuet, Magnetic measurement system for the dipoles of the Italian Hadron Therapy Center (CNAO), International Workshop on Magnetic Measurements IMMW14, Ferney-Voltaire, France, 26–29 September 2005.

Hall probes: physics and application to magnetometry

S. Sanfilippo

Paul Scherrer Institut, Villigen, Switzerland

"What I do teaches me what I am looking for"

Pierre Soulages, 1953, retrospective of the 'painter of black and light', 14 October 2009 – 8 March 2010 (Pompidou Centre, Paris)

Abstract

This lecture aims to present an overview of the properties of Hall effect devices. Descriptions of the Hall phenomenon, a review of the Hall effect device characteristics and of the various types of probes are presented. Particular attention is paid to the recent development of three-axis sensors and the related techniques to cancel the offsets and the planar Hall effect. The lecture introduces the delicate problem of the calibration of a three-dimensional sensor and ends with a section devoted to magnetic measurements in conventional beam line magnets and undulators.

1 Introduction

The measurement of magnetic fields plays a key role in various areas. The mastering of the magnetic field intensity allows the construction of alternators and of electrical motors, the storage of information on magnetic tapes or hard disks, and the study of phenomena in condensed matter physics like the flux distribution in superconductors. Magnetic fields of very strong intensity are used in particle accelerators or in Tokomaks to guide and focus beams with high energy and make them collide. Field measurements are also omnipresent in the domain of medicine to detect heart and brain activity. The Hall effect is an ideal magnetic field sensing technology. Discovered in 1879 by a professor of the Johns Hopkins University in the USA, E. H. Hall [1], its application remained confined to laboratory experiments until the 1950s as a tool for measuring the magnetic field or exploring the basic electronic properties of a solid. With the mass production of semiconductors and the development of the associated electronics, it became progressively feasible to combine a Hall effect sensing element with electronics in a single integrated circuit. This changed the end objective of the Hall effect devices. From the simple measurement of the magnetic field, they became a key element in the sensing of pressure, current, position or temperature via magnetic field detection. Hall sensors can be found nowadays in many products ranging from computers to automobiles, aircraft and medical equipment. An overview of the Hall effect devices used as magnetic sensors can be found, for example, in the bibliographic items [B1]–[B6].

The object of this lecture is to describe the principle of the Hall effect, the basic characteristics of a Hall device, the advantages and the limitations of this technique in the field of magnetic measurements and magnetic field mapping. Obviously, given the vastness of the domain of application, this cannot be an exhaustive review of Hall probe measurements. The progress on the state of the art and on the domain of application is a typical topic of discussions in conferences [2], workshops [3] and international journals [4]. We will focus on the case of a three dimensional axis Hall device with an overview of the recent progress that has been accomplished to minimize the planar Hall effect and the offset. The delicate problem of the calibration of a three dimensional Hall device is briefly introduced. The course ends with two examples of magnetic measurements with a Hall probe: in beam line magnets and in undulators.

2 The Hall generator

2.1 The Hall generator

A Hall generator consists of a sub-millimetre (semi)conductor element, the active part, typically ranging from 0.01 mm to 0.1 mm which is sensitive to magnetic field. It is equipped with two current electrodes (biasing contacts) to conduct the current I into the conducting plate and with two voltage electrodes to measure the Hall voltage V_H (sensing contact). The Hall element is affixed to a ceramic substrate which provides improved mechanical support and thermal stability and encapsulated. Figure 1 shows a typical Hall generator construction. In its basic form, a current supplied by a highly stable AC or DC current source feeds two terminals and the Hall voltage is read through the other two terminals by a high impedance voltmeter or amplified and conditioned by acquisition electronics. This device, combined with associated electronics, is used to measure the magnetic field based on the Hall effect and is commonly called Hall sensor. It is part of the category of transducers that convert a non-electrical energy (the magnetic field) into an electrical one (the Hall voltage). In this lecture the following definitions will be used [5]: the denomination “Hall generator” or “Hall device” will refer to a four terminal device containing a Hall element and exploiting the Hall effect. The Hall generator mounted on a holder will be called “Hall probe”. The term “Hall plate” reflects a conventional form of the Hall device. The denomination “Hall sensor” will be used to stress the application of a Hall generator as magnetic sensor. The combination of a Hall device with some electronic circuitry is called “Integrated Hall magnetic sensor” or “IC Hall sensor”.

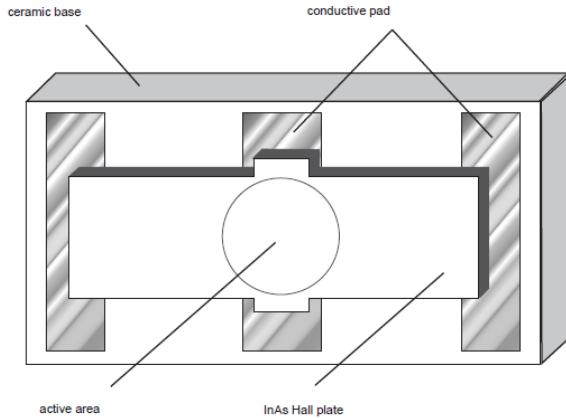


Fig. 1: Typical Hall generator construction (picture from Ref. [6])

2.2 The Hall effect (simplified approach)

The Hall effect is a galvanomagnetic phenomenon caused by the current flow deviation in presence of a magnetic field component perpendicular to the flow direction due to the Lorentz force. The response of this phenomenon, i.e., the voltage produced, was found to be proportional to the magnetic field intensity B . While the effect was originally detected in a thin metallic (gold) foil, modern Hall elements are nowadays mostly produced in semiconductors such as silicon (Si), indium arsenide (InAs), gallium arsenide (GaAs) or indium antimonide (InSb). This section describes the Hall effect in its simplest and classic form, where a long current-carrying strip is exposed to a magnetic field as shown in Fig. 2. The Hall generator consists of a thin semiconductor plate of dimensions (l, w, t) , equipped with four contacts. The control current I flows through the biasing contacts CC while the sense contacts, S1S2, are used for the measurement of the Hall voltage. The magnetic field \mathbf{B} is normal to the plate or making an angle γ (generally small) with respect to the normal of the strip. We assume that:

NB: The physical quantities in bold refer to vectors.

- the material is strongly extrinsic n-type, i.e., the majority carriers are electrons (particle charge e);
- the role of the holes is neglected;
- the velocity is the same for all the electrons (smooth drift approximation);
- the effect of the thermal agitation is neglected;
- the material is plate like and isotropic;
- the strip is infinitely long;
- the Hall contacts are extremely small.

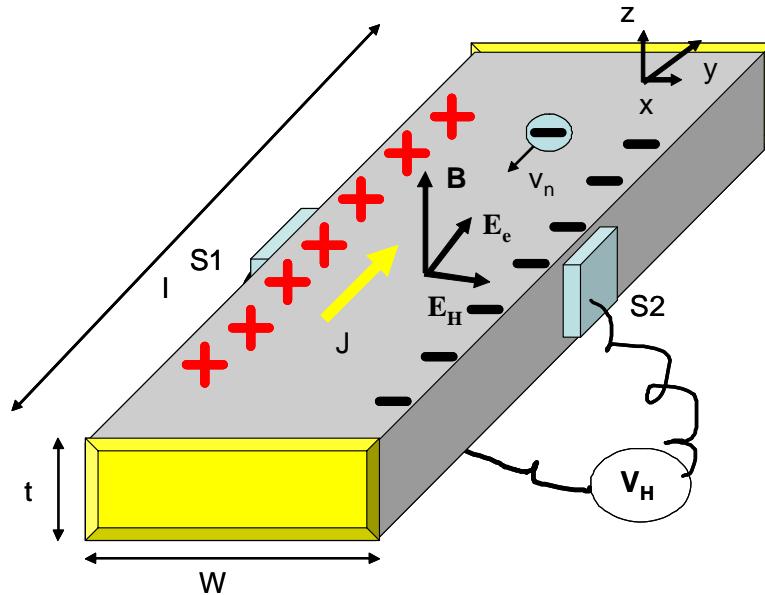


Fig. 2: Sketch of a Hall generator. An applied magnetic field \mathbf{B} (normal to the plate in that case) is applied and a current I (current density J) flows in the y direction through the CC contacts. The output voltage is measured through the S_1S_2 contacts

When no magnetic field is applied, the electrons move in the longitudinal direction (y direction) parallel to the longitudinal external electrical field \mathbf{E}_e . As soon as a magnetic field is applied, the trajectory is deflected and the electrons with a velocity \mathbf{v}_n move under the action of the Lorentz force \mathbf{F}_L in response to the magnetic field \mathbf{B} and the external electric field \mathbf{E}_e . \mathbf{F}_L is given by

$$\mathbf{F}_L = e (\mathbf{E}_e + \mathbf{v}_n \times \mathbf{B}) \quad e = -q \text{ for electrons .} \quad (1)$$

The electrons are pressed by the magnetic force towards one edge of the strip. Consequently, the concentration of the electrons at the other edge is decreasing. To re-establish the charge repartition, an electric field \mathbf{E}_H is created between the two edges and moves the carriers so as to decrease the excess of charges. Equilibrium is reached when the magnetic force pushing the carriers aside corresponds to this reacting electric force trying to push them back to the middle. The equilibrium is expressed by

$$\mathbf{E}_H + \mathbf{v}_n \times \mathbf{B} = 0 . \quad (2)$$

At this point, the electrons will again move in the longitudinal direction. This leads to the definition of the Hall field \mathbf{E}_H that is only a function of the velocity of the electrons and of the applied magnetic field. \mathbf{E}_H is given by

$$\mathbf{E}_H = -\mathbf{v}_n \times \mathbf{B} . \quad (3)$$

A measurable transverse voltage V_H called the Hall voltage is produced between the two edges of the strip. This represents the most tangible effect associated with the Hall effect. We will express the Hall voltage as a function of the current and the magnetic field.

The velocity of the electron is related to its mobility μ_n and the external electric field by

$$\mathbf{v}_n = -\mu_n \mathbf{E}_e . \quad (4)$$

The current density J is related to the electric field by Ohm's law:

$$\mathbf{J} = \sigma \mathbf{E}_e . \quad (5)$$

The material conductivity σ is expressed by

$$\sigma = n q \mu_n , \text{ where } n \text{ is the density of electrons} . \quad (6)$$

The associated current density for electrons is therefore given by

$$\mathbf{J} = n q \mu_n \mathbf{E}_e = -n q \mathbf{v}_n . \quad (7)$$

The Hall voltage V_H appearing across the two sense electrodes (S1S2) can be calculated along the width of the strip by

$$V_H = \int_{S1S2} \vec{E}_H \cdot d\vec{l} = -\frac{J}{nq} B w . \quad (8)$$

J is related to the biasing current I by $J = I / w t$ (see geometry in Fig. 2) .

For a homogeneous, isotropic, rectangular and infinitely long Hall generator, the Hall voltage can be related to the current and the magnetic field by combining Eqs. (8) and (9):

$$V_H = \frac{R_H}{t} I B . \quad (10)$$

R_H is named the Hall coefficient and is given **for the electrons** by

$$R_H = -I/n q = -\mu_n / \sigma . \quad (11)$$

More generally, if \mathbf{B} is slightly tilted with respect to the normal, making an angle γ with respect to the normal of the strip, the Hall voltage V_H is given by

$$V_H = \frac{R_H}{t} I B \cos \gamma . \quad (12)$$

R_H is the parameter that gauges the magnitude of the Hall effect.

Remark 1: The Hall voltage V_H depends linearly on

- the charge carrier velocity;
- the magnetic field intensity B ;
- the distance between the sense contacts.

Remark 2: The Hall constant R_H is related to the density of the electrons inside the support material and the mobility of the material [Eq. (11)]. This leads to one direct application of the Hall effect. The measurement of the Hall coefficient of materials with a known conductivity will indicate directly the mobility of the carriers responsible for the conductance.

- The minus sign above is obtained for electrons, i.e., negative charges.
- If positively charged carriers would be involved, the sign of the Hall constant would be positive.

Remark 3: In addition to the carrier mobility and the geometry (see Section 2.2), the sample thickness is of crucial importance. It determines the amplitude of the output signal following the relationship

$$V_H \propto 1/t. \quad (13)$$

Remark 4: Hall angle Θ_H

The total electrical field \mathbf{E} is the sum of the external field \mathbf{E}_e and of the Hall field \mathbf{E}_H which are not parallel. The Hall angle Θ_H is defined as the angle between \mathbf{E}_e and the total electric field \mathbf{E} (see Fig. 3). The angle value depends on magnetic field intensity and on the mobility of the charge carrier according to Eq. (14). The sign of the Hall angle coincides with the sign of the charge carrier.

$$\operatorname{tg} \Theta_H = \frac{|E_H|}{|E_e|} = -\mu_n B \quad (14)$$

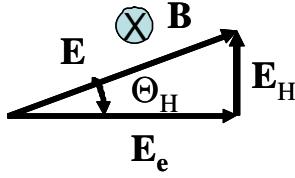


Fig. 3: Definition of the Hall angle in the case of a n-type semiconductor

2.3 Hall element geometry and choice of the material

2.3.1 Geometry

For real Hall generators, the actual dimensions and the finite size of the Hall contacts have to be taken into account. These two dependences can be regrouped in a dimensionless geometrical correction factor G ranging from 0 to 1 that multiplies the Hall voltage formula in Eq. (10).

$$V_H = G (l/w, \Theta_H, B) R_H I B / t. \quad (15)$$

It can be demonstrated that this factor depends on the geometry, on the dimensions of the active part, on the Hall angle Θ_H , and on the magnetic field. This factor determines the reduction of the Hall voltage due to the imperfect bias current confinement. It depends mainly on the ratio l/w . Figure 4 shows some typical shapes of Hall generators. The bridge shape (a) is a good approximation of a long Hall device and allows relatively large contacts. The parasitic effects from the contact resistance and the heating can therefore be minimized. The cruciform (b) and the Van der Paw (d) shapes are of particular interest. They offer the advantages that a high geometrical correction factor $G \sim 1$ can be achieved, the geometry is simple and compact and the shape is invariant for a rotation through $\pi/2$. The current and the sense contacts could be switched without changing the global symmetry. This

four-fold symmetry is favourable for the application of offset voltage compensation techniques like the spinning current technique (see Section 4.4). This geometry also allows a better definition of the active part centre.

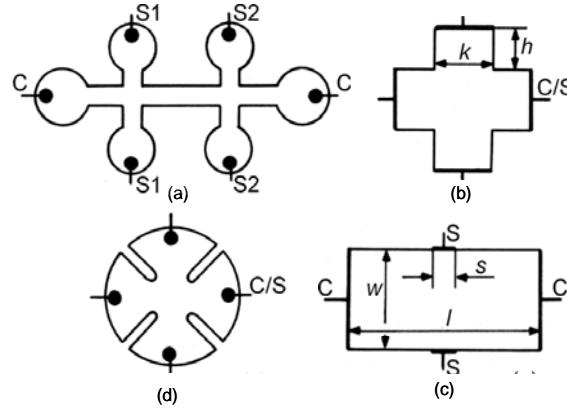


Fig. 4: Various shapes of Hall elements. C are the current contacts, S the sense contacts and C/S indicates that the two types of contacts are interchangeable [from Ref. B1, p.194]

2.3.2 Material

The choice of the proper material for the Hall element is of crucial importance. From Eq. (11), one can conclude that the Hall effect will be favoured in materials with high mobility and also low conductivity. Metals show low mobility and high conductivity and are therefore not a good choice. The sensors are usually prepared from n-type semiconductors where the dominant charge carriers are the electrons having much higher mobility than holes. Suitable candidates are single elements like Si and III-V compounds like InSb, InAs and GaAs. The III-V compounds combine high carrier mobility and reasonable value of conductivity. Silicon has moderate electron mobility but is compatible with the integrated circuit technology. This makes this material very attractive for the realization of Hall sensor chips. Table 1 gives the values of the semiconducting gap E_g and the mobility at room temperature of the various semiconductors used for Hall generators. R_H is also calculated for a given doping density.

Table 1: Gap and mobility of semiconductors at 300 K used to build a Hall plate (from Ref. [7]). R_H is calculated for a given level of doping

Material	E_g [eV]	μ_n [$\text{cm}^2\text{V}^{-1}\text{s}^{-1}$]	n [cm^{-3}]	R_H [cm^3C^{-1}]
Si	1.12	15 00	$2.5 \cdot 10^{15}$	2.5×10^3
InSb	0.17	80 000	$9 \cdot 10^{16}$	70
InAs	0.36	33 000	$5 \cdot 10^{16}$	125
GaAs	1.42	85 00	$1.45 \cdot 10^{15}$	2.1×10^3

2.4 Basic characteristics and parasitic effects on Hall probes

The quality of a Hall probe can be appreciated with the following criteria, see Refs. [B7, B8, B9]:

- the sensitivity to the magnetic field;
- the low zero-field offset and its small drift in time;
- low temperature-dependence coefficient;
- the small linearity error of the Hall voltage (% of non-linearity);
- the minimization of the thermal, electrical and frequency noise;

- very low power consumption;
- possibility to operate in a large temperature range (low temperature applications);
- and also ... the cost.

In the following, some of those basic features of a Hall sensor will be developed.

2.4.1 Magnetosensitivity

The magnetosensitivity (we will use the term of sensitivity) expresses the response of the output voltage to a magnetic field. The (absolute) sensitivity S can be formulated by

$$S = \frac{V_H}{B_{\perp}} \quad [\text{V/T}] \quad B_{\perp} \text{ represents the normal component of } \mathbf{B} \text{ to the Hall plate. (16)}$$

Two other figures of merit can also be defined:

- the supply current-related sensitivity S_I in Volts per unit field, per unit of bias current I_{HP} ;
- the supply voltage-related sensitivity S_V in volts per unit field, per unit of bias voltage mostly employed when the additional input is fed by a supply voltage V_s

$$S_I = \frac{S}{I_{HP}} \quad [\text{V/AT}] \quad \text{and} \quad S_V = \frac{S}{V_s} \quad [\text{V/VT}] . \quad (17)$$

Modern devices S show values typically ranging between 50 mV/T and 1 V/T at the nominal biasing current. The sensitivity drifts with time which necessitates periodic calibration, $V_H = f(B)$, of the Hall probe (see Section 5).

2.4.2 Offset voltage

The offset voltage V_{off} is a parasitic output voltage that appears in the absence of a magnetic field. Some origins of this voltage are

- a structural asymmetry of the active part (errors in geometry, non uniform doping density, contact resistance, etc.);
- alignment errors of the sense contacts, one being further upstream and the other one downstream with respect to the bias current [see illustration in Fig. 5 (left)].

The offset changes with time. The drift of the offset voltage originates from

- the ageing of the semiconductor that plays an important role for cryogenic Hall probes. Thermal cycles between 4.4 K and 300 K will cause damage in the active area (microcracks), changes in contacts, resin glue and package. The mechanical stress that appears will modify the repartition of the electron population in a many-valley semiconductor like n-type silicon. This evolution of concentration will change the Hall coefficient and hence the Hall voltage.
- The diffusion of the impurities from the contact into the active area, modifying the mobility and thus the Hall coefficient. This effect can be limited using a cruciform geometry.

It is practical to use the analogy of the resistive bridge to model the presence of the offset. Ideally the four resistors are equal. A resistance variation in one or more legs leads to bridge asymmetry resulting in the appearance of an offset [Fig. 5 (right)].

In conventional ‘plate-like’ Hall probes, the offset voltage is in most cases small, in the range of a microvolt. In some integrated circuit Hall sensors, however, it can reach a value of hundreds of microvolts. It is linearly dependent on the biasing current and is temperature dependent. The measured output voltage is therefore the sum of the offset and of the Hall voltages and the two contributions cannot be distinguished when $B \neq 0$. The offset can be measured in a zero-field gauss chamber or compensated by means of external circuits. Compensation techniques are presented in Section 4.4.1.

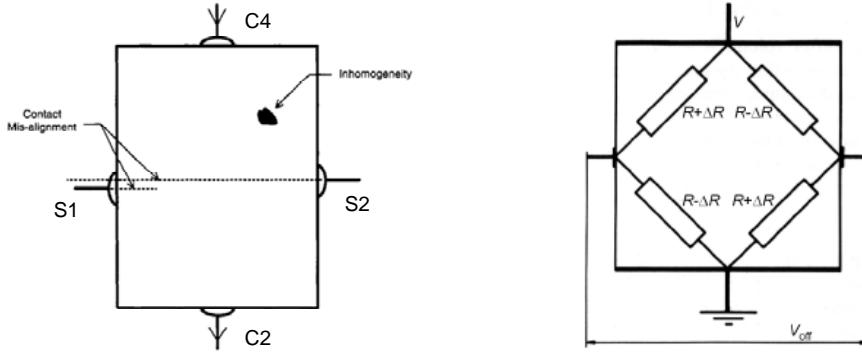


Fig. 5: Left: Offset resulting from the misalignment of sense contacts and from inhomogeneities in the material [B2, p.13]. Right: Analogy with an unbalanced bridge resistor [B1, p. 214]

Remark: Although the offset is usually expressed in terms of output voltage, it also needs to be considered in terms of magnetic field. A Hall element having a significant offset can still be valid for low magnetic field level applications when it features, at the same time, a very good sensitivity.

2.4.3 Temperature dependence

The sensitivity of the Hall probe can be affected by some environmental parameters such as the temperature. The main contribution to this effect originates from the thermal activated behaviour of the carrier concentration. Two cases have to be distinguished:

- The Hall element is made with materials with small band-gap (like InSb for example) and will display strong temperature dependence. It comes from the dependence of the intrinsic carrier’s density n_i that increases with the temperature following the activation law

$$n_i \sim T^{3/2} \exp(-E_g / 2 k T). \quad (18)$$

This dependence can be reduced by introducing donor impurities, trying to render the material extrinsic. In that case, the contribution to carrier concentration from thermal activation is negligible and the carrier concentration becomes practically unchanged within a temperature range called the exhaustion range. In this range, the Hall constant R_H is practically unchanged.

- The material that composes the Hall element has a large energy band-gap (like the InAs for example). The semiconductor stays extrinsic up to high temperatures despite low doping.

The temperature coefficient sensitivity and the temperature dependence of the offset voltage are usually expressed by the two following coefficients

$$\gamma_T = \frac{I}{V_H} \frac{\delta V_H}{\delta T} [\text{K}^{-1}] \quad \frac{\delta V_{offset}}{\delta T} [\mu\text{V/K}]. \quad (19)$$

As an illustration, Fig. 6 shows the Hall voltage versus temperature of a Hall generator made from InSb biased by two constant currents, 1 mA and 5 mA.

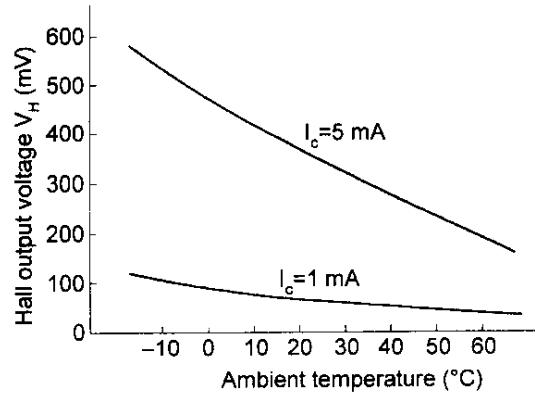


Fig. 6: Change of the Hall voltage with respect to temperatures [B2, p.191]

Like the sensitivity, the offset drifts over the temperature but randomly, varying from device to device. Some drifts are caused by piezo-resistive effects. The mechanical stress induced by the thermal expansion of the material leads to changes in the recorded voltages. These variations are not acceptable for some applications when high measurement stability is needed and/or the measurement is long in time. A temperature regulation has to be provided or a correction of the sensitivity must be made.

2.4.4 Non-linearity $V(B)$

The Hall voltage measured in a finite size Hall device is not a linear function of B . This non-linearity has various origins. One source is the non-linearity coming from the material property NL_M , i.e., caused by the field dependence of R_H . Another contribution, NL_G , comes from the magnetic field dependence of the geometrical factor G . NL_M and NL_G exhibit a quadratic magnetic field dependence but in opposite sign and of the same order of magnitude. The Hall generator is generally designed in such a way that these two effects could compensate each other to a large extent.

$$NL^* = NL_M + NL_G . \quad (20)$$

The non-linearity is estimated as the deviation of the real voltage at a given biasing current and magnetic field from the best linear fit V_{Hfit} of the measured value. It is usually expressed in terms of percentage by

$$NL = \frac{V_H(I,B) - V_{Hfit}}{V_{Hfit}} * 100 [\%] . \quad (21)$$

The non-linearity depends on the Hall generator properties and on the magnetic field interval. High linearity Hall devices will typically display a NL coefficient less than 0.2% at room temperature and for fields below 1 T. This error is determined during the calibration of the Hall probe V_H vs. B and will be corrected (see Section 5.1).

* NB: In case of vertical Hall sensors built using a CMOS technology (see Section 3.2.1), a third source of non-linearity is the variation with the magnetic field of the effective thickness of the sample $t(B)$. The width of the depletion layer between the two sensing contacts changes with B , leading to a variation of the sensitivity [8].

2.4.5 Noise effects

Noise that appears in the outputs of the signal voltage is one important limitation of the measurement precision. Various sources of inherent noise are generated by the Hall generator itself. The Johnson noise results from thermally induced motion of the electrons in the material. The resulting root mean square of this parasitic voltage V_{jn} is related to the resistance of the device, to the operating temperature and to the bandwidth of the signal by

$$V_{jn} = \sqrt{4kTRB_w} \quad (22)$$

where k is the Boltzmann constant ($1.38 \cdot 10^{-23} \text{ J K}^{-1}$); T is the absolute temperature in K; R is the resistance in ohms; B_w is the bandwidth in hertz.

To limit this source of noise, the bandwidth must be maintained small, just wide enough to pass the minimum required signal. Johnson noise is limited in the domain of hundreds of nanovolts using a 10 kHz bandwidth and a proper choice of the impedance.

The flicker noise has multiple origins and, in a first approximation, is inversely proportional to the frequency (it is also called $1/f$ noise). Working with low DC frequency signals (for example 1 Hz), this component becomes a significant problem. This noise depends on the material and on the fabrication technique of the Hall element and can therefore be improved. The noise voltage spectral density of a Hall device can be schematically described by the plot in Fig. 7. The corner frequency f_c that limits the two noise regions ranges from 1 to 100 kHz.

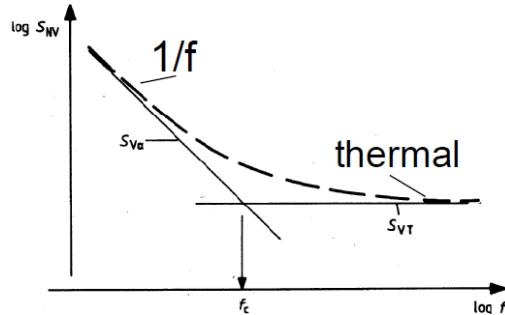


Fig. 7: Noise voltage spectral density of a Hall device [B1, p. 215]

2.5 ‘The planar’ Hall effect

This galvanomagnetic effect was investigated in 1954 by C. Goldberg and R.E Davis [9] and corresponds to an induced electric field perpendicular to the current in the current–magnetic field plane. The experimental conditions are similar to the ‘classical’ Hall effect but in that case, a voltage is measured normal to the direction of the current flow *with a magnetic field component in plane made by the current and the voltage measurement direction* (see geometry displayed in Fig. 8). The magnetic field \mathbf{B} can be decomposed into the following components: the in-plane component \mathbf{B}_p ($B_p \cos \phi, B_p \sin \phi$) and the component B_z normal to the plane of the plate, $B_z = B \cos \gamma$. The angles ϕ and γ correspond respectively to the angles between \mathbf{B}_p and the current direction (x direction) and between \mathbf{B} and \mathbf{z} (see Fig. 8). The electric field \mathbf{E} is expressed by [10, 11]

$$\mathbf{E} = \rho_0 \mathbf{J} - R_H [\mathbf{J} \times \mathbf{B}] + P_H (\mathbf{J} \cdot \mathbf{B}) \mathbf{B} . \quad (23)$$

With the configuration described in Fig. 8, the transverse coordinate of the electric field E_y is

$$E_y + R_H B \cos \gamma J + P_H J (B_p \cos \varphi) (B_p \sin \varphi) \quad (24)$$

where R_H is the Hall coefficient and P_H a coefficient called the planar coefficient, ρ_0 is the magneto-resistivity.

As shown in Eq. (24), the existence of a residual in-plane magnetic field component \mathbf{B}_p generates an additional electric field in the plane perpendicular to B_p and proportional to the square of its intensity. It results in the generation of an additional voltage U_p measured in the transverse direction (y-direction) which is added to the Hall voltage V_H

$$V_{output} = V_H + U_p = \frac{R_H I B \cos \gamma}{t} + \frac{P_H}{2t} B_p^2 \sin(2\varphi) I. \quad (25)$$

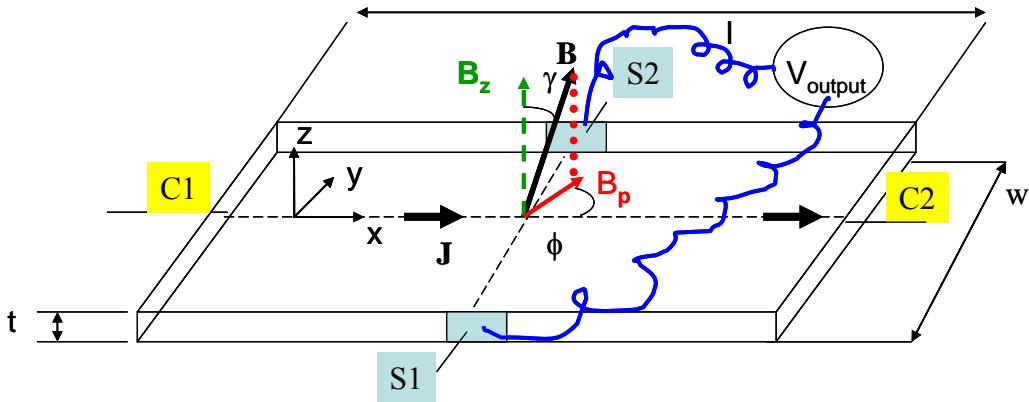


Fig. 8: Very long plate exposed to a magnetic induction \mathbf{B} with an in-plane component \mathbf{B}_p

U_p is proportional to B_p^2 and follows a $\sin 2\phi$ angular dependence where ϕ is the angle between the current and the in-plane component of the magnetic field, as depicted in Fig. 9. This superimposed voltage U_p will spoil the accuracy of the measured value coming from the field component perpendicular to the plane. Planar Hall effect voltage compensation techniques have to be used. As shown by Eq. (25), the planar voltage disappears for $\phi = 0^\circ$ or for multiples of 90° . If the orientation of the in-plane component is known, a rotation of the Hall probe parallel or perpendicular to this direction will cancel this effect. The amplitude of the planar voltage depends on the semiconductor used for sensing. Material with higher conductivity will display a smaller contribution (see amplitudes of U_p in Fig. 9).

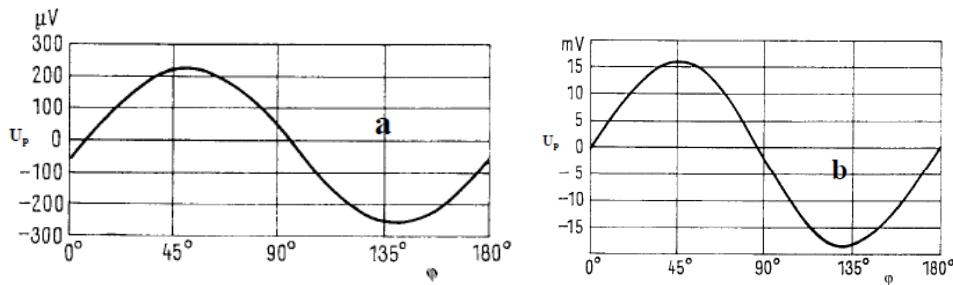


Fig. 9: Angular dependence of the planar voltage U_p for an InSb Hall probe for two different conductivities: (a) n-doped InSb $\sigma = 750 \Omega^{-1} \text{cm}^{-1}$, (b) intrinsic InSb, $\sigma = 200 \Omega^{-1} \text{cm}^{-1}$ (from Ref. [7])

Compensation techniques were proposed to cancel the planar Hall effect. The use of two identical Hall probes on the top of each other, but the lower turned by 90° w.r.t. the upper one, was

suggested by Turck [10] and described in Ref. [12]. The direction of current of probe 1 and probe 2 are perpendicular and make an angle ϕ_1 and $\phi_2 = \pi/2 - \phi_1$ with the planar component of magnetic field.

In that case, the compensation occurs because $\sin(2\phi_1) + \sin(2\phi_2) = 0$. The limitations of this arrangement are listed in Ref. [13]. In particular the Hall probe material and the geometry must be identical. Moreover the centre of the probe must be located on the same line perpendicular to the active part and the angle of rotation has to be exactly 90°. An arrangement using a single sensor, where the voltage and current connection are interchanged on the fly, is described in Ref. [14].

2.6 Quantum Hall effect (field measurements at cryogenic temperature)

The (integer) quantum Hall effect is observed in two-dimensional electron systems subjected to low temperatures and strong magnetic fields. The measured Hall conductivity σ shows in that case quantized values. This phenomenon was discovered in 1980 at the Grenoble High Magnetic Field Laboratory in France by Klaus von Klitzing [15]. Studying the Hall conductance of a two-dimensional electron gas at very low temperature, he observed that the dependence of the Hall voltage as a function of the magnetic field displays a staircase sequence of wide plateaux instead of the expected monotonic increase (see Fig. 10). These steps occur at precise values of conductance equal to ne^2/h , where n is the integer that characterizes each plateau, e the electron charge and h the Plank constant. The integer quantum Hall effect can be explained by the quantization of the electron orbits in two dimensions. These orbits correspond to Landau levels which have the discrete energy values

$$E_n = \frac{h}{2\pi} \omega_c (n+1/2). \quad (26)$$

$\omega_c = eB / m$ is the cyclotron frequency.

The application of strong magnetic fields leads to a degeneracy of each Landau level and the electrons have many possibilities for a state occupation. The change of electronic levels gives rise to ‘quantum oscillations’ called the Shubnikov–de Haas oscillations (see Fig. 10). Such oscillations limit the accuracy of the Hall voltages measurements. The following plot displays the deviation ΔV of the Hall voltage from a straight line as a function of the applied field resulting from these oscillations. The perturbation can reach up to a level of one per cent at high fields, depending on the semiconductor used for the sensing. A precise calibration at cryogenic temperatures is necessary for a low temperature and high accuracy of magnetic field determination.

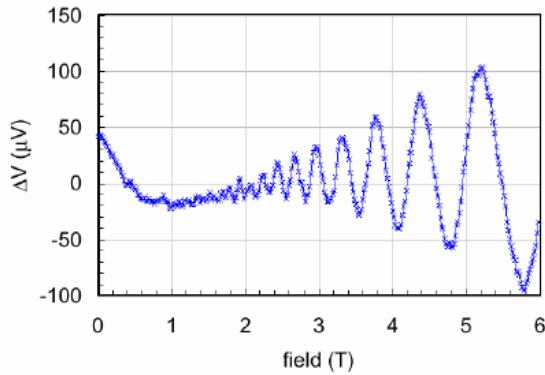


Fig. 10: Shubnikov–de Haas effect in a Hall generator at 4.2 K. The plot displays the deviation ΔV of the Hall voltage from a straight line as a function of the applied field (reprinted from Ref. [16])

3 Single-axis Hall probe

Hall probes, i.e., Hall generators mounted on a sample holder, are available in a variety of shapes and active area dimensions.

3.1 ‘Plate-like’ Hall probes

The use of a sensor with a ‘plate-like’ geometry is optimal for measurements of homogenous magnetic field (at the level of the active part), provided that the sensor can be positioned correctly. The families, transverse and axial, are defined by the direction of the flux lines through the probes (see Figs. 11 and 12). The Hall element is mounted at the end of a long flat stylus. The transverse type is designed for the measurement of the main magnetic field component perpendicular to the chip surface and is sufficient for many applications. This Hall plate has a quite simple structure and can be adapted to almost any kind of semiconductor technology without special requirements or additional steps. For this reason the devices are generally very cheap to manufacture.



Fig. 11: Transverse Hall probe (picture from Ref. [6])

The axial type is generally cylindrical, the Hall plate being mounted perpendicular to the axis of the cylinder. It is designed to measure fields parallel to the axis (Fig. 12).



Fig. 12: Axial Hall probe (picture from Ref.. [6])

At the beginning of the 1980s other geometries were proposed, based on the following general criteria established for good Hall devices [17]:

- In the active part, the current density and the magnetic field have to be as much as possible close to the orthogonality.
- The shape of the active region has to facilitate the measurement of the Hall voltage V_H .
- Offset voltage at zero fields has to be minimized, i.e., along the sensing contact (S1S2), the electrical field \mathbf{E} and the Hall field \mathbf{E}_H have to be as much as possible close to orthogonality.

3.2 Non ‘plate-like’ devices

The conventional shape of the Hall generator mentioned in Section 2.3.1 (rectangular, cross, bridge, etc.) is sometimes not adapted in the case of a non-homogeneous magnetic field or a particular shape of the field lines. A second class of Hall generators, called non-plate-like, features a tri-dimensional structure. They are used to fabricate sensors capable of measuring simultaneously two or three components of the magnetic field. The devices are structured in order to fulfil the three criteria presented above. Reference [17] is a survey of this second class of Hall generator.

3.2.1 Vertical Hall Device (VHD)

The vertical Hall device (VHD) was devised more than 20 years ago (see Refs. [18, 19]) to respond to a magnetic field parallel with the chip surface. The device is called ‘vertical’ as the region that generates the Hall effect is perpendicular to the chip plane. It can be obtained from a transverse Hall plate by the transformation schematized in Fig. 13:

- turning the horizontal Hall plate into a vertical position;
- dividing one biasing contact into two parts and arranging all the contacts in line on top of the plane.

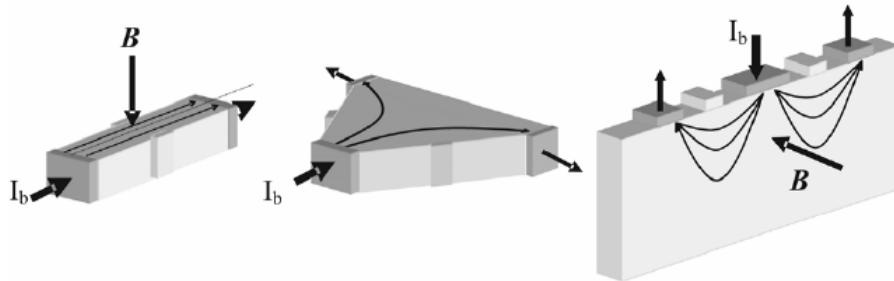


Fig. 13: Transformation of a plate-like Hall probe into a vertical one (from Ref. [20])

The advantage of this geometry is that all the electrical contacts are on one side of the device so it is compatible with standard one-side silicon technology. The use of vertical sensors allows one to fabricate the devices by evaporation, sputtering, and the pattern is structured using photolithography. With integrated circuit technology, not only thin devices with high sensitivity but also deeper structures can be obtained by implantation and thermal diffusion of impurity ions. Vertical Hall sensors with high current sensitivity $S_I = 400 \text{ V/TA}$ have been produced by Landis and Gyr (Switzerland). A patented technology called ‘Vertical Hall Technology’ based on n-type silicon wafers with low doping concentration like $2 \cdot 10^{14} \text{ cm}^{-3}$, was employed. The technology and the device are described in Ref. [20].



Fig. 14: Left: Vertical ‘bottomless’ Hall sensor made by the VH technology (from Ref. [20]). Right: Principle of a vertical Hall sensor. The current flow and the equipotential lines are shown. The Hall voltage is measured between the two probe contacts (from Ref. [20])

Figure 14 (left) displays the cross section of a vertical Hall sensor built with this process. Figure 14 (right) describes the principle layout of a vertical Hall sensor with \mathbf{B} perpendicular to the plane. The active zone belongs to the n-type substrate material, limited by a p-well ring. In the version presented in Fig. 14, the VHD has five contact terminals made by n+ implantation and arranged in a

line on top of the active diffusion region. The current is injected through the central contact to the two outside ones connected to ground. The Hall voltage is sensed between the probe contacts. The output voltage V_H is given by the integral of an equipotential line between the two sense contacts. However, production costs turn out to be high due to the special process of obtaining low doped and deep structures. Moreover, the sensor is open downward and allows a deep current flow (e.g. 30 μm). Other technologies like bipolar [21] or CMOS technology [22, 23] have been used to manufacture vertical devices with the key advantage of allowing the co-integration of the electronics on the same chip.

From Eqs. (10), (16) and (17), it turns out that current sensitivity S_I is inversely proportional to the product of the thickness t and the doping concentration n_d :

$$S_I \propto \frac{1}{tn_d}. \quad (27)$$

To increase the sensitivity, the first obstacle to overcome is therefore to limit the depth of the biasing current flow. Using the CMOS technology with an additional deep-n well, the device could be isolated from the substrate by the layer and the current will therefore be concentrated near the surface (between 0 μm and 3 μm). The second difficulty is to use a low doping n-layer. A patented method [24] involving a different structure of the implementation mask reduced the level of doping by almost one order of magnitude near the surface. A complete description of the CMOS process [22 and B3] and the performance of vertical devices fabricated with the CMOS technology can be found in Ref. [25]. Variations of the sensor layout have also been developed. A design with four contacts allowing the four-fold symmetry advantages for the offset compensation was proposed [8]. An improved version features a design with six contacts with two outside contacts short-circuited to minimize the offset voltage. Figure 15 displays the three design structures of a vertical Hall sensor, realized using the integrated circuit technology: five terminals (centre), four terminals (left), six terminals (right).

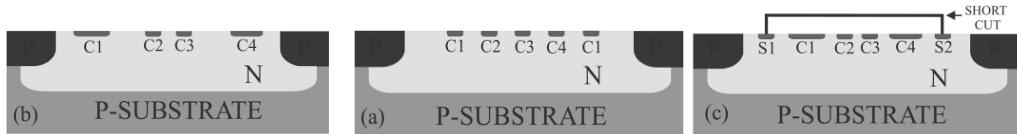


Fig. 15: Structure of a vertical sensor sensitive to a magnetic field parallel to the surface plane built using a derived CMOS technology. Three variations are presented: the design with five terminals (a), four terminals (b), six terminals (c) [from Ref. B1 p. 263]

This technology allows the manufacture of large quantities of Hall devices with very small dimensions (micro-sensors) and for a low price. Sensors fabricated with the integrated circuit technology have highly reproducible geometry and physical properties. Figure 16 is an example of a vertical device.

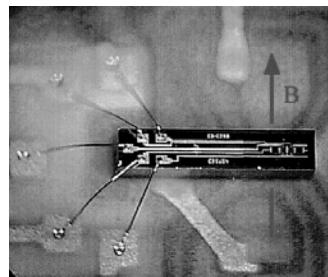


Fig. 16: Vertical device sensitive to a magnetic field parallel to the surface of the chip from Ref. [17]

3.2.2 Cylindrical devices

A cylindrical Hall device features a magnetic field sensitive volume in the form of a half-cylinder imbedded in the sensor chip. It results formally from the conformal deformation of a vertical device (see Fig. 17). This shape of sensor was devised to be sensitive to circular magnetic field form. A typical application is the current sensor.

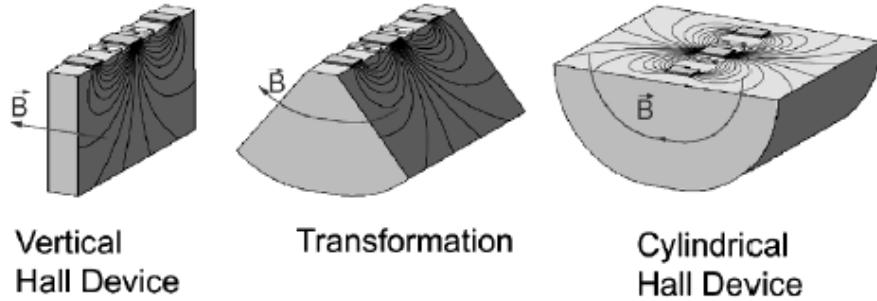


Fig. 17: Cylindrical device (right) obtained from a vertical Hall device (left). The sensitive volume of a cylindrical device is adapted to the measurement of a circular magnetic field [from Ref. B1 p. 265]

3.2.3 ‘Planar Hall’ device with open bottom

This Hall generator has the aspect of the Hall plate but features an open bottom structure, i.e., a structure not isolated at all from the bulk. The active zone is deep and can be compatible with a vertical Hall device to form a three-axis sensor. Figure 18 illustrates the structure of the so-called ‘planar Hall’ device made with the technology for vertical devices (see Section 3.2.1). The design has a cross shape to make the sensor invariant for a switch between the biasing and sensing contacts.

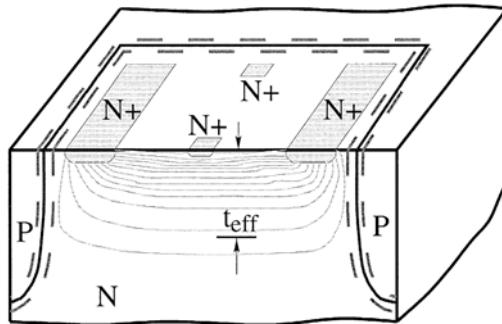


Fig. 18: Planar Hall device made with the technology of vertical devices. The effective thickness t_{eff} defines the depth down to which the current is forced to flow [26]

As for the vertical device, the active zone is limited laterally by the P-well ring but not vertically. To increase the sensitivity, the spread of the current into the bulk has to be limited to an effective thickness t_{eff} just below the chip surface. The reduction of the size of the sensor together with an additional surface doping, results in a well-defined conductive zone below the surface. The sensor can reach a current-related sensitivity of 0.33 V/mA T [26].

4 Two- and three-axis Hall effect devices

4.1 Two-axis device

A two-axis Hall device that measures simultaneously the two in-plane components B_x and B_y can be obtained by associating two vertical bottomless generators orthogonally. The combination to generate a two-axis device is displayed in Fig. 19.

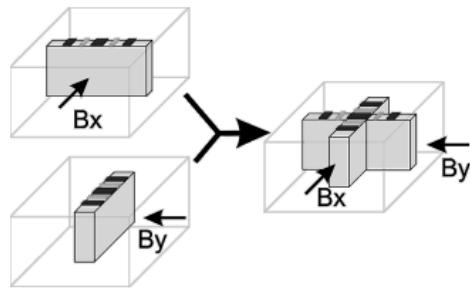


Fig. 19: Fabrication of a two-axis Hall device [from Ref. B1 p. 262]

Combining such a magnetic sensor with a permanent magnet, one can fabricate an accurate contactless angular position sensor [17].

4.2 Three-axis sensor design

4.2.1 Conventional geometry

A combination of three Hall magnetic generators oriented in the three main directions and supported by three orthogonal plates was and remains a classical way to measure simultaneously the three components of the magnetic field. This arrangement (see Fig. 20) has several disadvantages:

- The spatial resolution is limited by the distance between the single elements.
- The precision in the orthogonality of the three sensors will depend on the assembly procedure.
- The measurement is not ‘point like’ as the flux density is not measured at the same spot.
- The number of contact leads is high.

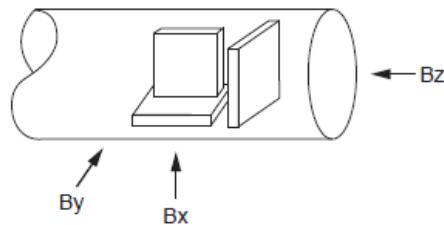


Fig. 20: Three-axis Hall probe built by associating three single-axis Hall probes (from Ref. [6])

The main advantage with this construction is that the accuracy of a single Hall probe can be practically kept in the three directions.

4.2.2 Geometry using vertical Hall devices, micro Hall sensors

With the development of vertical Hall devices it is possible to fabricate an integrated structure by merging two or three modules. The planar technology is based on the principle that the magnetic field comes from the decomposition of two perpendicular in-plane components and a third orthogonal to it. The resulting ‘new generation’ of Hall sensors features a reduced active measurement volume and an improved orthogonality of the axes of sensitivity. A tri-axial chip with a square shape, built using a small single volume of a semiconductor crystal, was devised in 1999 [20]. The working principle consists in the association of the vertical and plate-like structure. For the measurement of the in-plane components called B_x and B_y , the geometry is similar of that a vertical device. For the component perpendicular to the sensor plane, the geometry is derived from that of a Hall plate but with an open bottom. As displayed in Fig. 21, the structure consists of a square shape silicon block with four current contacts in the corners and four sense contacts between them.

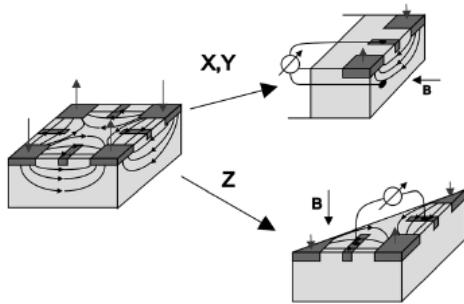


Fig. 21: Basic structure of a tri-axial Hall device. From the recorded voltages V_x , V_y , V_z , the three components of the magnetic field are obtained. The bias contacts are located on the corners of the square and the sense contacts are between them. The arrows indicate the flow direction of the currents [from Ref. B1, p. 266]

Similarly, three-axis Hall chips are made using two different types of devices as shown in Fig. 22. A VHD is merged with a planar Hall device whose geometry is described in Section 3.2.3.

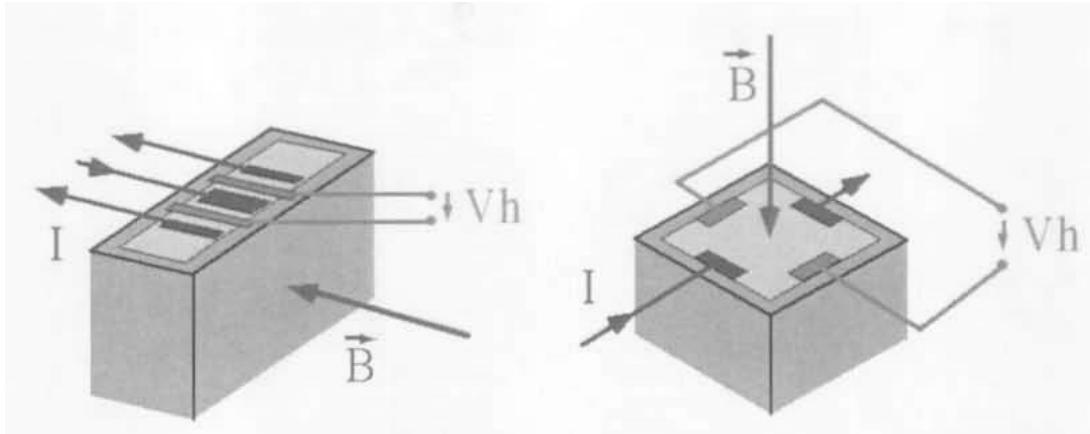


Fig. 22: Vertical Hall element (left) and a Hall planar device, i.e., a conventional Hall generator with an active part open to the bottom (right). The two types of element implemented in a chip will measure the three components of the magnetic field. The pictures are reprinted from Ref. [20]

The possibility of co-integrating a Hall generator and smart electronics on a single chip allows the fabrication of integrated Hall sensors, having dimensions typically between $10 \mu\text{m}$ and $100 \mu\text{m}$. A picture of such a Hall element fabricated with the CMOS technology is shown in Fig. 23. Hall cross shape sensors, featuring an active area as small as $0.1 \mu\text{m}^2$ square is even achieved. The performance, the technologies and the applications of the micro Hall devices are reviewed in Refs. [27, B5].

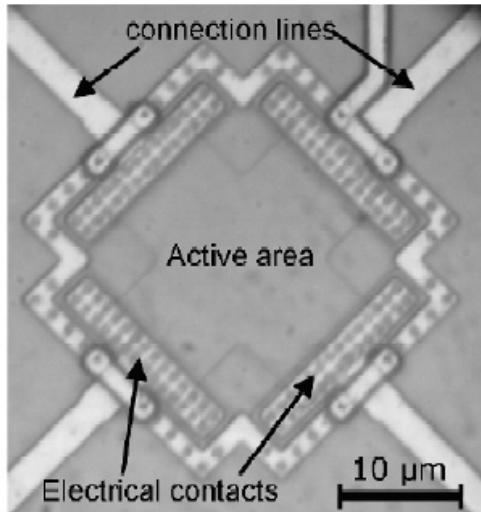


Fig 23: CMOS Hall sensor (from Ref. [B3], p. 4)

The small size of these micro-sensors facilitates the integration of a large number of Hall elements in the same chip surface. Therefore, another solution for multi-axis sensors consists in spacing the two types of device for each axis of sensitivity. To assure a measurement for each component in the same spot, the sub-sensors have to be arranged in pairs at equal distance from the measurement centre. For example, the structure shown in Fig. 24 is composed of eight Hall sensors on silicon [20]. The components that are parallel to the sensor surface are measured by two pairs of vertical Hall devices located on the opposite sides of the sensor system. The perpendicular component is measured by four horizontal devices located in the corner of the system.

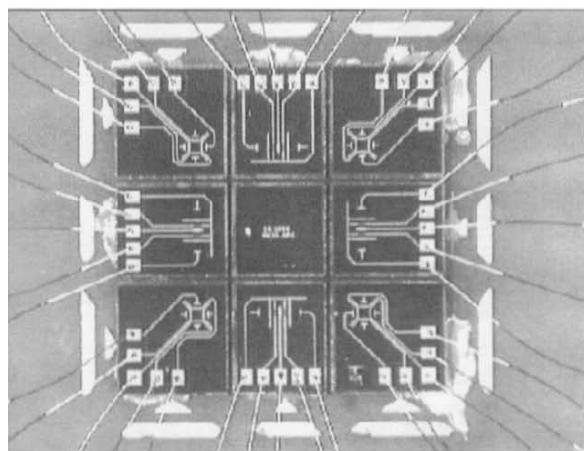


Fig. 24: Three-axis sensor made of eight sub-sensors. The two sensor types are arranged to share one common active volume, formed by the eight sub-sensors. The dimensions are of the active part are $2 \times 2 \times 0.2 \text{ mm}^3$ (from Ref. [20])

4.3 Example of a three-axis integrated circuit Hall sensor

The three-axis IC Hall sensor produced by the company SENIS is presented here as an illustration, see Refs. [28, 29]. The sensing part is composed of one single horizontal Hall element at the centre which senses the component perpendicular to the magnetic probe. Two pairs of vertical devices aim at measuring each in-plane component of a magnetic field (see Fig. 25). The devices are fabricated using the N-well CMOS technology. The dimensions of the sensing part are about $150 \mu\text{m} \times 150 \mu\text{m}$. The sensitive volume is $0.15\text{mm} \times 0.15 \text{ mm} \times 0.01 \text{ mm}$ and the mutual orthogonality of the three axes is below 0.5° . The sensing part is integrated into a complex electronic circuitry which supplies the current, provides the amplification and the cancellation of parasitic signals like the offset and the $1/f$ noise (see Section 4.4.1). A temperature sensor is also integrated in the chip. The dimensions of the chip are $6400 \mu\text{m} \times 6400 \mu\text{m} \times 550 \mu\text{m}$.

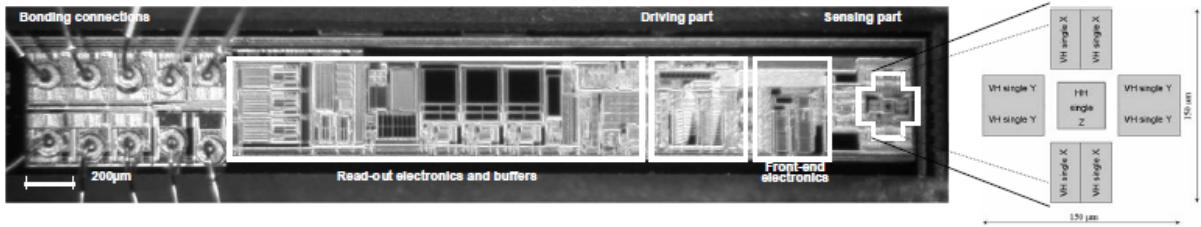


Fig. 25: Photo of a three-axis Hall probe chip produced by the company SENIS (from Ref. 29])

The Hall sensor features a spatial resolution of 0.15 mm , a negligible offset voltage and planar Hall effect. The IC Hall sensor has equipped a three-axis teslameter, commercially available with a standard accuracy of 0.1% in the magnetic field range between 20 mT and 2 T .

The performance of this teslameter is summarized in Table 2.

Table 2: Specifications of a commercial teslameter from SENIS [29]

Full scale (nominal)	2 T
Sensitivity to D.C magnetic field	5 V/T
Tolerances of sensitivity ($B = 1 \text{ T}$, D.C.)	0.1%
Temperature coefficient of sensitivity	$<100 \text{ ppm}/^\circ\text{C}$
Residual non-linearity (up to 2 T)	$<0.05\%$
Planar Hall effect: $V_{\text{plan}}/V_{\text{vert}}$ (1 T)	$<0.01\%$
Offset ($B = 0 \text{ T}$)	$<\pm 1 \text{ mV} (\pm 0.2 \text{ mT})$
Long term instability of sensitivity	$< 1\% \text{ over } 10 \text{ years}$

This new generation of Hall magnetometers equipped with IC Hall sensors is now commercially available, offering the possibility to measure all three components of the magnetic field even in highly inhomogeneous fields due to strong reduction of the active volume with respect to the conventional Hall probes. The technological developments underlying each of the components of this new type of magnetometer system are reviewed in Ref. [30].

4.4 Improvements in the three-axis IC sensor characteristics

The most important issues are offset, noise, temperature dependence, and planar Hall effect.

4.4.1 Offset reduction

Hall sensors made with CMOS technology suffer from large offsets which make them less appropriate in high accuracy applications. As mentioned in Section 2.4.2, the offset has many origins: material

inhomogeneities, material construction, mechanical stresses, and temperature variations. These resulting offset fluctuations can hardly be compensated even with frequent calibration. Offsets at the level of few mT are typically observed and realistic signal levels of a few μV used in some applications become then completely masked. The orthogonal coupling of the sensors is the traditional static offset cancellation (reduction) method. The coupling is done by the direct connection of two sensors, each having a biasing mode shifted by 90° [31]. Figure 26 describes the equivalent circuit using the electrical analogy of the bridge resistor. The offset cancellation is provided by the matching properties of the two sensors. For a successful compensation, the increment of resistance ΔR that appears in one leg of the bridge has to be equivalent in the two sensors.

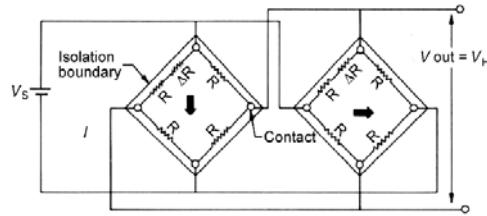


Fig. 26: Pairing of two matching sensors. They are connected electrically in parallel but with orthogonal current directions [31]

This compensation gives approximately a tenfold improvement over the uncompensated Hall probes, cancelling in particular the effect of mechanical stresses. This technique is limited by the mismatching between the Hall devices. The drift of the offset could not be strictly equivalent for the two Hall generators and the resulting compensation will only be partial. A step further is the use of the so-called ‘spinning current Hall plates’ developed at Delft University in 1992 [32]. As shown in Fig. 27, the Hall generator is symmetrical and the current terminals are periodically swapped with terminal voltages. In the first cycle, the biasing current flows between points 1 and 2. The output voltage between points 3 and 4 is $V_{out1} = V_H + V_{off}$. In the second cycle, the connections are commutated and an output voltage $V_{out2} = V_H - V_{off}$ is recorded between the points 1 and 2. The difference between the mean values V_{out1} and V_{out2} is equivalent to V_H . The offset reduction is achieved by the fact that Hall voltage does not depend on the current direction, but the offset voltage does. Using four terminal devices, the offset can be reduced to 0.5 mT. Spinning-current Hall plates with eight contacts were also designed. With the association of dynamic offset cancellation techniques like auto-zeroing and chopping [32], the offset can be reduced to values down to 1 μT [33].

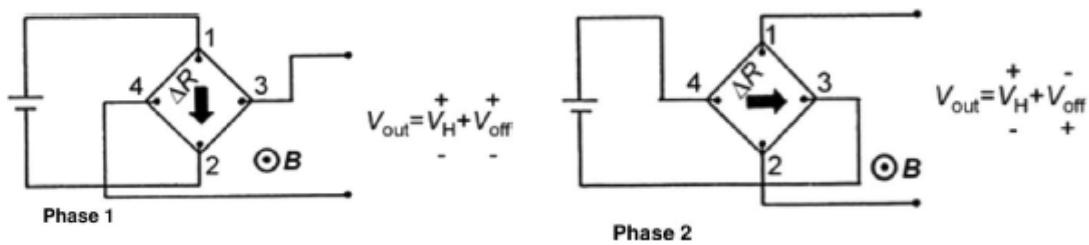


Fig. 27: Spinning current technique for offset compensation: (a) phase 1, current is flowing from 1 to 2 and the voltage is sensed between 3 and 4. (b) phase 2, the terminals are swapped (from Ref. [29])

4.4.2 Compensation of the temperature dependence of the sensitivity

The simplest way to compensate the variation of the sensitivity with temperature is the use of a resistor connected in parallel with the input terminal of a Hall device. The key point is that at constant bias current, the change in the input resistance of the Hall device will modify proportionally the output voltage. The input resistance, defined as the resistance measured between the current leads, will be used as a temperature indicator. As shown in Fig. 28, the current I is divided into two parts: I_{in} flows in the Hall element and I_R in the resistor. The change of the input resistance caused by the temperature variation will affect the current redistribution among I_{in} and I_R . Thus by adjusting the bias current I_{in} the change caused by the temperature effect is compensated.

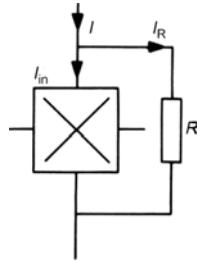


Fig. 28: Simplest compensation scheme for temperature effect compensation on sensitivity (from Ref. [B1], p. 281). The square with the cross is the symbol of the Hall device (the output is the result of the multiplication of the biasing current and of the magnetic field)

More sophisticated circuits can also be built with an operating amplifier and a differential magneto-resistor scheme [B5].

4.4.3 Reduction of the 1/f noise

The ‘spinning current’ technique (Section 4.4.1) also reduces the 1/f noise. The 1/f noise can be seen as a slowly fluctuating offset voltage, appearing alternately in the output voltage terminals when the bias current is rotated back and forth. Provided that the switching frequency is higher than the corner frequency f_c (see Section 2.4.5), the voltage noise contribution, together with the offset, can be separated from the useful signal and filtered. The spinning frequency is high, several hundred kHz.

4.4.4 Hall planar effect reduction in three-axis Hall sensors

Evaluation of the contribution of the planar Hall effect is of a crucial importance if one uses three-axis Hall sensors. For vertical devices made with silicon, Schott found a theoretical contribution of 12%. This estimation was not confirmed by the measurements, indicating, on the contrary, a lower contribution down to 1%. This strong reduction was attributed to the particular spatial geometry of the vertical Hall device and its resulting current paths in the vertical structure [34]. Unlike the case of the plate shape, only a short section of the current filaments are distorted by the planar flux density, thus reducing the planar Hall effect. Current spinning also cancels the planar Hall voltages. We refer again to the corresponding magneto-resistance model presented in Fig. 29. Let us take the case of a Hall element with a cross geometry. The current I is sent from one contact to its opposite one and the output signal is measured between the two other contacts. The model is made by four resistors having, for simplification, an equal value at zero-field. A planar magnetic flux density \mathbf{B}_p acting on the diagonal direction will unbalance the resistor bridge by changing the resistance values. To a first approximation, the magnetic field will indeed change the value of the equivalent resistor only if a field component perpendicular to the current flow through the resistor is present. This unbalance will create

a supplementary voltage U_p proportional to B_p^2 (as recalled in Section 2.5), appearing between two of the four terminals.

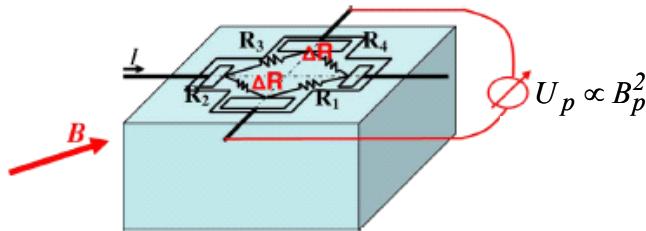


Fig. 29: Contribution of the Hall planar effect to the output voltage on a Hall element electrically modelled by a resistor bridge (picture from Ref. [29])

The similarities with the situation described in Fig. 5 for the offset led to the investigation of the impact of the ‘spinning current’ method on the planar Hall effect contribution. The result for a horizontal Hall plate is presented in Fig. 30. The amplitude of the planar Hall voltage V_{PH} is determined with and without spinning current effect compensation, for fields up to 2 T. The compensation works successfully with reduction of the ratio planar/conventional Hall voltage from 1.3% down to 0.02%.

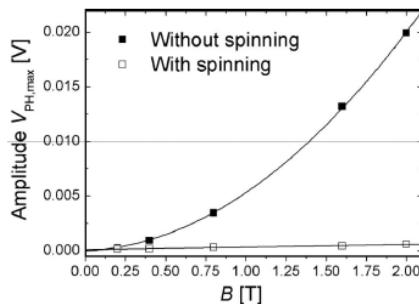


Fig. 30: Measured planar Hall voltage without and with the application of the spinning current technique [29]

4.5 Three-axis Hall sensors: limitations and the next challenges [35]

- For the Hall probes built in the conventional way (Section 4.2.1), one main improvement concerns the physical size. The merging of three single-axis sensors sets a lower limit on the size of around a few mm³. This not only limits the gap size that one can map but also the capacity of the sensors to be used in highly inhomogeneous fields. The distance between the three sensors leads to a different measurement of the field and therefore, the measurement is not a ‘three-axis’ consistent point measurement. The accuracy of the orthogonality is another critical issue. The orthogonality in a traditional three-sensor assembly is not easy to achieve with sensors that have such a small surface area. A tilt error of 1 μm on a sensor of 1 mm diameter may already be detectable.
- In the three-axis IC Hall sensor, *sensitivity and accuracy* are the major concern. The size constraint limitation combined with the less efficient vertical Hall probe geometry and the material choice for the integrated sensor (Si), renders this type of sensor difficult to use for high-accuracy field measurements (e.g., at 0.01% of the measurement range). Multi-axis IC sensors typically feature an accuracy between 0.1% and 1%, one or two orders of magnitude

less than conventional single-axis probes. A standard accuracy of 0.01% (of the measurement range) has been difficult to achieve up to now. Noise coming from the circuitry, drift of the offset due to temperature effects and difficulty in the calibration to remove the effects caused by the mutual interaction of the three axes are the major limitations.

- Cryogenic applications also need special attention. Some integrated sensors ‘freeze out’ at low temperatures, and are therefore not suitable for cryogenic applications. The oscillations coming from the quantum Hall effect (Section 2.6) have also to be considered. Cryogenic multi-axis sensors have therefore, up to now, been built in the traditional way (three single-axis sensors), with carefully chosen bulk semiconductor material, and with severe constraints on the mechanical stability to be able to survive the temperature cycling.
- Finally, the calibration of three-axis devices remains a complex and long process (see next section). Up to now, the end-user generally does not have the equipment, the time and the knowledge required and will rely upon the manufacturer. The calibration of the three major axes at high magnetic fields is not easy since the small gap of these magnets makes it difficult to insert the probe. In addition, a good mechanical reference surface is required in order to have a reproducible position of the probe in all three orientations. To improve the sensitivity down to 0.01%, the cancellation of the planar Hall effect and of the high order of perturbations is required. The method necessitates being able to accurately orient the calibration field, not only along the major axes but also at many angles in between. Furthermore, reliable calibration software that reconstructs coherent calibration parameters — usually nonlinear — based on all these measurements is needed. A considerable effort has to be made to simplify the calibration process using commercially available tools.

5 Calibration of the Hall probes

5.1 Single-axis Hall probe calibration

The calibration is necessary to determine the real output signal of the Hall probe. As recalled in Section 2.4.4, the Hall voltage is not a linear function of B and is a function of the temperature. These functional dependences of the Hall voltage have to be known before starting the measurements. The calibration is the test during which known values of the magnetic field are applied to the Hall probe and the corresponding output reading is recorded. This procedure takes place in all field ranges of interest and is repeated quite often to avoid problems with offset drifts. A successful calibration method consists in using as a reference the magnetic field produced by an NMR teslameter $B\text{-NMR}$. Typically, one needs:

- a Hall probe current source with a high stability;
- a high-accuracy voltmeter to measure the Hall voltage;
- an NMR teslameter with probe(s) to measure the applied magnetic field;
- a calibration magnet with its power supply. The required homogeneity of the magnetic field $\Delta B/B$ is 10^{-4} , 10^{-5} per mm for the locking of the NMR frequency;
- a temperature-stabilized room.

The Hall and NMR probes are placed in the gap of a magnet whose field magnitude and polarity can be changed while the readings of V_H and $B\text{-NMR}$ are recorded. At the Paul Scherrer Institut, the calibration field range is from ± 0.1 T to ± 2 T. The low end of 0.1 T is our measurement field limit of the NMR teslameter (Metrolab PT 2025 with probes 2-5). The high end at 2 T is due to the field non-homogeneity of the calibration magnet (Bruker dipole B-E 30hf), which at this level, is too high for the teslameter to obtain the NMR signal. The Hall and NMR probes are placed in the centre of the pole gap, each on one side of the magnet mid-plane and at equal distance from the

magnet poles (Fig. 31). In this way, both probes experience the same field value. Any possible variation of the magnetic field from mid-plane to pole does not therefore affect the calibration results. The Hall probe needs to be oriented perpendicular to the field, otherwise the measured value will correspond to $B \cos(\gamma)$, where γ is the angle between the field and the axis normal to the probe. An error of 1 degree reduces the V_H reading by 2.10^{-4} . The NMR is insensitive to small-angle errors.



Fig. 31: Left: Hall probe inside a 2 T calibration magnet. The magnet has a gap of 38 mm and a field homogeneity 0.6×10^{-5} in the centre. Right: Measurement of the Hall voltage V_H for a magnetic field read by a NMR probe B -NMR

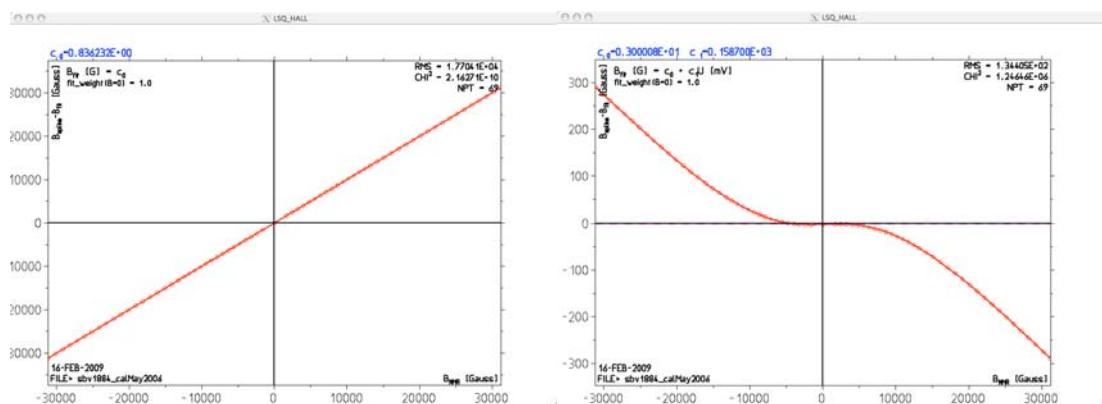


Fig. 32: Left: Hall voltage recorded versus magnetic field read by the NMR. Right: Deviation from the linearity $V_H = f(B)$ in the case of the Siemens SVB601 Hall probe. The non-linearity is less than 0.5% up to 2 T

During the calibration (and afterwards during the magnetic measurements) the environment air temperature should be stabilized and kept constant. While the Hall voltage is not only directly proportional to the magnetic field but also to the Hall probe current, a high -precision constant current source for the Hall probe is required (<100 ppm). In order to improve the accuracy of the V_H reading, a method with averaging of multiple measurements should be used. An integration time of 20 ms in the digital voltmeter (DVM) will filter the 50 Hz noise that is present in every electrical device so it should be chosen if possible. Figure 32 shows an example of deviation from the linearity $V_H = f(B)$ in the case of the single-axis Hall probe Siemens SVB601, used at the Paul Scherrer Institut.

Several methods are used to describe the non-linearity of the field versus Hall voltage. The simplest and easiest to use is

$$B = c_0 + c_1 V_H + c_2 V_H^2 + \dots c_n V_H^n, \quad (28)$$

Where the coefficients c_i are calculated with the least-squares fit of the tabulated measured V_H -Hall/B-NMR data.

The highest order term n is typically between 10 and 15. The higher the order of polynomial, the better the approximation. Above a certain order the improvement is, however, not significant. The highest order term can be determined when the reduced Chi-square χ^2 of the fit is near 1.

$$\chi^2 / ((N - n) B_{err}^2) \approx 1 \quad . \quad (29)$$

Here N is the number of measured pairs (V_H , B -NMR) and B_{err} the field measurement error.

5.2 Calibration of a multi-axis Hall sensor

In the case of a three-axis sensor, the difficulty of performing a precise calibration is increased because of the presence of the mutual coupling between axes. The classical operation consisting in the calibration of the probe along the three main axes is not sufficient to reach a level of precision better than 1% because contributions like the planar Hall effect, which are not symmetric around the axis perpendicular to the Hall probe plane, are zero on these main axes. A dedicated patented calibration device was developed at CERN [36, 37] in order to estimate and cancel out the high-order terms of the Hall voltage. In the proposed calibration technique, the sensor makes a three-dimensional scan, being turned in a constant and homogeneous magnetic field in steps of a fraction of π like $\pi/12$ for example. The operation is repeated for several field strengths and temperatures. Then the output voltage is expressed in orthogonal functions, the spherical harmonics for the spatial dependence, and the Tchebychev polynomials for the magnetic field and temperature dependence. The set of equations (e.g., set of voltages) obtained for the various angular steps and different magnetic fields is solved and reduced to the different components of the decomposition. The correspondence between the main spherical harmonics Y_{lm} and the components of the output voltage are presented in Table 3. This calibration process aims at correcting the errors coming from the non-linearity, the non-orthogonality of the sensors, the temperature sensitivity, and the planar Hall effects. The three spatial components of \mathbf{B} can be reconstructed with a precision of about 0.01%.

Table 3: Coefficients of the output voltage components resulting from decomposition in spherical harmonics. The biasing current flows in the x direction and a classical ‘Hall voltage’ is measured in the y direction

Coefficient Y_{lm}	Significance
Y_{00}	Offset
Y_{10}	Hall sensitivity proportional to B_z
Y_{22}	Planar hall contribution, proportional to the product B_x, B_y
Y_{32}	‘3D Hall effect’, proportional to the product B_x, B_y, B_z

To perform the rotation, a device called ‘the calibrator’, using three pick-up orthogonal coils (coil 1, 2, 3) placed on a rotating platform to measure each component of magnetic field, was built [38]. In the coil support plate, one or several B -sensor cards containing the Hall sensor to be calibrated are fixed with a dowel pin (see Fig. 33). The B -sensor card, build by NIKHEF Amsterdam, contained in that case three Siemens KSY44 single-axis probes glued on a glass cube and all the electronics, in particular a 24-bit ADC. The B -sensor cards were rotated continuously and slowly in a constant and homogeneous field. It was produced either by a magnet at CERN reaching 1.4 T, or for magnetic fields up to 2.5 T by a 130 mm-aperture solenoid at the High Magnetic Field Laboratory of Grenoble (France). The reference value of the magnetic field was obtained from a NMR probe and the angle of rotation measured with a precision of 0.02 mrad. The temperature is stabilized at level of $\pm 0.02^\circ\text{C}$. The Hall probes, calibrated with this procedure, successfully mapped the axial magnetic field of the 2 T solenoid of the ATLAS experiment at the LHC. More details about the reconstruction parameters and the results can be found in Ref. [39]. The recent improvements performed on the calibrator are presented in Ref. [40]. However, the major difficulty remains the size of the apparatus. A suitable

calibration magnet should have a large aperture or gap (~ 100 mm), a good homogeneity on a large scale, and produce a high magnetic field. The use of this technique at the level of mass production implies

- the construction of ‘light’ calibrators equipped with miniaturized sensor cards;
- simplified electronics and reliable calibration software that reconstructs coherent calibration parameters — usually non-linear — based on all these measurements.

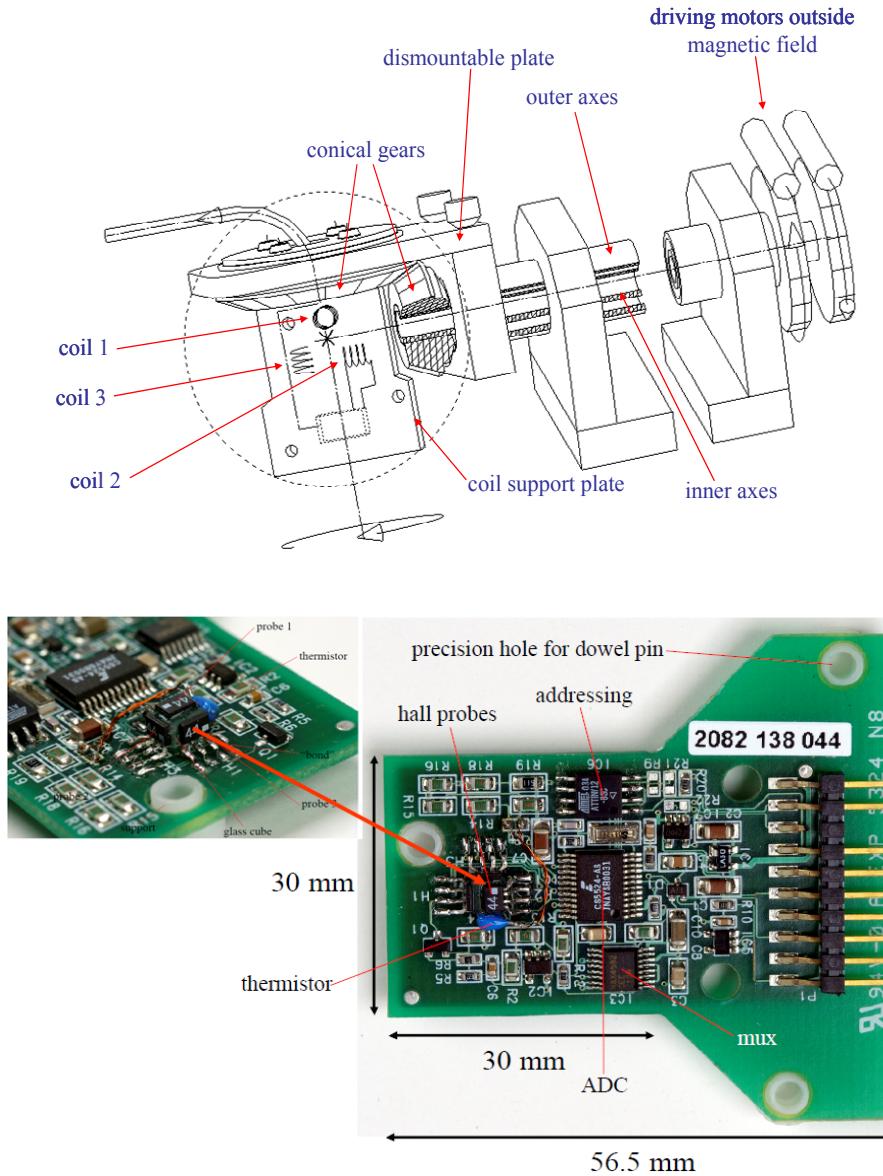


Fig. 33: Calibrator (top) and B -sensor card built by NIKHEF Amsterdam (bottom) containing the Siemens KSY44 probes glued onto a glass cube [38]

6 Application to magnetometry

6.1 Field measurement accuracy

The best method for measuring a magnetic field has to be selected after balancing the requirements of the characteristics of the magnetic field to be mapped [41]. The following parameters have to be considered:

- the quantity to measure, e.g., field components, total field, integral, gradient;
- the field measurement range;
- the reproducibility and accuracy;
- the mapped volume and field geometry;
- the time bandwidth;
- the environment (air, vacuum, cryogenic);
- the use, e.g., research and development, production.

Figure 34 reports a summary of the capabilities of the various methods to measure a magnetic field as discussed in the lectures by L. Bottura and K.N. Henrichsen [16, 42], in terms of measurement accuracy and measurement range. The measurement with Hall probes is particularly suitable for a broad range of magnetic fields up to 10 T or more. The absolute accuracy of the measurement is affected by the parasitic effects discussed in Section 2.4: non-linearity, temperature dependence, alignment errors, and the stability of the offset. Using a reproducible and periodic calibration, a control of the temperature and various compensation techniques, an accuracy value of 10^{-4} can be routinely achieved with a single-axis Hall probe. Commercial multi-axis sensors achieve routinely an accuracy of 10^{-2} . Precautions in the calibration and techniques to reduce the noise and the offset drifts have to be applied to approach the level of 100 ppm.

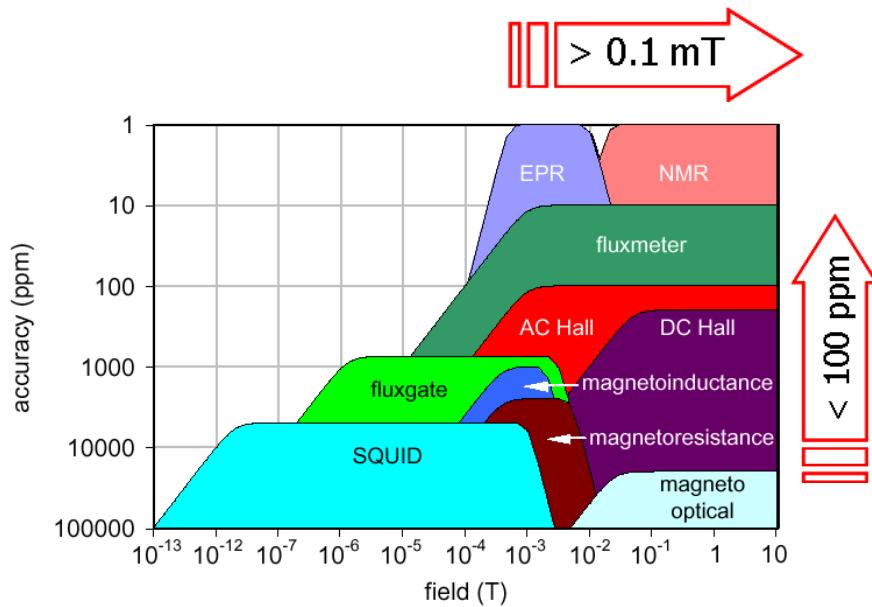


Fig. 34: Accuracy of magnetic field measurements performed with different techniques as a function of the magnetic field level [16, 42]

Going more into details, Table 4 summarizes a list of advantages (+) and drawbacks (-) in magnetic field measurements using a Hall effect device.

Table 4: Magnetometry using a Hall device, advantages and drawbacks

(+)	(-)
<ul style="list-style-type: none"> ▪ Easy to use, easily portable/ moved ▪ Particularly suited for complex geometries ▪ Inexpensive, big market available ▪ Fast measurement (instantaneous response) ▪ Cover a broad range of magnetic fields ▪ Medium accuracy for single component measurement (~0.01%), resolution ~0.5 G ▪ It is a field mapper: it can measure the three components ▪ Can be used for time varying magnetic fields ▪ Works in non-uniform fields ▪ Works at cryogenic temperatures 	<ul style="list-style-type: none"> ▪ Temperature sensitivity ▪ Non-linearity $V_H = f(B)$ ▪ Offset ▪ Drift of the sensitivity with time ▪ Noise ▪ Lower accuracy for multi-axis integrated circuit Hall sensor: 1% to 0.1% ▪ Cross-talk between axes (planar Hall effect, higher order of perturbations) ▪ Calibration (delicate for multi-axis sensors) ▪ Quantum Hall effect (cryogenic temperature)

6.2 Magnetometry in room-temperature beam line magnets

Hall probes are particularly adapted for making:

- point like measurements,
- magnetic field integrals,
- field maps.

As an illustration, the construction of a field map performed at the Paul Scherrer Institut (PSI) in a resistive beam line dipole is briefly presented. The measurement system consists of an automated five-axis positioning bench adaptable with different types of Hall probes (transverse, axial). The probes are positioned such as to measure the main component of the magnetic field. Highly linear Siemens SBV 601-S1 probes with an active area of 2.6 mm² and a sensitivity of 0.06 V/T measure the component of the magnetic field perpendicular to the active region with an accuracy of 0.1 G. Table 5 summarizes the performance of the system.

Table 5: Field measurements using the Siemens SBV 601-S1 Hall probes

Hall Probe	Siemens SBV 601S1
Semicond. material	InAs
I max	400 mA
U_{Hall}	60 mV@1T
Longitudinal range	2100 mm
Horizontal range	650 mm
Vertical range	360 mm
Long./Transv./Vert. Resolution	10 mm
Maximum calibrated Field	3.1 T
Hall Probe absolute accuracy	100 ppm
Hall probe resolution	1 μ T
Temperature sensibility	70 ppm/°C

An overview of the measurement system with a picture of a Hall probe is shown in Fig. 35. The probe is attached to a measuring titanium arm that can move with five degrees of freedom (three directions of translation and rotation around two axes). The device slides on compressed air pads over a flat, precise, machine slide granite block.

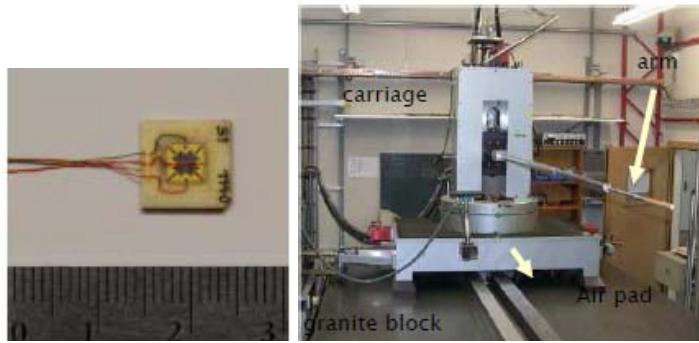
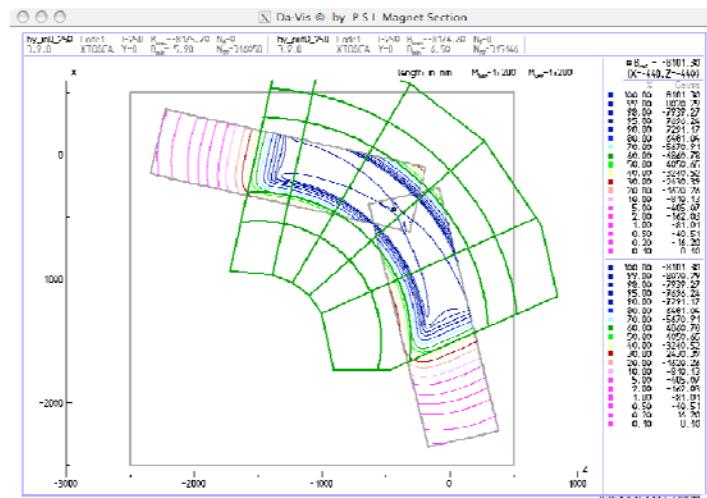


Fig. 35: Left: Single-axis Hall probe. Right: Hall probe measurement system developed at the Paul Scherrer Institut

The measurements are performed in a flying mode at a speed of 50 mm/s. The system is placed in a temperature controlled hutch ($\pm 0.1^\circ\text{C}$). The voltages are recorded using HP/Agilent digital multimeters interfaced to a computer via a GPIB gateway. The data are post-processed off line using the calibration procedure described in Section 5.1. For the analysis of a beam line magnet, the knowledge of the main field component is often enough. In the magnetic mid-plane of dipoles, quadrupoles, sextupoles and solenoids, the field mapping of the main component in the magnetic mid-plane provides information like:

- effective magnetic length,
- edge angles,
- edge curvatures,
- beam bending angle,
- beam focusing strength,
- beam vertex position,
- magnetic axis.

Figure 36 shows the results for the main component measured in the mid-plane of a 90° bending magnet. It has been measured from each end and the field maps are here rotated with appropriate angle to each other and put together on the same plot. The magnet outlines are also plotted in thick lines.



The other field components off the mid-plane are mathematically describable. If necessary, and provided there is good accuracy of the measured main field component, the other two field components could be reasonably well calculated. Two methods are used for creating the full field map in a volume, by extrapolation of the mid-plane fields (already used in other programs like TRANSPORT/TURTLE) and by interpolation between the fields in the mid-plane and two other planes (method developed by the PSI Magnet Section). The extrapolation method is based on the Taylor expansion, therefore relying on high-order derivatives, i.e., the differentiations of the mid-plane fields. All field components off the mid-plane are calculated and the measurement error propagates fast with the distance from the mid-plane. The interpolation method is less sensitive to measurement error and produces more accurate results. The main field component is approximated with an even-order function given by

$$By(x,y,z) = B_0(x,z) + B_2(x,z) y^2 + B_4(x,z) y^4; (y=0 \text{ is the mid-plane}). \quad (30)$$

This equation is solved for every (x,z) point from the main field component B_y measured in three planes yielding the coefficients B_0 (equal to the measured field), B_2 and B_4 . The scalar potential at any point in space is described by

$$V(x,y,z) = B_0(x,z) y + B_2(x,z) y^3/3 + B_4(x,z) y^5/5. \quad (31)$$

The first derivative of potential in any direction is the field component in that direction, thus containing the calculation error to the lowest possible differentiation.

The B_x and B_z field components calculated with the interpolation method on a plane $y = 5 \text{ cm}$ are shown in Fig. 37 (left and right).

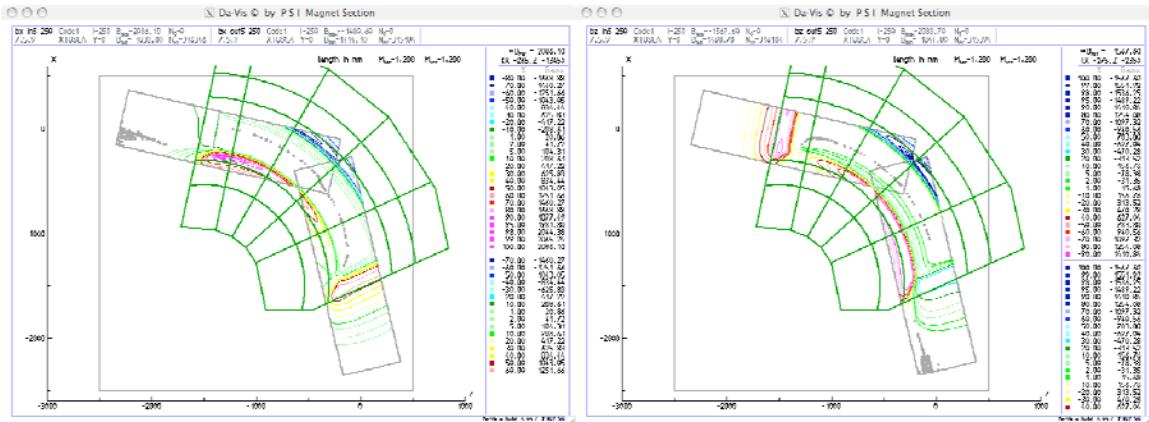


Fig. 37: Transverse field components, B_x (left), B_z (right)

6.3 Hall probe measurements in undulators

6.3.1 Introduction

Undulators are powerful generators of synchrotron radiation at storage rings. They are also one of the key components in the production of laser-like radiation down to the angstrom wavelength regime with X-ray free electron lasers. Details concerning the science and the technology of undulators can be found in Refs. [B10, B11, B12] and are also presented in these proceedings [43]. We recall briefly the minimum of information to help in their understanding. Insertion devices can generate a sinusoidal planar field with high peak intensity (up to 2 T) and short period using a periodic magnetic field produced by arrays of magnets of alternating polarity. The electrons ‘wiggle’ back and forth and the acceleration produced causes the emission of radiation at each pole (see Fig. 38).

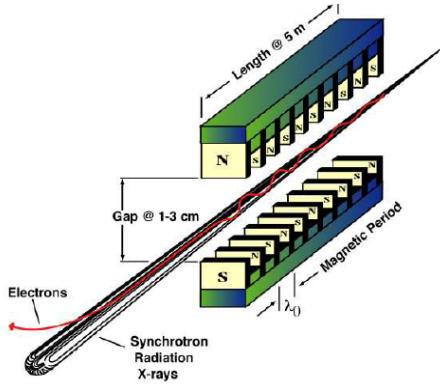


Fig. 38: Schematic of a planar undulator

In the case of undulators, the light emitted at each pole interferes with that emitted from others and a quasi-monochromatic spectrum will be observed. At a resonance wavelength λ given by Eq. (32), interference occurs because the emitted light overcomes the electrons by exactly one wavelength. λ is related to the geometric period length λ_0 of the dipole arrangement but shrunk by $1/2\gamma^2$ ($\gamma = E/E_0$ is the Lorentz factor). This reduction is caused by the conjunction of two relativistic effects. A first shrinking factor of $1/\gamma$ is due to the Lorentz contraction, the electrons see the Lorentz contracted period. Another $1/2\gamma$ results from the Doppler effect when the radiation emitted in the reference frame of the electrons is observed in the laboratory frame. The produced wavelength depends on the period and the energy, but also on the deflection parameter K defined in Eq. (32) and on the observation angle Θ . The number n is the harmonic number of the emitted radiation ($n = 1; 3; 5; \dots$) at the undulator axis.

$$\lambda = \frac{\lambda_0}{2n\gamma^2} (1 + K^2/2 + \gamma^2 \Theta^2) \quad (32)$$

$$K = 0.0934 B_0 \lambda_0 \quad (B_0 \text{ is in T, } \lambda_0 \text{ in mm}) . \quad (33)$$

Here K is related to the magnet periodicity and to B_0 , the peak magnetic field on the undulator axis. As shown by Eqs. (32) and (33), λ can be tuned by adjusting the magnetic field. This adjustment is done by changing the magnetic gap between the lower and upper arrays of the magnet blocks (see Fig. 39). When the undulator is linearly polarized, electrons wiggle in one plane. Undulators can also be elliptically polarized, so electrons travel in helixes generating elliptically polarized light. The case of $K_z = K_x$ corresponds to a circular polarized light.

Technologies used to build the magnets for the undulator are:

- The Pure Permanent Magnet (PPM) technology:
Magnets are made of NdFeB (remanent field $B_r = 1.2\text{--}1.4$ T) or of Sm₂Co₁₇ ($B_r = 1.05$ T). To generate a sinusoidal field, two arrays of permanent magnets should ideally be positioned with the easy axis rotating through 360° per period along the direction of the electron beam. In practice, a good approximation is done by splitting the system into rectangular magnet blocks, M per period (see Fig. 39).
- The Hybrid technology:
The vertical magnetized magnets are replaced by high permanent iron poles to increase the on-axis flux density.
- The electromagnet technology for long period undulators.
- The superconducting technology with NbTi as superconductor, for fields higher than 2 T.

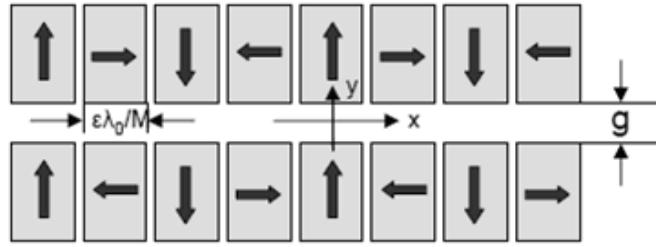


Fig.39: Approximation of a permanent magnet undulator by splitting the system into rectangular magnet blocks, four magnet blocks per period (from Ref. [43])

6.3.2 Magnetic field errors in permanent magnet undulators

The maximum field strength and the characteristics in terms of field errors are figures of merit of the undulator (for a given period length and a given gap). Field errors from permanent magnet undulators are not iron dominated. They come from:

- machining tolerances,
- inhomogeneities of magnetization within the magnet block,
- error in positioning and orientation of the blocks.

Field errors are divided into two categories: the field integral errors and the phase errors.

Knowledge of the first and the second field integral (see formulas in Fig. 40) is essential but not sufficient to characterize the magnetic field in the undulator. Measurements of these two quantities give information of the total change in angle and in position of the beam trajectory at the exit of the device. The ideal undulator is transparent to the electron beam, the first and second field integrals in the undulator vanish with zero errors in the block-to-block magnetization and block positioning. In practice, for a better beam stability, it is important to tune the undulator to smaller values. The integral measurements are usually performed using flipping coils, moving or pulse wire techniques, see Refs. [44, 45].

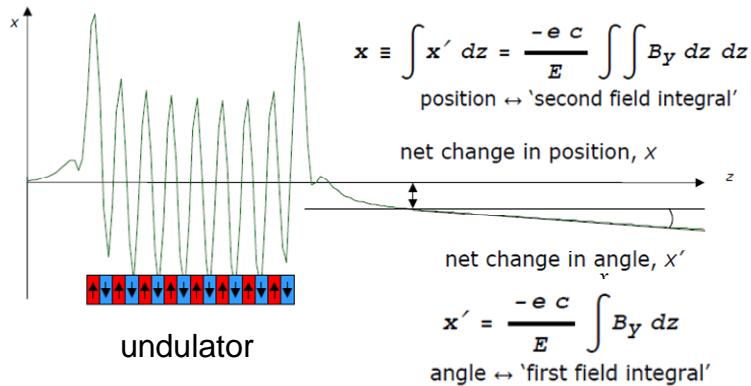


Fig. 40: First and second field integral measured along an undulator producing a magnetic field in the y direction (from Ref. [44]). The first field integral gives information on the angle and the second field integral on the position of the electron at the exit of the undulator

Knowledge of the local shape of the magnetic field along a straight line close to the electron trajectory is also important. It provides information about another class of error, the phase error. The

phase error indicates the deviation from perfect matching in phase between the electron and the emitted radiation from pole to pole. The slippage between the electron and the light is exactly one period of the emitted radiation (λ) when one period of the undulator (λ_0) has been traversed for an ideal magnetic field, i.e., the electron falls behind the light by one period (λ). A phase error of ± 1 degree, for instance, indicates that there is an offset by $\pm 1/360$ from one period to the next. It can be characterized as the error in the periodicity of the produced electrical field (see Fig. 41) and has to be kept small to avoid an incomplete constructive interference. A definition of the phase error can be found in Ref. [46]. It is related to the square of the first field integral via the following formula:

$$\varphi(z) = \frac{2\pi}{\lambda_0} \left[\frac{1}{(1 + K^2/2)} (z + e^2/m^2 c^4 J(z)) - z \right], \quad (34)$$

where $J(z)$ is the integral of the square of the first field integral

$$J(z) = \int_0^z \left[\int_0^{z'} B_y(z') dz' \right]^2 dz'. \quad (35)$$

The phase error will attenuate and spread the angular spectral flux on the peak of the spectrum. The origins of this field imperfection are the magnetic peak field fluctuations, the period and the field shape fluctuations. These errors are deduced from local magnetic measurements, performed most widely with a Hall device. After being measured, this error is corrected by shimming.

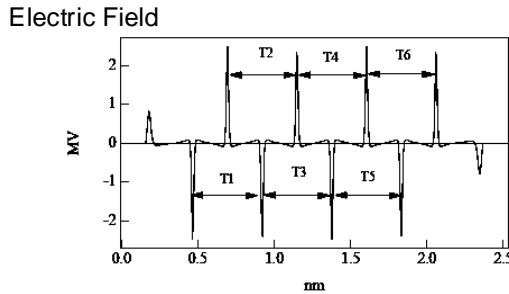


Fig. 41: Effect of the phase error observed in the horizontal electrical field produced on-axis by a single electron (from Ref. [47])

6.3.3 Specificity of magnetic measurements in permanent magnet undulators

The choice of the Hall sensor, the design of the Hall probe bench, and the measurement process have to be adapted in the case of undulator measurements, see Refs. [47, B11]. Insertion devices specifically feature a small gap (several millimetres), a great length (several metres) and a field shape with large longitudinal gradient. This results in the following remarks:

- The measuring device has to be sufficiently small to be inserted in undulators with a small gap. Hall probes with a small size are also preferred for undulators with very short periods. A drastic size reduction will affect the sensitivity of the Hall element.
- On-the-fly scanning is essential to reduce the sensor vibration and the measurement time.
- Temperature stabilization is not critical if the temperature does not change during the measurement over the length of the undulator.

- A high level of accuracy in the Hall probe position on the longitudinal axis is critical, in particular for undulators producing a strong magnetic field gradient.
- The measurement of the three magnetic field components is useful and sometimes required. The measurement accuracy will, however, be limited by the planar Hall contribution. In a device with a vertical field, the determination of the horizontal field integral from numerical integration of the local field can be largely offset. The error is larger in the case of helical undulators. A three-dimensional sensor with a low cross-coupling between the axes (small planar coefficient) should be selected and compensation techniques have to be applied. Investigations on planar Hall effect compensation were carried out on the undulators for the FEL project at the Advance Photon Source (Argonne National Laboratory, USA). Horizontal magnetic field measurements in presence of a vertical field component with a strong gradient were performed [48], using a two-axis Sentron Hall probe based on vertical devices. It was also reported that the planar Hall voltage can be successfully cancelled using a two-sensor device, one Hall probe positioned at 90° with respect to the other [49]. Cancellation techniques like the switching of the voltage/current terminals explained in Section 4.4.4 can be used but will require two undulator scans. At the ESRF (Grenoble, France), the calibration of the Hall probe includes the determination of the planar Hall coefficient, by measuring the three components of the field at various positions in known field geometry of the undulator [50]. The Apple 2 undulators of PETRA 3 (DESY, Germany) were measured using a three-dimensional IC Hall sensor [43].

6.3.4 Hall probe measurements in undulators at the Paul Scherrer Institut

An example of local field measurements performed in undulators of the Swiss Light Source at the Paul Scherrer Institut is presented for illustration. The design of the Hall probe bench was performed at the ESRF [51]. As displayed in Fig. 42, the bench is based on a granite support mounted on vertically adjustable feet featuring a cross section of 350 x 600 mm² (width, height). The measuring bench is 4 m long and has an effective length capable of hosting insertion devices with up to 3.5 m magnetic length. The upper granite surface has a flatness of 15 µm and is equipped with guide rails assembled with a parallelism of 5 µm. An Anorad linear motor equipped with a Heidenhain encoder is directly mounted on the granite. The longitudinal position is read out via an optical linear encoder calibrated with a laser interferometer. A carriage is mounted on the guide rails and includes two transverse stages with a travel range of 250 mm. The sensor consists of three one-dimensional Bell Hall elements mounted on a print board (see Fig. 42), fully calibrated up to 1.8 T with a NMR reference magnet. The probes feature a sensitivity of 1 T/V and a residual non-linearity of 0.05% after correction. The bias current is 5 mA. The Hall voltage is read by three Keithley precision voltmeters and the output noise for a 20 ms integration time is less than 0.06 G. They measure all three components of the magnetic field along the aperture of the insertion device under test with an accuracy of 2×10^{-4} T. The sensor position is known within 20 µm after software correction based on laser calibration. The functionalities of the measuring bench were implemented in C and C++ with an interface made in Igor PRO. The magnetic field measurements are post-processed with the SRW code developed at ESRF [52].

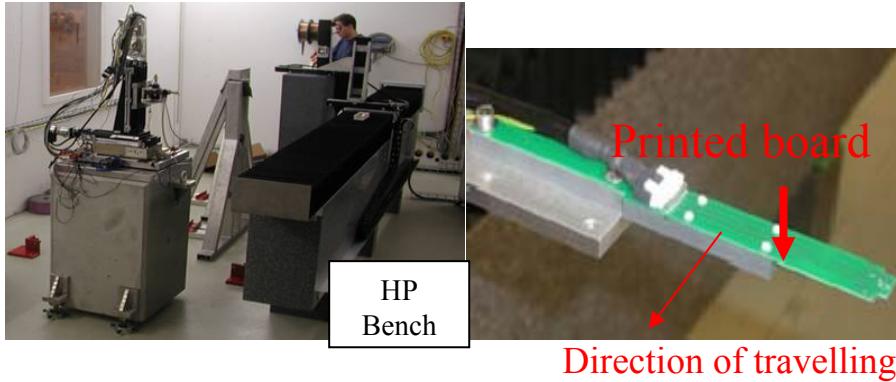


Fig. 42: Left: View of the Hall probe bench. Right: Hall probe keeper holding the printed circuit with the Hall elements

The Hall probes travel along the undulator at a typical speed of 10 mm/s. Some 2500–5000 points are recorded. The typical shape of the magnetic field as a function of the longitudinal position measured in a planar undulator is presented in Fig. 43. The first field integral is calculated from the data with an accuracy of 0.5 Gm but has to be calibrated with the result obtained by another measuring method because of the non-linearity of the probes. This measurement aims to investigate the existence of random local angular ‘kicks’ along the undulator axis. The post-processing of the data results in the measurement of the phase error along the undulator axis (see example in Fig. 43).

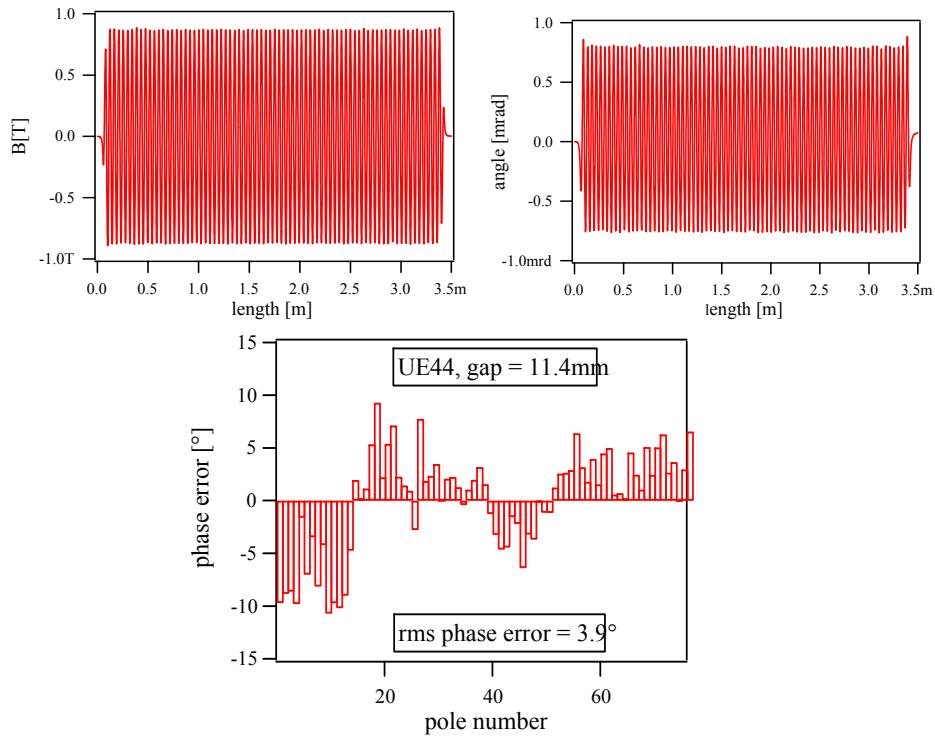


Fig. 43: The first integral and the longitudinal variation phase error can be calculated from the measured field along the longitudinal direction

7 Conclusions

This lecture reviewed the properties of Hall generators and their use in magnetic field measurements. Hall probes are particularly adapted for point, high speed, and large magnetic field volume

measurements. It also offers the advantage of using simple measuring arrangements and low costs. Single-axis Hall probes typically offer an accuracy in the range of 100 ppm. They can be used at cryogenic temperatures and at high magnetic fields. Drawbacks are the need for periodical calibrations and for temperature control, and the existence of parasitic contributions like the planar Hall voltages that have to be compensated. Considerable improvements have been made to reduce the size of the magneto-transducers and develop relatively easy-to-use systems to measure all three magnetic field components in a small spot. The evolution brought by the integration of electronics in vertical (silicon) Hall elements results in the development of the so-called integrated circuit Hall sensors, capable of performing a 3D field mapping in a sub-millimetre size. Techniques have been developed to minimize the offset value and its drift: the planar Hall effect and the $1/f$ noise. The development of these multi-axis micro-sensors has extended the application of Hall effect sensors, for example, in domains like geomagnetism, spatial positions of objects, contactless automation. A new generation of Hall magnetometers based on IC sensors is now available, see Refs. [29, 30, 41, 53]. They are compact and flexible and are developed with standardized interfaces. For magnetometry, however, the drawback of the three-axis IC sensor is the limited accuracy w.r.t. a conventional single-axis Hall probe. The challenges will therefore concern:

- The improvement, by at least one order of magnitude, of the accuracy of such multi-axis sensors. They feature routinely an accuracy between 0.1% and 1% of the measurement range.
- The ability to perform regular three-dimensional calibrations using standard procedures, interfaces and analysis software.
- The capacity to cover a wider range of magnetic fields, in particular the low field domain for medical applications.
- The lowering of the cost that will enable customers to use three-axis instruments instead of the ‘classical’ single-axis approach.

8 Acknowledgements

The author would like to thank V. Vrankovic and T. Schmidt for their contribution to this lecture. The author is grateful to P. Keller (Metrolab Instruments Geneva), to D. Popovic-Renella and R. S. Popovic (Senis Gmb Zürich) for the valuable discussions concerning the three-axis Hall sensors and for the products they allowed me to show as examples during the lecture. The author is also grateful to D. George and M. Calvi for having the patience to correct this document.

References

- [1] E. H. Hall, On a new action of the magnet on electric currents, *Am. J. Math.*, 2, pp. 287–292, 1879.
- [2] Proc. International Magnet Technology Conferences, MT-1 (1965) to MT-20 (2007).
- [3] Proc. International Magnetic Measurement Workshops, MMW-1 (1981) to IMMW-16 (2009).
- [4] Sensors and Actuators A, Physical, Elsevier Science.
- [5] B. Berkes, Hall generators, CERN Accelerator School on Magnetic Measurement and Alignment, Montreux, Switzerland, 1992, CERN-92-05, 167–192.
- [6] J. Boettger, The Hall effect gaussmeter, Sensor Magazine, December 1997, <http://archives.sensorsmag.com/articles/1297/he1297/main.shtml>.
- [7] J. Kviticovic, Hall generators, CERN Accelerator School on Measurement and Alignment of Accelerator and Detector Magnets, Anacapri, Italy, 1997, CERN-98-05, 233–249.
- [8] E. Schurig, Highly sensitive vertical Hall sensors in MOS technology, Thesis 3134, EPFL, Switzerland, 2004.

- [9] C. Goldberg and R. E. Davis, New galvanomagnetic effect, *Phys. Rev.* 94, pp. 1121–1125, 1954.
- [10] B. Turck, Evaluation précise des 3 composantes d'un champ magnétique à partir des tensions prélevées d'une sonde de Hall, *Nuclear Instruments and Methods*, 105, pp. 97–10, 1972.
- [11] E. H. Putley, *The Hall Effect and Related Phenomena* (Butterworth, London, 1960).
- [12] D. Swoboda, The polar coordinate magnetic measuring system for the axial field spectrometer magnet at CERN, *CERN ISR-Boom/81-16*, 1981.
- [13] B. Berkes, New system for precise Hall probe measurements for low intensity magnetic fields, IMMW16, 2009, available at <http://immw16.web.psi.ch/Presentations/>.
- [14] I. B. Vasserman, Cancelling the planar Hall effect voltage by interchanging connections on a Hall sensor, IMMW16, 2009, available at <http://immw16.web.psi.ch/Presentations/>.
- [15] K. von Klitzing, G. Dorda, and M. Pepper, New method for high-accuracy determination of the fine-structure constant based on quantized Hall resistance, *Phys. Rev. Lett.* 45, pp. 494–1402, 1980.
- [16] L. Bottura and K. N. Henrichsen, Field measurement methods, CAS School on Superconductivity and Cryogenics for Particle Accelerators and Detectors, Erice, Italy, 2002, CERN-2004-008, pp. 118–148.
- [17] R. S. Popovic, Not-plate-like Hall magnetic sensors and their applications, *Sensors and Actuators A*, 85, pp. 9–17, 2000.
- [18] Ch. Roumenin and P. Kostov, Planar Hall effect device, Bulgarian Patent N° 37208, 1983.
- [19] R. S. Popovic, The vertical Hall-effect device, *IEEE Electron Devices Lett.* 5, pp. 357–358, 1984.
- [20] C. Schott, Accurate magnetic field transducer based on silicon Hall sensors, Thesis 1985, EPFL, Switzerland, 1999.
- [21] K. Maenaka, T. Ohgusu, M. Ishida, and T. Nakamura, Novel vertical Hall cells in standard bipolar technology, *Electronics Letters* 23, pp. 1104–1105, 1987.
- [22] O. Brand and G. K. Fedder, *CMOS-MEMS Advanced Micro and Nanosystems*, vol. 2 (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2005), pp. 1–57.
- [23] E. Schurig et al., A vertical Hall device in CMOS high voltage technology, *Sensors and Actuators A*, 97–98, pp. 47–53, 2002.
- [24] Procedure for the production of Hall elements, European Patent No. 1045 461 A2.
- [25] U. Falk, A symmetrical vertical Hall effect device, *Transducers'89*, International Conference on Solid-State Sensors and Actuators, Montreux, Switzerland, vol. 2, Ed. S. Middelhoek (Elsevier Sequoia, Lausanne, 1990).
- [26] Z. Randjelovic, A. Pauchard, Y. Haddab, and R. S. Popovic, A non-plate like Hall sensor, *Sensors and Actuators A* 76, pp. 293–297, 1999.
- [27] G. Boero, M. Demierre, P.A. Besse, and R. S. Popovic, Micro-Hall devices: performance, technologies and applications, *Sensors and Actuators A*, 106, pp. 314–320, 2003.
- [28] P. Kejik, E. Schurig, F. Bergsma, and R. S. Popovic, First fully CMOS-integrated 3D Hall probe, *Transducers 05*, International Conference on Solid-State Sensors, Actuators and Microsystems, Seoul, Korea, 2005 (IEEE, Piscataway, NJ, 2005), pp. 317–320.
- [29] D. Popovic et al., Three-axis teslameter with integrated Hall probe, *IEEE Trans. Instrum. Meas.*, Vol. 56, pp. 1396–1402, 2007.
- [30] P. Keller, A new generation of Hall magnetometers: enabling technologies, *Magnetic Business and Technologies*, pp. 19–21, June/July 2008.

- [31] J. Maupin and M. Geske, The Hall effect in silicon circuits, Symposium on the Hall Effect and its Applications, Baltimore, 1979, Eds. C. L. Chien and C. R. Westgate (Plenum, New York, 1980).
- [32] P. J. A. Munter and D. C. van Duyn, Spatial symmetry of transduction effects in Hall plates, Sensors and Actuators A, 31, pp. 206–209, 1992.
- [33] A. Bakker and J. H. Huijsing, A CMOS spinning-current Hall sensor with integrated submicrovolt offset instrumentation amplifier, Proceedings of ProRISC99/SAFE99, Mierlo, The Netherlands, 1999 (Utrecht, 1999), CD-ROM, pp. 17–20 (ISBN: 90-73461-18-9).
- [34] Ch. Schott, P.-A. Besse, and R.S. Popovic, Planar Hall effect in the vertical Hall sensor, Sensors and Actuators, 85, pp. 111–115, 2000.
- [35] P. Keller, Metrolab Instruments, Geneva, Switzerland, private discussion, 2009.
- [36] F. Bergsma, Calibration of Hall sensors in three dimensions, IMMW13, 2003, available at <http://www-conf.slac.stanford.edu/immw/>.
- [37] F. Bergsma, Progress on the calibration of the 3D Hall probes, IMMW14, 2005, available at <http://immw2005.web.cern.ch/>.
- [38] F. Bergsma, 3D Hall sensor applications, IMMW15, 2007, available at <http://immw2005.web.cern.ch/>.
- [39] M. Alekса et al., Measurement of the ATLAS solenoid magnetic field, J. Instrum. 3 S08003, 2008.
- [40] F. Bergsma, Developments in field mapping at CERN, IMMW16, 2009, available at <http://immw16.web.psi.ch/Presentations/>.
- [41] P. Keller, Technologies for precision magnetic field mapping, available at <http://magnetismagazine.com/images/Presentations/Philip%20Keller.pdf>.
- [42] L. Bottura, Overview of measurement methods, Proc. CERN Accelerator School Magnets, Bruges, Belgium, 2009.
- [43] J. Bahrdt, Permanent magnets including wigglers and undulators, Proc. CERN Accelerator School Magnets, Bruges, Belgium, 2009.
- [44] B. Shepherd, Magnetic measurements of insertion devices, Lecture at Accelerator Science and Technology Centre, 2003.
- [45] G. Rakowsky, Measurement of small-gap insertion device, IMMW12, 2001, available at <http://www.esrf.eu/Accelerators/Conferences/IMMW12/present/Rakowsky.pdf>.
- [46] F. Ciocci, G. Dattoli, E. Sabia, and M. Sassi, Comments on phase error for Accel undulator factory test, SPARC-FEL-07/001 note, 2007.
- [47] P. Elleaume, Specificity of magnetic measurement for insertion devices, IMMW12, 2001, available at <http://www.esrf.eu/Accelerators/Conferences/IMMW12/present/Elleaume.pdf>.
- [48] I. Vasserman, Test of horizontal magnetic field measurements in the presence of a strong vertical field, FEL Conference 2004, Trieste, Italy, pp. 527–529.
- [49] I. Vasserman, B. Berkes, J. Xu, and J. Kvitkovic, Compensation of the planar Hall effect voltage using a new two-sensor Hall probe design, Proceedings PAC09, Vancouver, BC, Canada (URL <http://trshare.triumf.ca/~pac09proc/Proceedings/papers/we5rfp062.pdf>).
- [50] J. Chavane and P. Elleaume, The ESRF insertion devices, J. Synchr. Rad. 5, pp. 196–201, 1998.
- [51] J. Chavane and C. Penel, ESRF insertion device field measurement bench, ESRF note, available at <http://www.esrf.eu/Industry/files>.
- [52] O. Chubar, P. Elleaume, S. Kuznetsov and A. Snigirev, Physical optic computer code optimized for synchrotron radiation, available at http://ftp.esrf.eu/pub/UserReports/24682_A.pdf.

- [53] S. Lozanova, S. Noykov and Ch. Roumenin, A novel 3-D Hall magnetometer using subsequent measurement method, Sensors and Actuators A, 153, pp. 205–211, 2009.

Bibliography

Many books refer to the Hall effect device. Those of this list were used to prepare this lecture. A large amount of the information and pictures in this paper originates from Ref. [B1].

- [B1] R. S. Popovic, Hall Effect Devices, second edition (Institute of Physics Publishing, Bristol, 2004).
- [B2] E. Ramsden, Hall Effect Sensors – Theory and Application (Newnes, Oxford, 2006).
- [B3] S. C. Mukhopadhyay and R. Yueh-Min Huang, Sensors: Advancements in Modeling, Design, Issue, Fabrication and Practical Applications (Springer, Berlin, 2008).
- [B4] S. Y. Yurish and M. T. S. R. Gomes, Smart Sensors and MEMS (Kluwer, Dordrecht, 2004).
- [B5] Ch. S. Roumenin, Microsensors for magnetic fields, *in* MEMS a Practical Guide to Design and Analysis and Applications, Eds. J. G. Korvink and O. Paul (Andrew Publ., 2005), pp. 453–520.
- [B6] Honeywell, Hall Effect Sensing and Application, MICRO SWITCH Sensing and Control manual, available at
<http://content.honeywell.com/sensing/prodinfo/solidstate/technical/hallbook.pdf>.
- [B7] J. Fraden, Handbook of Modern Sensors: Physics, Designs, and Applications, third edition (Springer, New York, 2003).
- [B8] P. Ripka and A. Tipek, Modern Sensors Handbook (ISTE, London, 2007).
- [B9] P. Ripka, Magnetic Sensors and Magnetometers (Artech House, Boston, 2001).
- [B10] J. A. Clarke, The Science and Technology of Undulators and Wigglers (Oxford University Press, 2004).
- [B11] H. Onuki and P. Elleaume, Undulators, Wigglers and their Applications (Taylor & Francis, London, 2003).
- [B12] P. Schmüser, M. Dohlus, and J. Rossbach, Ultraviolet and Soft X-ray Free-Electron Lasers: Introduction to Physical Principles, Experimental Results, Technological Challenges (Springer, Berlin, 2008).

Magnet stability and reproducibility

N. Marks

ASTeC, STFC, Daresbury Laboratory and University of Liverpool,
The Cockcroft Institute, Daresbury, Warrington WA4 4AD, U.K.

Abstract

Magnet stability and reproducibility have become increasingly important as greater precision and beams with smaller dimension are required for research, medical and other purpose. The observed causes of mechanical and electrical instability are introduced and the engineering arrangements needed to minimize these problems discussed; the resulting performance of a state-of-the-art synchrotron source (Diamond) is then presented. The need for orbit feedback to obtain best possible beam stability is briefly introduced, but omitting any details of the necessary technical equipment, which is outside the scope of the presentation.

1 Introduction

Magnet stability and reproducibility clearly are important issues that strongly affect the performance of accelerator magnets and influence their effectiveness in interacting with the beam as required. Skilful design and high-quality engineering are worthless if variations of the magnet performance over time (i.e., lack of stability) or between the different elements in lattice (i.e., poor reproducibility) have a deleterious effect on beam properties. Notwithstanding, the quest for ultra-high stability is a relatively recent phenomenon that has appeared with the advent of larger and larger accelerators generating beams with smaller and smaller cross-sections. The subject is also quite specialized, with a relatively small number of individuals practising their expertise in this area; and it spans a wide range of sciences and technologies, for example, from geology and seismology, to mechanical and electrical engineering and, of course, accelerator theory.

For these reasons, the presentation and this summary article rely heavily on published material or private communication from a number of experts in this field, who are acknowledged at the end of the paper.

2 Beam stability requirements

Starting at the delivery end of the accelerator complex, the degree of beam stability required at the interaction point (whether on a target, a sample, at a patient, or on another beam), depends very much on the purpose of the accelerator. In this paper the requirements typical of some of the most demanding applications — for example, particle physics research and materials analysis — will be considered. In all facilities, the beam stability is an important component of the accelerator specification, but the areas cited perhaps present the biggest challenges. It is clear that the higher precision demanded by the research community can be met by reducing beam size only if the beam position and angle is correspondingly stable and movement and deflection is restricted to some small fraction of the beam emittance.

The resulting demand for magnet stability and operational reproducibility then depends substantially on the lattice parameters; this issue will be addressed in Section 3, whilst the parameters that drive these requirements — beam geometries and required positional and directional stability — are detailed below, with examples chosen from two existing facilities and a proposed future project.

2.1 A ‘state-of-the-art’ synchrotron source — Diamond

The magnet lattices of the electron storage rings, that are the core of a synchrotron source, have evolved significantly since the first dedicated sources were designed with simple FODO structures in the 1970s. The state-of-the-art facilities now have specialized straight sections housing insertion devices, and many quadrupole families producing beams with very small cross-sections. These small dimensions produce very high brightness photon beams for experimenters but, correspondingly, the beam position and angle at each source point needs to be highly stable.

The situation is exemplified by ‘Diamond’ the 3 GeV synchrotron source recently commissioned in the UK. This storage ring has a complex lattice of 48 dipoles and 240 quadrupoles [1], giving the stored electron beam the horizontal and vertical β values shown in Fig. 1.

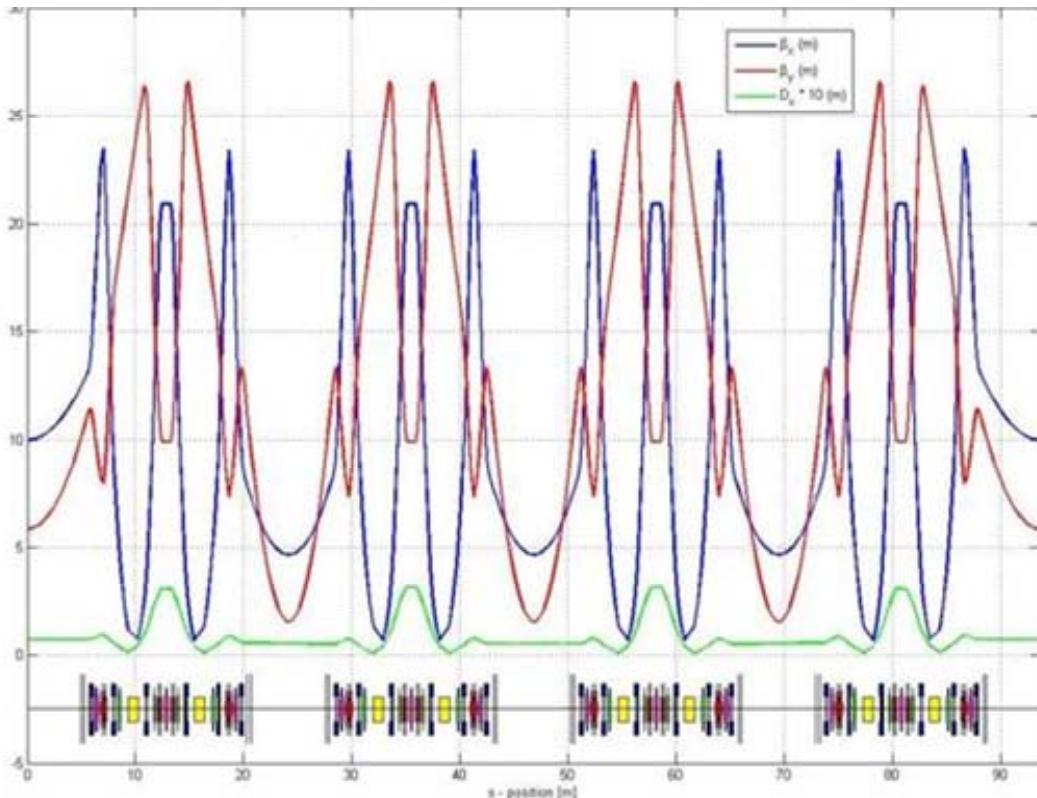


Fig. 1: Part of the Diamond magnet lattice showing the curves of β_x (blue) and β_y (red)

It can be seen that the lattice focuses the beam in both planes in the straight sections; the beam dimensions and divergences at these points are given in Table 1, together with the beam emittances in the lattice.

Table 1: Electron beam parameters in the third-generation (2005) 3 GeV synchrotron source, Diamond

	Horizontal	Vertical
Beam size fwhh (μm)	123	6.4
Beam divergence (μrad)	23	4.2
Beam emittance (nm rad)	2.7	0.03

In the Diamond project, it was determined that the stability of the electron beam at the photon source points needed to be 10% of the beam dimension figures above [2], giving the tolerances on electron beam stabilities given in Table 2.

Table 2: Electron beam stabilities required in the 3 GeV synchrotron source, Diamond

	Horizontal	Vertical
Beam positional variation (μm) \leq	12.3	0.64
Beam angular variation (μrad) \leq	2.3	0.42

2.1.1 Comparison with an earlier SR source

It is of interest to compare the Diamond beam dimensional data with that from the Daresbury SRS, an early dedicated source initially commissioned in the 1970s and upgraded some ten years later to meet the requirements for a second-generation light source. The upgraded lattice gave the electron beam parameters of Table 3.

Table 3: Electron beam sizes and emittance in the 2 GeV second-generation (1988) source, SRS

	Horizontal	Vertical
Beam size fwhh (mm)	2.6	0.24
Beam emittance ($\mu\text{m rad}$)	0.11	

It can be seen that emittances had been reduced by over two orders of magnitude in less than twenty years, with correspondingly increased demands on beam stability. Given the same criterion as applied to Diamond, the tolerance on beam movement in the SRS was of the order of 0.3 mm horizontally and 0.024 mm vertically — values that could be achieved at that time with conventional technology.

2.2 The CERN LHC

The Large Hadron Collider (LHC) is, at the time of writing, being commissioned at CERN. The storage ring will collide beams travelling in opposite directions, with beam cross-sections in the interaction regions determining the experimental luminosity. The expected sizes of a 7 TeV beam in the region of the CMS interaction point [3] are shown in Fig. 2. The expected r.m.s. values at a number of collision points are given in Table 4.

Table 4: Expected r.m.s. beam sizes at interaction points in the LHC

CMS & ATLAS (protons)	16 μm
LHC b (protons)	22–160 μm
ALICE (ions)	16 μm
ALICE (protons)	> 160 μm

It can be seen that at the narrowest interaction points, the LHC beam will be approximately an order of magnitude smaller than the horizontal electron beam size in Diamond, resulting in a requirement for correspondingly greater stability.

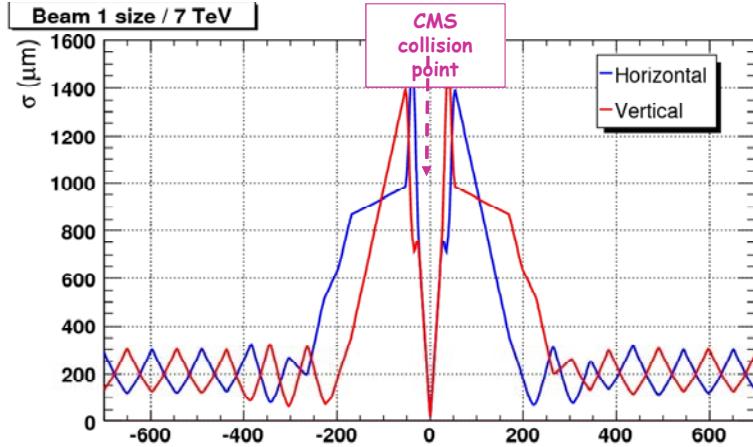


Fig. 2: Expected rms beam sizes at 7 TeV at the CMS collision point in the CERN LHC

2.3 The proposed International Linear Collider (ILC)

Considerable work has already taken place to design a future particle physics facility that will collide 250 GeV electrons together, with an upgrade path to 500 GeV per beam; a diagram of the proposed layout is given in Fig. 3.

Whilst the project is yet to be funded, the critical issue of the required electron beam dimensions has been addressed [4]. The cited presentation explains:

'After acceleration, the beams are not ready to deliver the full luminosity required for the physics studies – the beam size must be reduced from 1 micrometre at the end of the acceleration unit to just a few nanometres at the interaction point (they) need to go through several optics correction points....'

The presentation does not indicate the tolerance that this microscopic beam places on positional and angular stability but it is clear that any movement of the same order as the total beam size would inhibit the collision process, destroy the luminosity, and therefore would be unacceptably large.

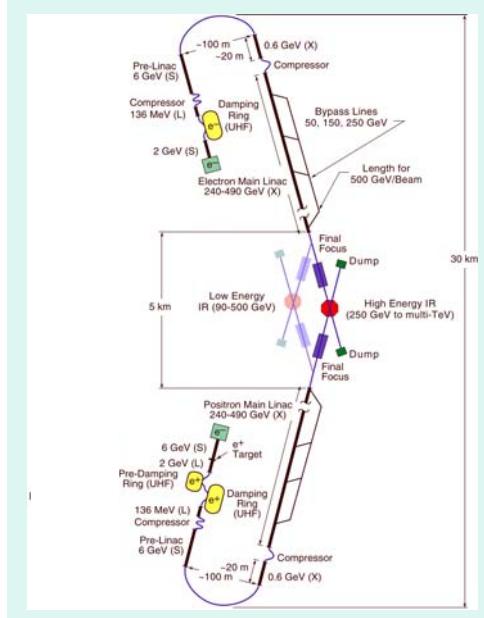


Fig. 3: Proposed layout of the 30 km long e- / e+ International Linear Collider (ILC)

The implication from these three examples is that the most demanding current facilities require beam stabilities of the order of a few microns in the horizontal plane and less than one micron in the vertical. In the future, these tolerances will become significantly tighter, with beam dimensions reducing to the nanometre level.

The implication of these beam requirements on magnet stabilities will now be considered.

3 Magnet stabilities

In this section, the effect of magnet positional and angular vibration (due to instabilities) and static displacement (reproducibility factors) on beam position will be considered. The discussion will mainly be concerned with the consequences of such movements of dipoles and quadrupoles, with sextupoles briefly considered.

3.1 Dipole magnets

In a circular accelerator, the dipoles generate the vertical field needed to deflect the beam and generate a quasi-circular complete 360° orbit; they are therefore the principal magnets present in the accelerator ring. In many accelerators, pure dipole fields are required to satisfy this bending role, the beam focusing being provided by separate quadrupole magnets (a ‘separated-function’ lattice). However, in a minority, a spatial transverse gradient is added to provide some focusing as well as bending, producing a ‘combined-function’ magnet. The issues dealt with in this section apply only to magnets generating a pure dipole field; where combined-function magnets are being used, the stability criteria relevant to quadrupoles, as described in the next section, will also apply and must be duly considered.

Whilst these issues are discussed in the context of a circular accelerator, they apply equally to beam-lines within which the charged particles are transported in a single pass.

3.1.1 Positional and angular displacements

- Small horizontal and vertical transverse displacements: the dipole design is intended to generate a uniform vertical field over the ‘good field’ aperture of the magnet in both the horizontal and vertical planes. Thus, with any small transverse displacement, the vertical field at the beam remains unchanged and hence has a negligible or zero effect on the beam; large displacement would result in some part of the beam path being outside the good field region of the magnet, but such movements are gross and very much larger than the instabilities that are considered.
- Longitudinal displacement: this also does not affect the integrated field strength of the magnet experienced by the beam and therefore the dipole still will generate the correct beam angular bend; however, it changes the position of the azimuthal centre of that bend in the lattice, creating a loss of symmetry in the complete circular ring; this will produce some closed-orbit distortion (i.e., horizontal displacement) of the beam around the complete ring; the degree of such distortion will completely depend on the lattice and needs to be examined for each particular design; it should also be noted that if the displacement is static (i.e., a positional reproducibility error), small distortions can be corrected by the beam-steering d.c. correctors, whilst longitudinal vibration (instability) could be expected to be sufficiently small to generate no appreciable distortion; notwithstanding, good communication between magnet engineers and lattice designers will be needed to resolve this issue and it is likely that a more stringent tolerance would be defined on longitudinal dipole placement than in the two transverse planes.
- Twist about the longitudinal axis (‘roll’ — see Fig. 4 for definition of angular errors); roll is a far more serious issue; a roll error of θ in a dipole will generate a small horizontal field:

$$B_x = B_y \sin \theta;$$

the horizontal field B_x will produce a vertical bend in the beam and, whilst θ will be very small, the main bending field B_y will be large and an unacceptable vertical steering effect may be produced, significantly distorting the orbit in the vertical plane; the principal issue will be static — the accuracy with which the magnet is surveyed into position — but dynamic effects (vibration) must also be carefully considered; an error that must be minimized, with a tolerance band to be specified by the lattice designers.

- Twist about the radial axis ('pitch'): this will produce an axial field component which will change horizontal and vertical focusing and will couple horizontal and vertical betatron oscillations; this is an undesirable effect in any accelerator but a critical issue in high-energy lepton machines where the vertical beam size is determined principally by this coupling effect; again, this error must be minimized to meet lattice specifications.
 - Twist about a vertical axis ('yaw'): possibly less critical than roll or pitch, as it does not couple motion in the two transverse planes; it does produce an entry and exit angle at each end of the magnet which will result in some transverse focusing effects which will, in principle, cancel each other out; a further topic for consultation with lattice experts.
-

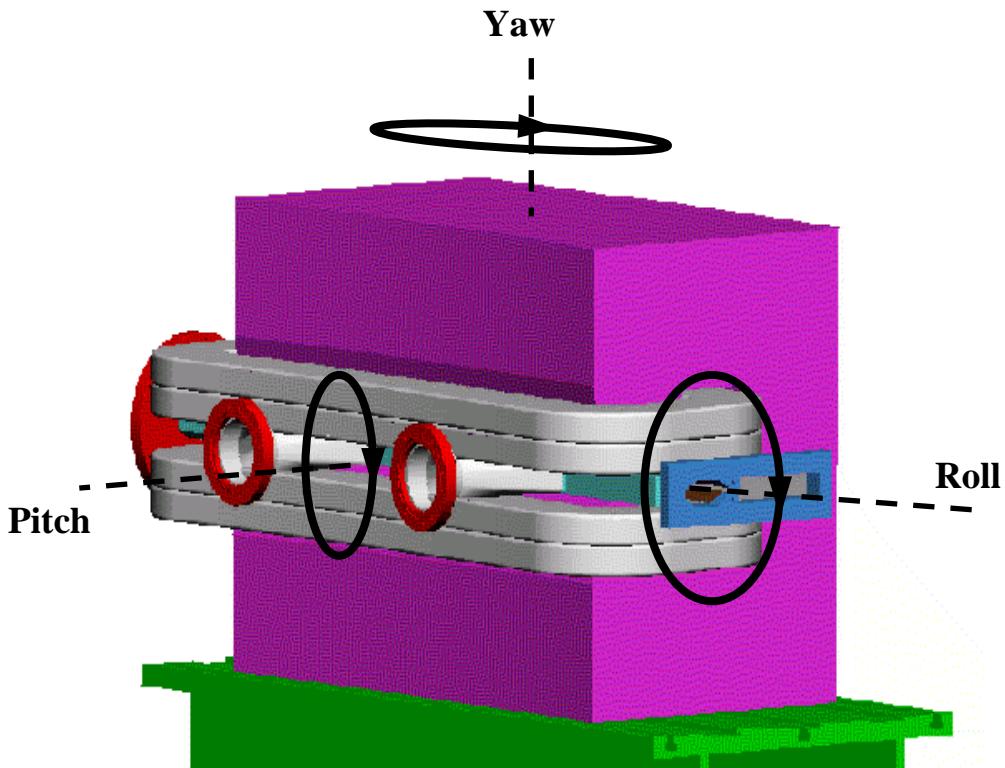


Fig. 4: Definition of terms used to describe angular errors in magnet alignment

3.1.2 Dipole strength variations

The beam deflection produced by a dipole magnet is proportional to its 'magnetic strengths' — the azimuthal integration of the vertical field through the length of the magnet:

$$\int B_y \cdot dz$$

Variations in strength are produced by:

- errors in reproducibility in magnet geometry, in both the inter-pole gap (or coil configuration for a s.c. magnet) and physical length, occurring during manufacture; these can be minimized by choosing the lattice position for the dipoles once their magnetic strengths have been measured after production;
- current leakage from the circuit powering the magnets' coils;
- where dipole string is separated into separate circuits, inequality of the output current from different power converters.

It is this parameter that is one of the most critical in a circular accelerator, for variations in bend-strengths within a lattice made up of (nominally identical) separate dipoles, will induce major orbit distortions. Again, the tolerance that can be placed on an acceptable variation in strength, dipole to dipole, depends critically on the lattice details but it is possible to give some general values for accelerators of different sizes. A tolerance of $\pm 1:10^4$ is typical for smaller machines, with greater reproducibility being required for larger installations; for example, the high-current dipole power converters for the LHC require a d.c. stability of better than 10 ppm [5].

3.2 Quadrupole magnets

3.2.1 *Quadrupole transverse positional variation*

Quadrupole magnets are required to generate zero field at the correct beam centre, with a linear field gradient across the beam, in both horizontal and vertical planes, to focus particles that are off-centre. The zero field point (the ‘magnetic-centre’ where $B_x = B_y = 0$) therefore needs to be accurately located on the correct, undeviated closed orbit. Any subsequent transverse displacement of the magnet will consequently result in the deflection of the central closed-orbit, moving the beam position in horizontal or vertical position (depending on the direction of the magnet displacement) at all other parts of the ring.

The amplitude of such beam displacement depends critically on the lattice design but, in a strong-focusing configuration, it will be very much greater than the magnet displacement that is causing the orbit distortion. The ratio of beam movement to quadrupole displacement is known as the ‘**amplification factor**’. For a simple FODO lattice (such as the Daresbury SRS) this factor is between 10 and 20; for a modern, complex lattice with low-beta insertion points, it is significantly larger. In the Diamond facility, for example, the amplification factors in the storage ring are

horizontal amplification factor:	60;
vertical amplification factor:	45.

Hence, transverse displacements in quadrupole magnets produce large orbit distortions and therefore are critical issues in determining beam stability. Continuing with the Diamond facility as an example and taking the figures for the required **beam** stability given in Table 2, the corresponding **quadrupole magnet** stabilities are as shown in Table 5:

Table 5: Quadrupole magnet transverse stabilities required in the 3 GeV synchrotron source, Diamond

	Horizontal	Vertical
Quadrupole transverse positional variation (μm) \leq	0.2	0.015

This clearly is a very demanding stability requirement and it is reasonable to question whether a positional stability of 0.015 μm in the vertical plane is achievable. This will be resolved in the later stages of this paper.

3.2.2 Variation in quadrupole strength

Static variation of strength from quadrupole to quadrupole, due to small errors in engineering reproducibility during construction and assembly of the magnets, will distort the beta values around the lattice and lead to different beam sizes at different circumferential positions around a circular machine. As with other instabilities, the amplitude of this distortion is dependent on the lattice and the acceptable level on the accelerator application. These effects must be estimated and the level of inter-quadrupole strength variation considered during magnet construction. Strength measurements on the complete quadrupole production sequence will indicate whether this target has been achieved, and adjustment of the quadrupole magnetic lengths by small modification at the magnet ends can be carried out if necessary.

Dynamic variations in strength of individual quadrupoles, caused by variation in magnet excitation of the separate power sources for those quadrupoles, are more serious and will lead to instability in beam dimensions at different parts of the ring. This is controlled by ensuring that the specification for the power converters provides for sufficient stability in output current. It should be noted that in some circumstances beta variations around a lattice are required (in a light source with different insertion devices in the various straights for example) and separate power sources are then essential, with the potential of varying their output currents individually.

Dynamic variations in strength of a complete quadrupole family will result in changes of the beam's 'tunes' — the number of betatron oscillations per revolution — in both the radial and vertical planes. This is highly undesirable, for it can cause the beam to engage a resonance resulting in beam transverse disturbances and loss. It must be minimized to an acceptable level, again by imposing a tight tolerance on the stability of the quadrupole power supplies and ensuring that the equipment meets this specification.

3.3 Summary of instability effects

The conclusion of the above discussion is that magnet positional and amplitude stability is important in all accelerators and critical in many. It should be borne in mind however, that:

- static positional displacements are corrected during installation and survey;
- static amplitude variations should be corrected by measurement at the end of manufacture;
- uniform positional displacement of ALL accelerator magnets by the same amount in the same direction is not a problem.

So independent (between magnet and magnet), dynamic (time varying) instabilities are of major concern, with quadrupole position being the most critical issue and presenting the biggest problems.

4 The causes of magnet-induced beam instabilities

Having examined

- the typical beam stabilities required in various types of facility,
- the relationship between the variations in magnet strengths and positions and the beam instabilities that they cause,
- the resulting magnet positional tolerances required in a modern state-of-the art facility,

we now need to explore further the root causes of the magnet instabilities, so that effective counter-measures can be considered.

An overview of this topic was presented at the 2009 Particle Accelerator Conference in Vancouver [6]. The speaker detailed the causes of positional variation as a function of disturbance frequency, as measured in the ALS at the Lawrence Berkeley National Laboratory, shown in Fig. 5.

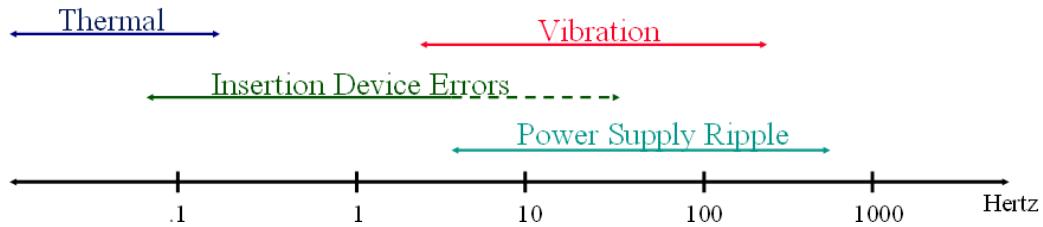


Fig. 5: Spectrum of the causes of magnet instability seen in the ALS (LBNL)

The paper also gave further details of the amplitudes of the disturbances and their dominant causes, at different frequencies, as shown in Table 6 below.

Table 6: The amplitudes and dominant causes of instabilities perceived in the ALS (LBNL)

Frequency	Magnitude	Dominant Cause
Two weeks (A typical experimental run)	$\pm 200 \mu\text{m}$ Horizontal $\pm 100 \mu\text{m}$ Vertical	1. Magnet hysteresis 2. Temperature fluctuations 3. Component heating between 1.5 GeV and 1.9 GeV
1 Day	$\pm 125 \mu\text{m}$ Horizontal $\pm 50 \mu\text{m}$ Vertical	Temperature fluctuations
8 Hour Fill	$\pm 50 \mu\text{m}$ Horizontal $\pm 20 \mu\text{m}$ Vertical	1. Temperature fluctuations 2. Feed forward errors
Minutes	1 to 5 μm	1. Feed forward errors 2. D/A converter digitization noise
.1 to 300 Hz	3 μm Horizontal 1 μm Vertical	1. Ground vibrations 2. Cooling water vibrations 3. Power supply ripple 4. Feed forward errors

Based upon these data, the following sources of magnet instability will be examined and, bearing in mind the discrepancy between the above figures and the required stabilities given in Table 5, possible counter-measures discussed:

- ground vibration;
- thermal instabilities;
- water vibration;
- power supply instabilities and ripple.

4.1 Ground vibration

Seismologists and surveyors express ground vibrations measurements as the ‘**power spectral density**’ (PSD), which has dimensions of $(\text{length})^2 (\text{frequency})^{-1}$; in practical circumstances, units of $(\mu\text{m})^2 (\text{Hz})^{-1}$ are often used. For a particular site, the PSD is measured as a function of frequency and the spectra plotted for the horizontal and vertical planes. The r.m.s value of the physical movement of the ground (z_{rms}) is then obtained by taking the square-root of the integrated spectrum (S_x) between defined frequency limits (f_1 and f_2):

$$z_{rms}(f_1, f_2) = \sqrt{\int_{f_1}^{f_2} S_x(f) df} .$$

4.1.1 The PSD spectrum

A typical power spectral density plot is shown in Fig. 6; this was obtained from bed-rock on the Daresbury Laboratory site [7] but has details that are common in all such measurements.

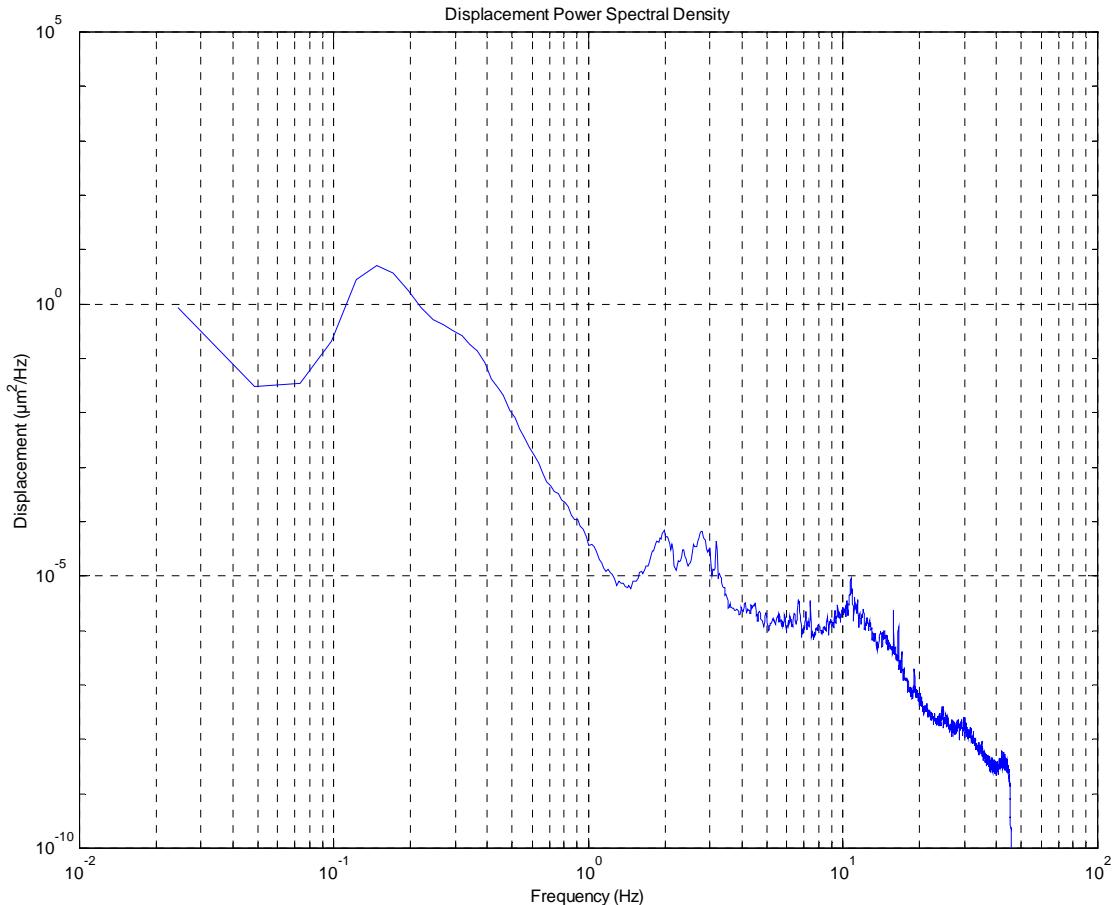


Fig. 6: Ground power spectral density in $(\mu\text{m})^2(\text{Hz})^{-1}$ as a function of frequency (Hz), as measured on bed-rock below the Daresbury Laboratory [7]; the axes are logarithmic; the ordinate varies between 10^{-10} and 10^5 ; the abscissa between 10^{-2} and 10^2 .

The peak between 0.1 and 0.3 Hz is called the ‘microseismic peak’ and is caused by the pounding of ocean waves on the coast-line; it is present on all sites, however far they may be from the sea!

The higher frequencies, at 1 Hz and above, are ‘technical and cultural noise’. The amplitude of these disturbances varies significantly between sites, depending on the locality. The proximity of heavy industrial plant, railway tracks or high speed roads will increase this noise, and a laboratory site will also generate its own ‘cultural’ noise. It is a point of some surprise to learn that an overflying, wide-body aeroplane will depress the ground beneath by up to 4 μm .

The ultra low-frequency ‘earth-tide’ is semi-diurnal ($c. 2 \times 10^{-5}$ Hz) and is large: $c.0.6$ m peak to peak, though, owing to its very long wavelength, this does not cause problems.

4.1.2 *Ground motion wavelength*

To disturb the beam, magnet vibrations need to be of different amplitudes over the circumference of an accelerator i.e., to be incoherent. The degree of incoherence due to ground motion depends completely on the wavelength of the disturbance. Seismologists have shown that ground disturbances can be categorized into bulk waves and surface waves; there are two types of each and they have different wave velocities and different wavelengths, details depending crucially on the underlying sub-soil and bed-rock formation.

However, Holder [7], when studying the effects of ground vibration on the stability of the lattice magnets in the Diamond storage ring with a 150 m diameter, concludes that ‘ground waves with wavelengths of significantly greater than 300 m will not be a problem’ and consequently ‘the low frequency limit, below which the lattice will move coherently, is about 1.5 Hz’. Clearly, the value of this lower limit depends on the size of the accelerator, being higher for small accelerators, but substantially lower for large machines. He then proceeds to warn that ‘above this limit particularly important frequencies exist that give ground wavelengths that are the same order as the betatron wavelengths and therefore cause resonant beam excitations’.

The message is therefore clear — for all but the largest accelerators, the large microseismic peak does not constitute a significant problem, but great attention must be paid to the effects generated by the technical and cultural noise at frequencies of the order of a few Hz and higher. It should be noted from Fig. 6 that there is a plateau in the PSD spectrum between approximately 1 Hz and 10 Hz. Above that the amplitude is decreasing rapidly — approximately three orders of magnitude between 10 Hz and 50 Hz. Whilst the magnitude of the technical and cultural noise will vary from site to site, this rapid decrease is a standard feature. Hence, excitations in the region of 1-10 Hz have the potential to be very damaging.

4.1.3 *Girder resonance*

The lattice magnets will be mounted on girders to provide as rigid a support as possible and to facilitate positional adjustments. An engineering FEA model of a Diamond girder, supporting a dipole, four quadrupoles, and three sextupoles on the single girder, is shown in Fig. 7. Ground vibration will be transmitted to the magnets through the girder mounts, so these mechanical engineering components feature strongly in the understanding of the effect of ground movement on the accelerator.

The girders will have a number of resonant frequencies, corresponding to the normal modes of oscillation that can exist in their geometric structure. Steier [6] explains that in the design of some earlier accelerators, massive support girders were used, resulting in low resonant frequencies that were in the ‘danger zone’ between 1 Hz and 10 Hz. Later, lighter girders were designed but these were still weighted down by mechanical components intended to provide the precision adjustment of magnet position, with the consequence of the resonant frequencies still being low. However, he adds that the latest synchrotron radiation storage rings (Diamond, Soleil and NSLS II) avoid this problem and have successfully ‘lifted’ the girder resonance frequencies into the tens of hertz.

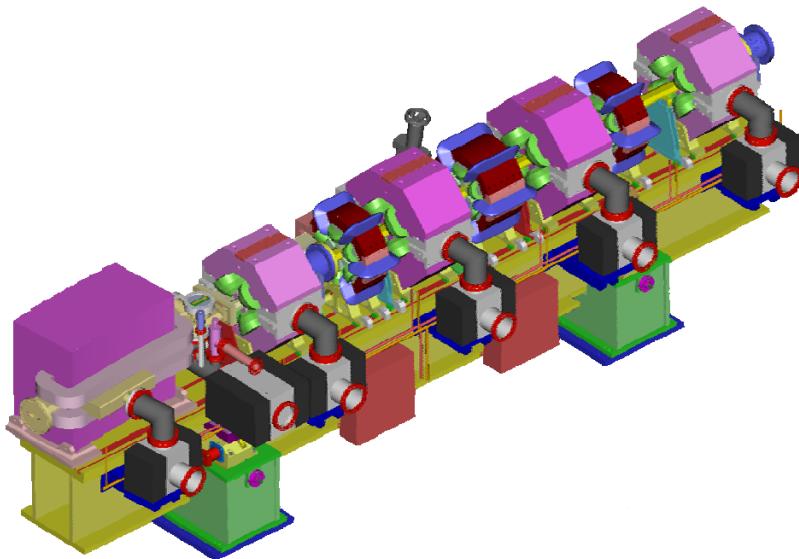


Fig. 7: Engineering model of a Diamond girder supporting a dipole, four quadrupoles, and three sextupoles

Huang [8] presents data on the design of the Diamond lattice girders, firstly studying their static deflections. He indicates that the 561 m storage ring uses 72 magnet support girders level between 2 planes, 1 mm apart; the average height difference between adjacent girders is approximately 0.1 mm with a predicted annual variation in level of approximately 0.4 mm. Using finite element analysis (FEA) techniques he predicts the vertical static deflection of the magnets on the girders shown in Fig. 8; the calculated maximum static deflection is 48 μm .

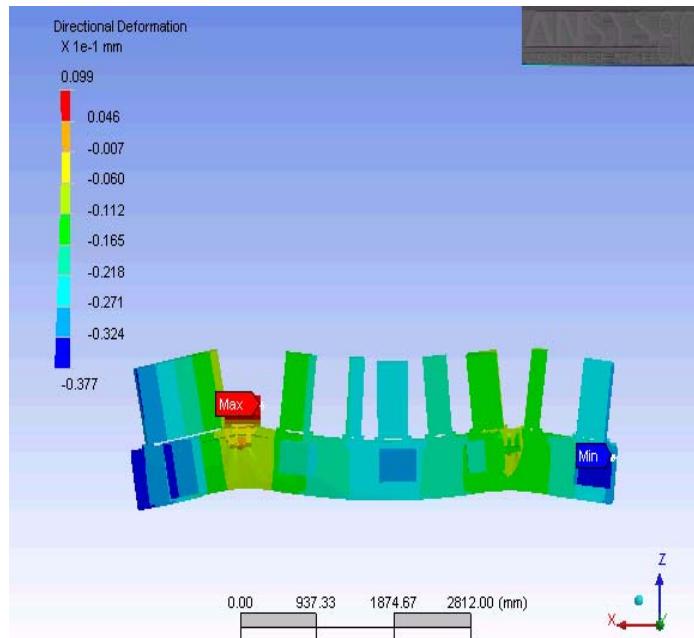


Fig. 8: Predicted static vertical deflections of the Diamond lattice girders when loaded with magnets, according to Huang [8]

Clearly, these deflections will need to be taken into account when the magnets are being surveyed into position.

The vertical and horizontal resonance spectra of the girders were also studied using the FEA codes and, knowing the ground power density spectrum on the Diamond site, the consequential spectrum and amplitude of vibration for eight different magnets, positioned on the girders, predicted. The resultant power spectral densities (in units of mm^2/Hz) in the vertical and horizontal planes are shown in Figs. 9 and 10.

The predicted vertical resonances are at 41, 51, 53, 63, 73, and 88 Hz, with the horizontal resonance spectrum being similar. Clearly, the resonance frequencies have been raised well above the 10 Hz target minimum.

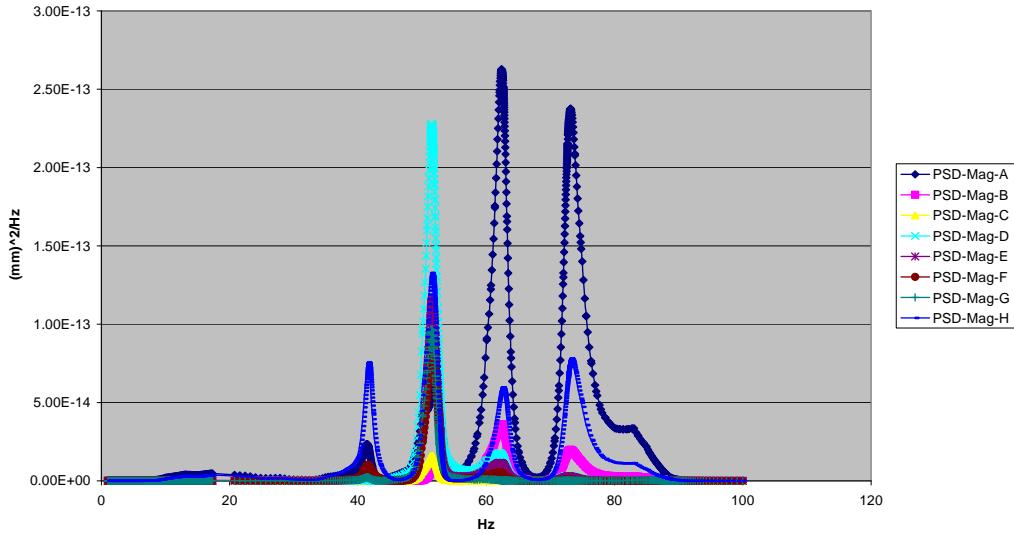


Fig. 9: Predicted spectrum of vertical vibrations at the positions of eight magnets in the Diamond lattice, according to Huang [8]; the amplitudes are expressed as the ‘power spectral density’ in units of $(\text{mm})^2/\text{Hz}$.

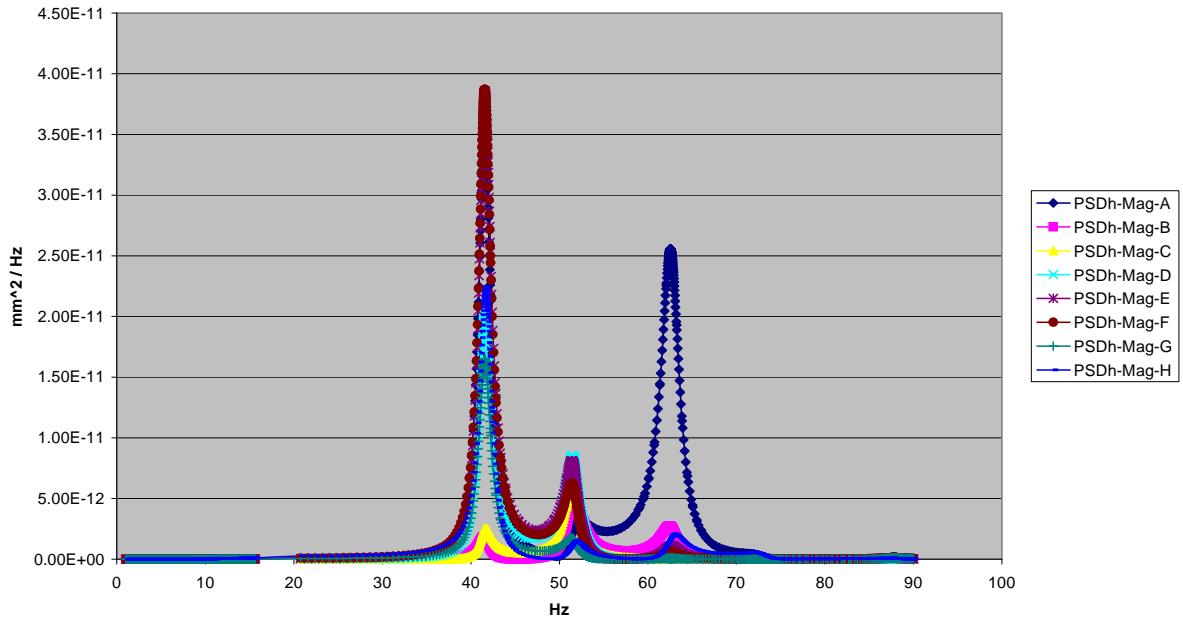


Fig. 10: Predicted spectrum of horizontal vibrations at the positions of eight magnets in the Diamond lattice, according to Huang [8]; the amplitudes are expressed as the ‘power spectral density’ in units of $(\text{mm})^2/\text{Hz}$.

4.2 Thermal instabilities

With the linear coefficient of expansion of steel being $c.1.3 \times 10^{-5} /K$, the effects of temperature change on magnet strength and position are illustrated by the following example:

A quadrupole with:

- inscribed radius of 40 mm;
- pole length of 100 mm;
- magnetic centre 200 mm above support feet;

then:

- strength change due to pole expansion = 0.007% / K;
- movement of magnetic centre = 2.5 $\mu\text{m}/\text{K}$.

So during power-up, with a potential temperature increase of the magnet yoke of up to 10 K, the quadrupole centre would move by up to 25 μm — far in excess of the tolerances required as given in Table 5 — and the strength increase by nearly 0.1% — also unacceptably large.

It is clearly crucial to control the temperature of the accelerator environment — both the air temperature in the tunnel and the water circulated for equipment cooling. Also, the temperature rise in the magnets themselves must be minimized.

4.2.1 Ambient temperature control

The degree of temperature control now economically possible in the large enclosed space of an accelerator lattice tunnel and the circulated cooling water is indicated by Steier [6] by the data presented in Fig. 11, which shows the temperature of the cooling water (left-hand graphs) and tunnel air temperature (right-hand graphs) at the ALS over 24 hours.

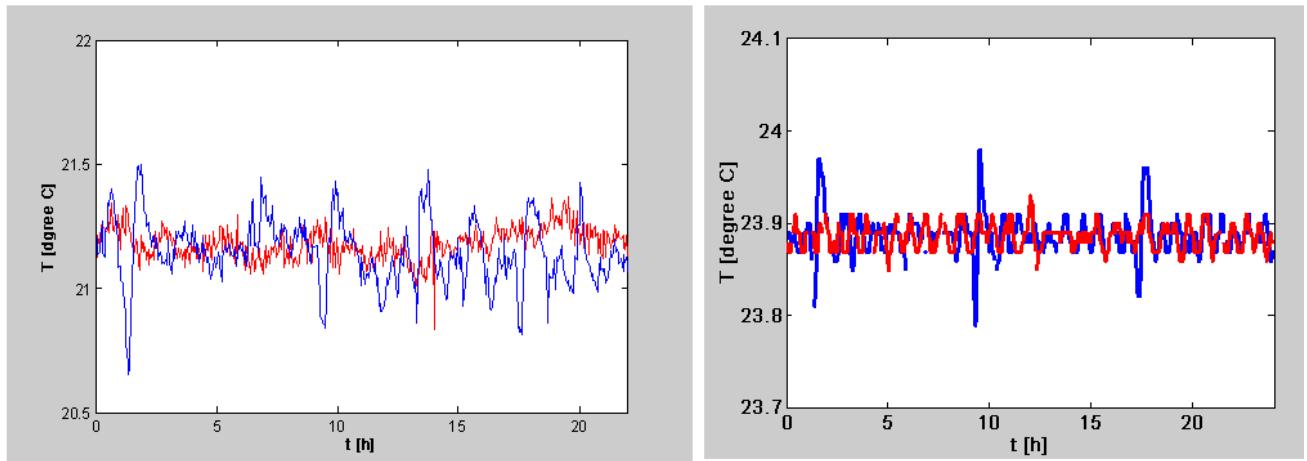


Fig. 11: Temperature of the ring cooling water (left-hand) and tunnel air temperature (right-hand) at the ALS over 24 hours; the red curves are with the tunnel room removed.

The mean and maximum temperature variations in the Diamond storage rings are provided by Kay [9]:

- tunnel air temperature: $22.0 \pm 0.1^\circ\text{C}$;
- demineralized water supply temperature: $22 \pm 0.3^\circ\text{C}$.

4.2.2 Magnet temperature rise

However stable the ambient air temperature and the cooling water input temperature, the magnet yokes will run above that value. It is therefore important to limit the temperature rise in the magnet coils, as heat conducted into the yoke will result in dimensional changes to the yoke with corresponding movement to the magnetic centre. The power dissipation in the coils is governed by the root mean square (r.m.s.) of the coil current and the copper cross-section of the total coil.

The r.m.s. current depends on the waveform of the coil current; in the case of a storage ring, where, during operation, the magnets are kept at a fixed dc current, the r.m.s. current is obviously equal to the d.c. value. It is also standard knowledge that for a pure alternating waveform, the r.m.s. current is the peak divided by the square root of 2. But in a booster synchrotron, where the magnets are excited by a monodirectional waveform that comprises a d.c. and an a.c. component, the r.m.s. value is given by the following:

$$I_{\text{rms}} = \sqrt{I_{\text{dc}}^2 + (I_{\text{ac}}^2)/2} .$$

In the particular case of a fully biased sin-wave,

$$I_{\text{dc}} = I_{\text{ac}} = (1/2) I_{\text{peak}}$$

and

$$I_{\text{rms}} = I_{\text{dc}} (\sqrt{3/2}) = I_{\text{peak}} (1/2) \{\sqrt{3/2}\} ,$$

where I_{ac} is the peak of the a.c. component and I_{peak} is the current maximum in the waveform.

The choice of conductor cross-section in the coils is determined by economic criteria. The ampere-turns needed are fixed by the magnet geometry and required field value. There is a high value of current density at which cooling becomes difficult, but it is usual to operate well below such a critical value. Examining the economics of building and then operating the magnet systems gives the following conflicting requirements:

- to minimize the capital cost of the coils and yoke, and hence of the whole magnet, as small an amount of conductor as possible should be chosen;
- but this then gives a higher current density, increased power losses, and expensive power expenditure, so lifetime operational costs increase.

Hence, it is usual to consider the total lifetime cost of the magnet system, both construction and operational costs, and to determine the optimum value of current density that minimizes total expenditure. The situation is demonstrated in Fig. 12.

In many cases, accelerator project managers choose to design the magnets with current densities below the optimum value, increasing the capital values of the magnet but reducing the temperature rise in the magnets (and decreasing the power loss to make the project more environmentally friendly). This clearly is the route to improve the stability of the magnet.

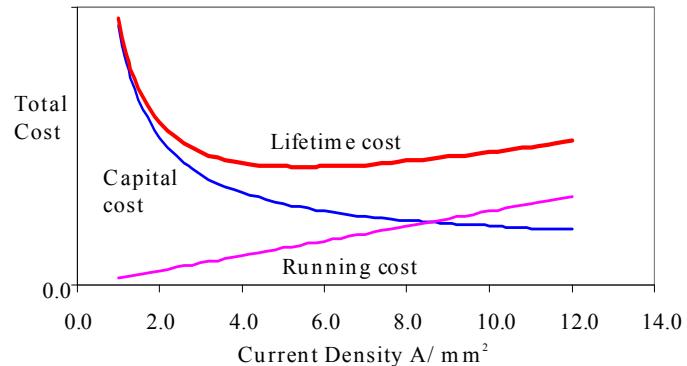


Fig. 12: Optimization of current density in the magnet coils, to limit lifetime cost made up of capital and operational components.

As an example, the Diamond quadrupoles were designed with a current density, at maximum gradient, of $\sim 2.5 \text{ A/mm}^2$. This is above the economic optimum value, which is generally regarded as being in the region of 3.5 to 4 A/mm^2 . This gave a maximum temperature rise of 10°C — and perhaps that is a little too high!

4.3 Water system induced vibration

This source of instability can really be classified as part of the technical and cultural noise mentioned in Section 4.1.1. However, it is different from other such sources of instability in that it is generated by a system that is part of the accelerator complex and which is, therefore, under the control of those engaged in the accelerator design or operation.

There are two sources of vibration associated with the cooling water that is needed to be circulated through the coils' hollow conductors:

- the mechanical vibration generated by the water pumps and transmitted to the magnets through the water channels; the pumps have rotating systems and, however well balanced these may be, there is always some residual vibration;
- the water passing through the hollow conductor is required to remove heat from the conductor inner tube surface and to do this it must have a velocity which produces turbulent flow; laminar flow does not break the water's boundary layer at the tube walls and therefore does not efficiently remove the dissipated heat; the turbulence generates mechanical vibration.

Figure 13 shows the horizontal spectra of girder vibration with water on and off on a beam-line at the Diamond Light Source [10].

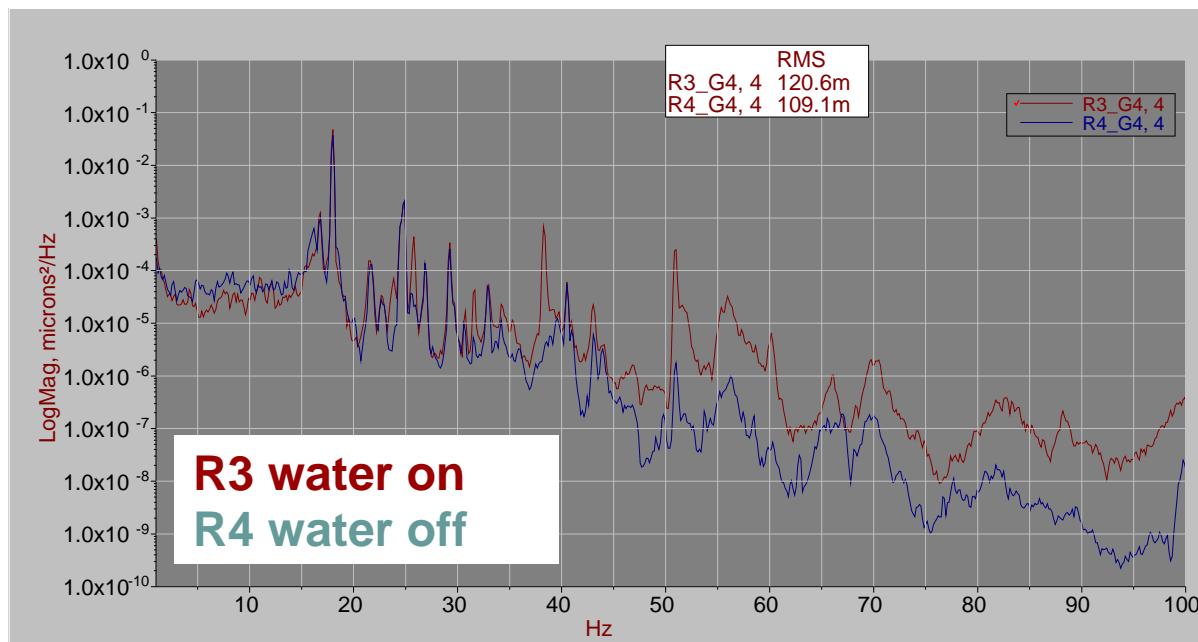


Fig 13: Spectra of horizontal girder vibration on a beam-line at the Diamond Light Source [10]

It can be seen that, at low frequencies, there is little difference but above approximately 50 Hz there is greater than an order of magnitude increase in some parts of the spectrum when the water is circulating. But it must be appreciated, however, that these relatively large differences are in regions where the amplitudes are two to three orders of magnitude down on the spectra at low frequencies. They therefore make little contribution to the overall r.m.s. (integrated) values, which are:

$$R3: 120 \times 10^{-9} \text{ m};$$

$$R4: 109 \times 10^{-9} \text{ m}.$$

The additional integrated r.m.s. vibration induced by the cooling system is therefore of the order of 10% of the total; not large, but significant and certainly worth minimizing. This therefore also points to minimizing the thermal losses in the coil by designing coils with low current densities that require lower volumes of cooling water and smaller pumps.

4.4 Power supply instabilities and ripple

Irrespective of whether the magnets require steady, direct current excitation or an alternating, biased, current waveform, they will be connected to ‘power converters’ that are fed from the alternating public supply, at an appropriate high voltage, and deliver the necessary power to the magnets, at the right impedance level, with the correct waveform, stabilized, smoothed and controlled. Such operation requires the use of power switching devices and fast feedback servosystems that have a means of controlling the high current and voltage output.

Before the advent of switched power electronics, control was through the use of switching valves (mercury arc rectifiers and thyratrons), which could withstand the high currents and moderate to high voltages. They were, however, relatively slow, commutation usually being possible at a maximum frequency of *c.* 300 Hz. Later, the first solid-state power controlled switches, thyristors, offered similar performance but at reduced cost and with simpler auxiliary apparatus.

Such systems were used to power accelerator magnets up to the 1960s and later. They all generated sharp switching spikes in their output voltage, at their fundamental switching frequency (usually 300 Hz) and higher harmonics. Large smoothing filter circuits were necessary and operational performances with a stability at best of 1:10⁴, with similar levels of current ripple were achieved.

The situation is now substantially improved by the invention and commercial availability of far faster power switches, which can be commanded to ‘switch off’ as well as to become conducting; a facility not available in the older power switching devices. These new devices — insulated gate bipolar transistors (IGBTs) — can now switch currents well into the thousands of amps at kilovolt potentials, in a few microseconds, though it is necessary to ensure that excessive power dissipation does not occur at the junctions during the switching transients.

These devices are now used in modern ‘switch-mode’ power converters. A simplified schematic diagram of such a device, to generate a d.c. output, is shown in Fig. 14.

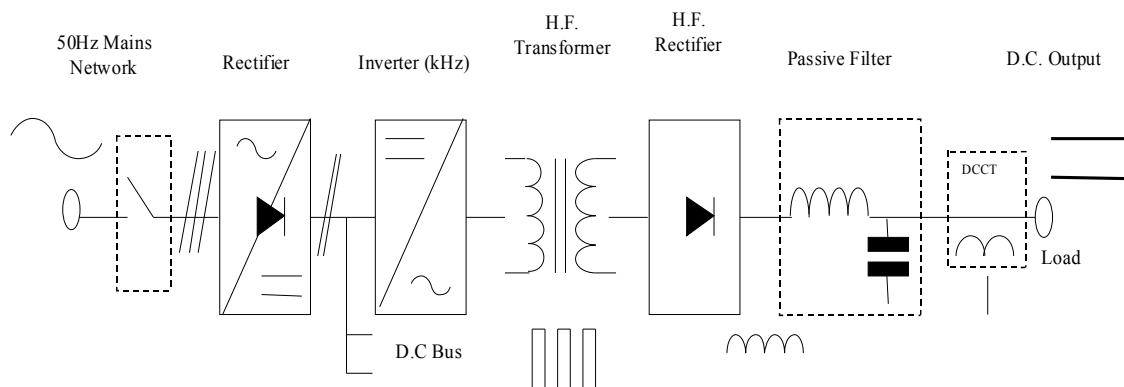


Fig 14: Simplified schematic diagram of a typical modern ‘switch-mode’ power converter

The front-end is powered from the unregulated public supply network. After the usual isolating switch-gear, the power is rectified through a simple unregulated diode system. The output is then fed

to a high-frequency inverter, using IGBTs to generate an alternating voltage at frequencies up to 40 kHz (depending on the application and power levels). This is then transformed to meet the output impedance level that is needed for the load. As the size and cost of a power transformer varies roughly as the inverse of its operating frequency, this transformer could be an order of magnitude smaller than one handling the same power at the network supply frequency. The alternating power is again rectified — providing a voltage on the output rail which has a very much higher frequency ripple and which therefore can be smoothed with much smaller filters. Finally, the output current is passed through a direct-current current transformer (DCCT) which provides the signal for an ultra-fast feed-back loop to control the output to a very high stability. This is done through the inverter, which can control its voltage output with a response frequency that is commensurate with its switching time (a few tens of microseconds).

Such systems can provide current stabilities of $1:10^5$, nearly as standard, and even of the order of a few ppm at somewhat higher cost; their output ripple is very low (of the same order as their stability) and, in the event of an external fault being detected in the load circuit, they can cut the output power far faster than a circuit-breaker or a fuse; power for power, they are no more expensive than the old power converter systems.

It is clear that they should be the power converters of choice for a modern accelerator magnet system.

5. Achieved stability: the current ‘state of the art’

This final section presents the level of stability now being achieved, using the Diamond Light Source as an example. It also points the way to possible further improvement using dynamic beam position control, a technique outside the remit of this paper.

5.1 Minimizing the instabilities

Section 4 above gives an overview of the most prevalent sources of instability and gives an indication of the most productive ways of minimizing their effects. These can be briefly summarized as applying the following ‘due diligence’ provisions during project planning and engineering design:

- choose a site with low ground vibration, examining and measuring the ‘technical and cultural’ noise that is present and assessing the impact of the measured spectra on engineering components and, consequentially, the beam;
- design the magnet girders to have high resonant frequencies of *c.*50 Hz or above, obtaining girder response spectra which have such values, well beyond the microseismic peak and in the region where the technical and cultural noise spectrum is also decreasing strongly with frequency;
- design for highly stable temperature control in the accelerator tunnel;
- minimize the waste heat load from auxiliary equipment that is conducted into the accelerator tunnel;
- minimize magnet temperature rise by using low current densities (particularly quadrupoles), adopting a value well below that indicated as being the economic optimum between capital and running cost;
- use as low water velocity as possible whilst still maintaining turbulent flow in the magnet’s cooling water channels, in order to scourge the heat from the conductor inner wall;
- choose water pumps that have minimum mechanical vibration both through their mounts and also transmitted to the water;

- mechanically insulate magnets from water pumps, feed-lines, etc.;
- use best quality, low ripple, highly stable, state-of-the-art power supplies.

5.2 Actual performance

Given all the provisions outlined above, what has been achieved in the state-of-the-art synchrotron source Diamond? Figure 15 shows the horizontal and vertical electron beam displacement power spectral density as measured at Diamond [11]. These data were obtained with the feedback orbit control systems (see below) disabled.

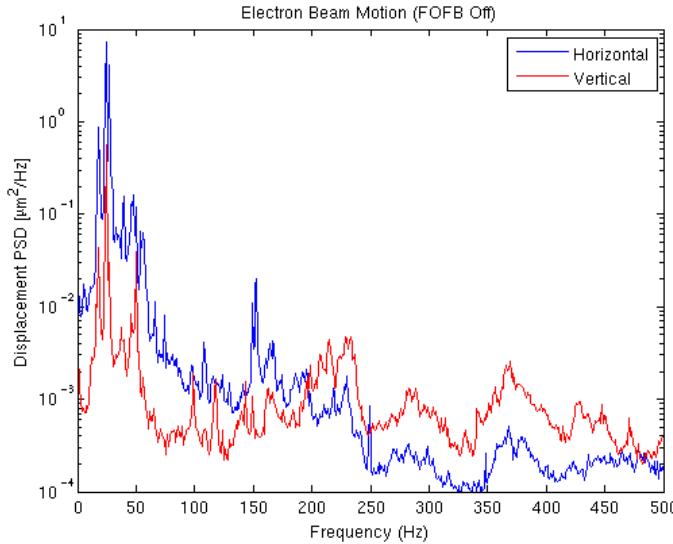


Fig 15: The horizontal and vertical electron beam displacement power spectral density measured at the Diamond Light Source with orbit feedback systems disabled

Integrating these data gives the total r.m.s. displacement spectra, as shown in Fig. 16.

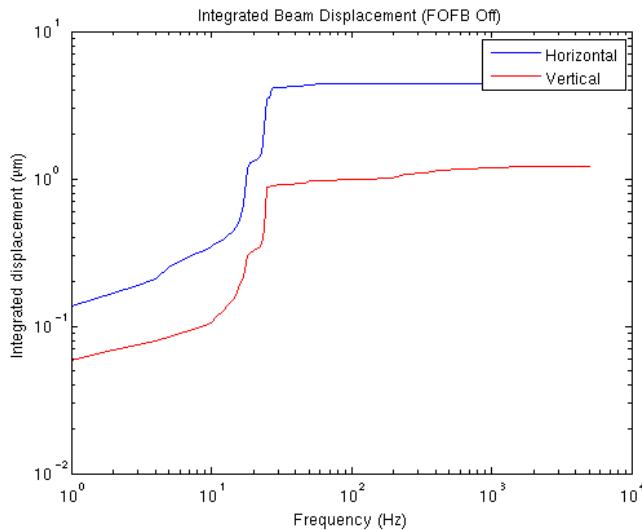


Fig 16: The integrated horizontal and vertical electron beam r.m.s displacement, measured at the Diamond Light Source with orbit feedback systems disabled

Note amplitude (r.m.s.) of integrated vibrations:

- | | |
|------------|---------------------------------------------------------------|
| horizontal | 4 μm ; with a target value of 12.3 μm ; |
| vertical | 1 μm ; with a target value of 0.64 μm . |

So the horizontal vibration is well within specification, but the vertical disturbances need to be further reduced by a factor of about 0.5. This indicates that the reduction of vibration achieved by the best available mechanical and electrical engineer provisions is not fully adequate to meet the stringent requirements of beam stability in the accelerator. However, the use of beam-position feedback systems can then give a further reduction in beam motion. This is demonstrated in the figures below, which show the integrated horizontal and vertical positional motion (Fig. 17) and the horizontal and vertical angular motion (Fig. 18) of the electron beam in 24 straights of the Diamond Light Source, with and without ‘fast-orbit feedback’ (FOFB).

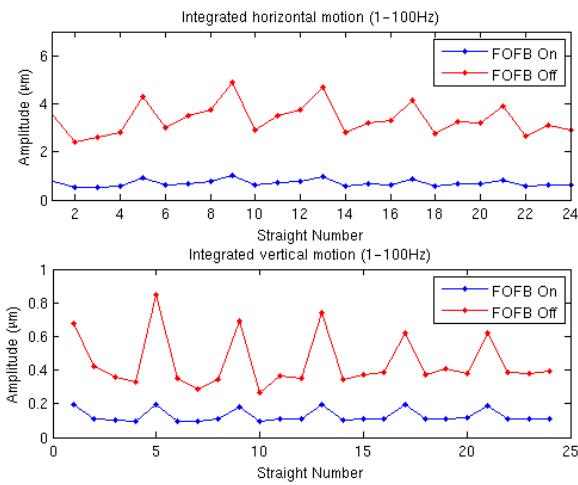


Fig 17: The horizontal and vertical positional motion of the electron beam in 24 straights of the Diamond Light Source, with and without ‘fast-orbit feedback’ (FOFB)

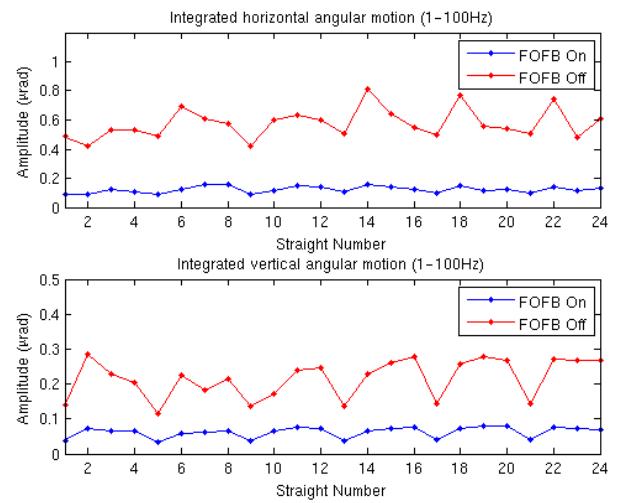


Fig 18: The horizontal and vertical angular motion of the electron beam in 24 straights of the Diamond Light Source, with and without ‘fast-orbit feedback’ (FOFB)

The four sets of data demonstrate clearly that the fast-orbit feedback systems produce substantial improvement in beam stability and thereby meet the original specification in the vertical and horizontal direction. However, it is not the purpose of this paper to explore FOFB systems any further.

6 Conclusion

By using best engineering practice and modern techniques for the design, construction, and operation of magnets and their power supplies, beam disturbance due to magnet instability and poor reproducibility can be minimized. But beam-position feedback systems will generally be needed as the final stabilizing influence on the beam and these are now extensively used in light sources, colliders, and other such facilities.

Acknowledgements

In preparing the presentation and this paper, I made use of much published material — including very new data presented at the Vancouver PAC (June 09) — and also information directly provided by colleagues. I am therefore happy to acknowledge help, either directly or through study of their papers, from Roberto Bartolini, DLS; Nicolas Delerue, University of Oxford; Richard Fielder, DLS; David Holder, Cockcroft Institute and University of Liverpool; Hou-Cheng Huang, DLS; James Kay, DLS; Ian Martin, DLS; Christopher Steier, LBNL; Jörg Wenninger, CERN.

References

- [1] R. Bartolini, Diamond Light Source status and future challenges, <http://www.isis.rl.ac.uk/accelerator/lectures/DLRALJAW/2009/Bartolini.ppt#309,1,Diamond Light Source Status and Future Challanges>.
- [2] R. Bartolini, H. Huang, J. Kay and I. Martin, Analysis of beam orbit stability and ground vibrations at the Diamond Storage Ring, WEPC002, Proc. EPAC08, Genoa, pp. 1980–2.
- [3] J. Wenninger, The LHC Accelerator Complex; <https://jwenning.web.cern.ch/jwenning/documents/Lectures/HCSS-June07-partA.ppt#256,1,The LHC Accelerator>.
- [4] N. Delerue, John Adams Institute, University of Oxford, A laser-wire to measure the beam sizes at the ILC, Proc. IoP HEPP Conference, Dublin, 21–23 March 2005.
- [5] F. Bordry and A. Dupquier, High current, low voltage power converters for LHC, Proc. EPAC'96, Sitges, 3 pp.
- [6] C. Steier, State of beam stability and control in light sources, Proc. PAC09, Vancouver.
- [7] D. Holder, Basics of site vibration measurement as applied to accelerator design, CLRC Daresbury Lab, AP-BU-rpt-001, Jan. 2000.
- [8] H-C Huang, DLS, Girder deflection and vibration analysis, MENG-FEA-REP-011.
- [9] J. Kay, DLS, private communication.
- [10] H-C Huang, Proc. Vibration Workshop, 1 April 2008, DLS.
- [11] I. Martin, DLS, Fast orbit feedback performance measurements, AP-SR-REP-0160.

Participants

AKCOLTEKIN, E.	ACCEL Instruments GmbH, Bergisch Gladbach, DE
ANTOINE, S.	SIGMAPHI, Vannes, FR
BALHAN, B.	CERN, Geneva, CH
BARRIERE, J.-C.	CEA Saclay, Gif-sur-Yvette, FR
BAUCHE, J.	CERN, Geneva, CH
BEECKMAN, W.	SIGMAPHI, Vannes, FR
BENNA, M.	ACCEL Instruments GmbH, Bergisch Gladbach, DE
BIRCH, S.	ISIS, Didcot, UK
BOCIAN, D.	Fermilab, Batavia, US
BODKER, F.	Siemens, Jyllinge, DK
BUCK, B.	MIT, Hampstead, US
BUDZ, P.	ACCEL Instruments GmbH, Bergisch Gladbach, DE
CAPELLUTO, A.	ASG Superconductors, Genoa, IT
CASTRO-GARCIA, P.	DESY, Hamburg, DE
CASTRONOVO, D.	Sincrotrone Trieste SCpA, Trieste, IT
CHAU, L.P.	Institut fuer Angewandte Physik, Frankfurt-am-Main, DE
DEFERNE, G.	CERN, Geneva, CH
DUNKEL, O.	CERN, Geneva, CH
ELLWOOD, G.	STFC Rutherford Appleton Laboratory, Didcot, UK
ESHRAQI, M.	CERN, Geneva, CH
FELDMEIER, E.	HIT Betriebs GmbH, Heidelberg, DE
FISCARELLI, L.	CERN, Geneva, CH
FOREST , F.	SIGMAPHI, Vannes, FR
FORTON, E.	IBA, Louvain-la-Neuve, BE
GABARD, A.	PSI, Villigen, CH
GARCIA PEREZ, J.	CERN, Geneva, CH
GLOVER, M.	STFC Rutherford Appleton Laboratory, Didcot, UK
GOELLER, S.	Siemens AG, Erlangen, DE
GOLLUCCIO, G.	CERN, Geneva, CH
GROTH LADEFOGED, P.	Danfysik A/S, Roskilde, DK
HUANG, J.-C.	NSRRC, Hsinchu City, TW
HUGHES, M.	STFC Rutherford Appleton Laboratory, Didcot, UK
HUGON, C.	CEA Saclay, Gif-sur-Yvette, FR
IUNGO, F.	LNF-INFN, Frascati, IT
JAGO, S.	STFC Rutherford Appleton Laboratory, Didcot, UK
JOHANSSON, M.	Scanditronix Magnet AB, Vislanda, SE
KLEEVEN, W.	IBA, Louvain-la-Neuve, BE
KRAMER, T.	EBG MedAustron, Wiener Neustadt, AT
KUBO, F.	ACCEL Instruments GmbH, Bergisch Gladbach, DE
KUMAR, S.	Inter University Accelerator Center, New Delhi, IN
KUO, C.Y.	NSRRC, Hsinchu City, TW
LAWRIE, S.	STFC Rutherford Appleton Laboratory, Didcot, UK
LEIBROCK, H.	GSI, Darmstadt, DE
LERAY, M.J.	SIGMAPHI, Vannes, FR
LONGHI, E.	Diamond Light Source, Didcot, UK
LOVERIDGE, P.	STFC Rutherford Appleton Laboratory, Didcot, UK

MARINOV, K.	STFC Daresbury Laboratory, Warrington, UK
MARTINO, M.	CERN, Geneva, CH
MASSANA, V.	CELLS-ALBA, Bellaterra, ES
MENOTTI, E.	University of Trieste, Basovizza, IT
MODENA, M.	CERN, Geneva, CH
MOLLEN, A.	CERN, Geneva, CH
MUELLER, H.	GSI, Darmstadt, DE
MUSARDO, M.	Sincrotrone Trieste SCpA, Trieste, IT
NANDI, C.	Variable Energy Cyclotron Centre, Kolkata, IN
PARADIS, Y.	IBA, Louvain-la-Neuve, BE
RAMOS, D.	CERN, Geneva, CH
RIAL, E.	Diamond Light Source, Didcot, UK
RIGLA PEREZ, J.	Instituto de Tecnicas Energeticas - UPC, Barcelona, ES
ROTTENBACH, A.	Siemens AG, Erlangen, DE
RUIZ DE VILLA, E.	Elytt Energy SL, Madrid, ES
SARDONE, M.	LNF-INFN, Frascati, IT
SIDOROV, S.	PSI, Villigen, CH
SPATARO, C.	Brookhaven National Laboratory, Upton, US
STOCKNER, M.	EBG MedAustron GmbH, Wiener Neustadt, AT
STODEL, M.	GANIL/CNRS, Caen, FR
STRUIK, M.	CERN, Geneva, CH
SUGITA, K.	GSI, Darmstadt, DE
TASSISTO, M.	ASG Superconductors, Genoa, IT
TOMLINSON, S.	STFC Rutherford Appleton Laboratory, Didcot, UK
WEST, S.	STFC Rutherford Appleton Laboratory, Didcot, UK
WETERINGS, W.	CERN, Geneva, CH
WIESNER, C.	Institut fur Angewandte Physik, Frankfurt-am-Main, DE
WITTE, H.	John Adams Institute, Oxford, UK
ZIMMERMANN, H.	Siemens AG Healthcare Sector, Erlangen, DE