CERN ACCELERATOR SCHOOL

MEASUREMENT AND ALIGNMENT OF ACCELERATOR AND DETECTOR MAGNETS

Europa Palace Hotel, Anacapri, Italy
11-17 April 1997

PROCEEDINGS
Editor: S. Turner
ABSTRACT

These proceedings present the lectures given at the eleventh specialised course organised by the CERN Accelerator School (CAS), the topic this time being 'Measurement and Alignment of Accelerator and Detector Magnets'. A similar course was already presented at Montreux, Switzerland in 1992 and its proceedings published as CERN 92-05. However recent progress in the field, especially in the use of superconducting magnets, has been so rapid that a revised course had become imperative. The lectures start with basic magnet theory and the motivation for magnet measurements followed by a review of superconducting magnets and their field dynamics. After a review of measurement methods, the details of search and harmonic coils, magnetic resonance techniques and Hall generators are given followed by methods to minimise errors in mechanical equipment, series production and detector magnet measurements. Turning to magnet metrology and alignment, first data quality control is explained followed by the setting of reference targets, and the alignment methods for accelerators and experiments including alignment by feedback. Finally seminars are presented on the biological effects of magnetic fields and on superconducting magnet fabrication and quality control.
CAS
CERN Accelerator School

MINNETS

11-17 April 1997
Europa Palace Hotel, Aqaeapri, Italy

This course is intended for staff in laboratories, universities,
and companies manufacturing associated equipment.

Further information and application forms from:
Mrs. S. von Wartburg,
CERN Accelerator School,
AC Division,
CH-1211 Geneva 23

E-MAIL:
Suzanne.von.Wartburg@cern.ch

FAX:
++41 22 767 5460

CLOSING DATE FOR APPLICATIONS: 1 FEBRUARY 1997
# MEASUREMENT AND ALIGNMENT OF ACCELERATOR AND DETECTOR MAGNETS

Hotel Europa Palace, Anacapri, 11–17 April 1997

<table>
<thead>
<tr>
<th>Time</th>
<th>Friday 11 April</th>
<th>Saturday 12 April</th>
<th>Sunday 13 April</th>
<th>Monday 14 April</th>
<th>Tuesday 15 April</th>
<th>Wednesday 16 April</th>
<th>Thursday 17 April</th>
</tr>
</thead>
<tbody>
<tr>
<td>08.30</td>
<td>Basic theory of magnets</td>
<td>Metrology and quality control</td>
<td>Breakfast</td>
<td>Setting reference targets</td>
<td>Search coils II</td>
<td>Mechanical equipment</td>
<td>Finding the axis</td>
</tr>
<tr>
<td>09.30</td>
<td>A. Jain</td>
<td>J. Iliffe</td>
<td>Breakfast</td>
<td>R. Ruland</td>
<td>M. Green</td>
<td>G. Moritz</td>
<td>P. Sievers</td>
</tr>
<tr>
<td>09.40</td>
<td>Motivation for precision</td>
<td>Time varying fields</td>
<td>Excursion</td>
<td>Search coils I</td>
<td>Harmonic coils I</td>
<td>Harmonic coils II</td>
<td>Alignment by feedback</td>
</tr>
<tr>
<td>10.40</td>
<td>E. Wilson</td>
<td>L. Bottura</td>
<td>Coffee</td>
<td>M. Green</td>
<td>A. Jain</td>
<td>A. Jain</td>
<td>M. Ross</td>
</tr>
<tr>
<td>11.00</td>
<td>Sources of field errors</td>
<td>Overview of measurement methods</td>
<td>Excursion</td>
<td>Sinking and ageing</td>
<td>Hall generators</td>
<td>Magnetic resonance techniques</td>
<td></td>
</tr>
<tr>
<td>12.00</td>
<td>A. Devred</td>
<td>K. Herichsen</td>
<td>MID-DAY BREAK</td>
<td>M. Mayoud</td>
<td>J. Kvitkovic</td>
<td>C. Raymond</td>
<td></td>
</tr>
<tr>
<td>16.00</td>
<td>Materials</td>
<td></td>
<td></td>
<td>Series measurements</td>
<td>Detector alignment</td>
<td>Detector magnet measurement</td>
<td></td>
</tr>
<tr>
<td>17.00</td>
<td>J. Billan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.30</td>
<td>Seminar Biological effects of magnetic fields</td>
<td>S. Zannella</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.30</td>
<td>Welcome Reception</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

EVENING MEAL  BANQUET  MEAL
FOREWORD

The aim of the CERN Accelerator School to collect, preserve and disseminate the knowledge accumulated in the world's accelerator laboratories applies not only to accelerators and storage rings, but also to the related sub-systems, equipment and technologies. This wider aim is being achieved by means of the specialised courses listed in the Table below. The latest of these was on the topic of Measurement and Alignment of Accelerator and Detector Magnets, Europa Palace Hotel, Anacapri, Italy, 11–17 April 1997, its proceedings forming the present volume.

In the first of the CAS courses on magnetic measurement and alignment the emphasis was on the basic principles and on the classical methods of achieving these tasks. Progress, however, has been so rapid in these fields over the last few years—especially regarding the use of superconducting magnets—that the International Magnet Measurement Workshop asked CAS to organise another course, again starting with the basic theory, but this time including the latest measurement and alignment techniques for both conventional and superconducting accelerator and detector magnets. The lectures resulting from this request were presented at the course in Anacapri and have been written-up for the use of present and future workers in these disciplines.

CAS wishes to thank the International Magnet Measurement Workshop, especially its representative K.N. Henrichsen, for their encouragement in organising this course. Together with the generous support we have learned to expect from the CERN Management, we also acknowledge the very considerable financial and organisational support of the INFN, Sezione di Napoli, the Universita degli Studi di Napoli, Federico II, the Dipartimento di Scienze Fisiche, Napoli—especially Prof. V.G. Vaccaro—and the local government of the Naples region. The management and staff of the Europa Palace Hotel also played an important part in ensuring that the course ran so smoothly. The CAS Advisory, Programme and Local Organising committees also deserve our sincere thanks for their guidance and attention to detail. However, the tremendous amount of work carried out by the lecturers of this course in preparing, presenting and writing-up their topics was paramount to the success of this course and deserves not only the thanks of CAS and the participants to the course but also of the many people who will use the proceedings in the future. Finally, all the effort which went into preparing this course was rewarded by the keen interest of the participants who came from all over the world to attend this course.

S. Turner, Editor
LIST OF SPECIALISED CAS COURSES AND THEIR PROCEEDINGS

<table>
<thead>
<tr>
<th>Year</th>
<th>Course</th>
<th>Proceedings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1989</td>
<td>Synchrotron radiation and free-electron lasers</td>
<td>CERN 90-03 (1990)</td>
</tr>
<tr>
<td>1990</td>
<td>Power converters for particle accelerators</td>
<td>CERN 90-07 (1990)</td>
</tr>
<tr>
<td>1991</td>
<td>RF engineering for particle accelerators</td>
<td>CERN 92-03 (1992)</td>
</tr>
<tr>
<td>1992</td>
<td>Magnetic measurement and alignment</td>
<td>CERN 92-05 (1992)</td>
</tr>
<tr>
<td>1993</td>
<td>RF engineering for particle accelerators (repeat of the 1991 course)</td>
<td>CERN 96-02 (1996)</td>
</tr>
<tr>
<td>1994</td>
<td>Cyclotrons, linacs and their applications</td>
<td>CERN 96-03 (1996)</td>
</tr>
<tr>
<td>1995</td>
<td>Superconductivity in particle accelerators</td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>Synchrotron radiation and free-electron lasers</td>
<td>(To be published)</td>
</tr>
<tr>
<td>1997</td>
<td>Measurement and alignment of accelerator and detector magnets</td>
<td>Present volume</td>
</tr>
</tbody>
</table>
CONTENTS

Foreword

A.K. Jain
Basic theory of magnets
- Multipole expansion of two-dimensional field 1
- Two-dimensional behaviour of the integral field 3
- Normal and skew components 4
- Transformation of field parameters 6
- Relationship between current and magnetic field 10
- The complex potential 11
- Field due to some simple current distributions 11
- Generating pure dipole and quadrupole fields using overlapping cylinders 16
- Multipole expansion of field due to a current filament 17
- Generating a pure 2m-pole field 20
- Current symmetries and allowed harmonics 21

E.J.N. Wilson
The motivation for magnet measurements 27
- Introduction 27
- Magnetic rigidity 28
- Focusing 29
- The gutter analogy 30
- Transverse equation of motion 31
- Circle diagram 32
- Closed orbit distortion 33
- Closed orbit in the circle diagram 34
- Sources of distortion 34
- Uncorrelated errors 34
- Magnification of errors 35
- Multipole field expansion 36
- Working diagram 38
- Phase space trajectory for 1/3 resonance 39
- Injection studies at FNAL 41

A. Devred
Review of superconducting storage-ring dipole and quadrupole magnets 43
- Types of storage-ring magnets 43
- Storage rings and superconductivity 44
- Conductor and conductor insulation 47
- Magnetic design 51
- Field quality 58
- Mechanical design 62
- Magnet cooling 66
- Quench performance 67
- Quench protection 68
- Brief summary 71
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect of coil construction errors</td>
<td>195</td>
</tr>
<tr>
<td>Deviation of the rotation axis from the magnetic axis</td>
<td>207</td>
</tr>
<tr>
<td>Calibration of harmonic coils</td>
<td>213</td>
</tr>
<tr>
<td><strong>C. Reymond</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Magnetic resonance techniques</strong></td>
<td>219</td>
</tr>
<tr>
<td>Basic NMR reminder</td>
<td>219</td>
</tr>
<tr>
<td>NMR magnetometer principle</td>
<td>221</td>
</tr>
<tr>
<td>Measuring range and accuracy</td>
<td>225</td>
</tr>
<tr>
<td>Measured-field property</td>
<td>226</td>
</tr>
<tr>
<td>Some special measurement conditions</td>
<td>227</td>
</tr>
<tr>
<td><strong>J. Kvitkovic</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Hall generators</strong></td>
<td>233</td>
</tr>
<tr>
<td>Introduction</td>
<td>233</td>
</tr>
<tr>
<td>Galvanomagnetic phenomena</td>
<td>233</td>
</tr>
<tr>
<td>Fabrication of Hall generators</td>
<td>235</td>
</tr>
<tr>
<td>Parameters of Hall generator</td>
<td>238</td>
</tr>
<tr>
<td>Hall magnetometry</td>
<td>243</td>
</tr>
<tr>
<td>Hall generator applications</td>
<td>245</td>
</tr>
<tr>
<td>Conclusions</td>
<td>247</td>
</tr>
<tr>
<td><strong>P. Sievers</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Finding the axis</strong></td>
<td></td>
</tr>
<tr>
<td>Contribution not received</td>
<td></td>
</tr>
<tr>
<td><strong>G. Moritz</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Mechanical equipment</strong></td>
<td>251</td>
</tr>
<tr>
<td>Introduction</td>
<td>251</td>
</tr>
<tr>
<td>The influence of mechanical tolerances on the accuracy of magnetic</td>
<td>251</td>
</tr>
<tr>
<td>measurements</td>
<td></td>
</tr>
<tr>
<td>Strategies for reducing the influence of mechanical errors</td>
<td>254</td>
</tr>
<tr>
<td>Special equipment versus universal equipment</td>
<td>256</td>
</tr>
<tr>
<td>Mechanical components</td>
<td>256</td>
</tr>
<tr>
<td>Benches for the measurement of the mechanical properties of magnets</td>
<td>256</td>
</tr>
<tr>
<td>Benches for the measurement of the magnetic properties of magnets</td>
<td>257</td>
</tr>
<tr>
<td><strong>P. Wanderer</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Magnet measurements for series production</strong></td>
<td>273</td>
</tr>
<tr>
<td>Introduction</td>
<td>273</td>
</tr>
<tr>
<td>Planning for series measurements</td>
<td>273</td>
</tr>
<tr>
<td>Low rate production measurements</td>
<td>277</td>
</tr>
<tr>
<td>High rate production</td>
<td>279</td>
</tr>
<tr>
<td>After completion of magnet production</td>
<td>285</td>
</tr>
<tr>
<td><strong>D. Newton</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Detector magnet measurements</strong></td>
<td>287</td>
</tr>
<tr>
<td>Detector magnets and accelerator magnets</td>
<td>287</td>
</tr>
<tr>
<td>Stored energies and pole-tip forces</td>
<td>288</td>
</tr>
<tr>
<td>Detector magnets and external measuring machines</td>
<td>288</td>
</tr>
<tr>
<td>Solenoid magnets and an internal measuring machine</td>
<td>290</td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Use of the field map</td>
<td>292</td>
</tr>
<tr>
<td>The field map with hindsight</td>
<td>294</td>
</tr>
<tr>
<td>Prospects for future detector magnets</td>
<td>295</td>
</tr>
<tr>
<td><strong>J. Iliffe</strong></td>
<td></td>
</tr>
<tr>
<td>Quality control of observational data in metrology</td>
<td>297</td>
</tr>
<tr>
<td>Introduction</td>
<td>297</td>
</tr>
<tr>
<td>Systematic errors</td>
<td>298</td>
</tr>
<tr>
<td>Random errors</td>
<td>299</td>
</tr>
<tr>
<td>Gross errors</td>
<td>304</td>
</tr>
<tr>
<td>Conclusions</td>
<td>308</td>
</tr>
<tr>
<td><strong>R.E. Ruland</strong></td>
<td></td>
</tr>
<tr>
<td>Setting reference targets</td>
<td>309</td>
</tr>
<tr>
<td>Introduction</td>
<td>309</td>
</tr>
<tr>
<td>Reference targets</td>
<td>309</td>
</tr>
<tr>
<td>Trajectory design space</td>
<td>313</td>
</tr>
<tr>
<td>Fiducialization</td>
<td>319</td>
</tr>
<tr>
<td>Superconducting magnet fiducialization monitoring</td>
<td>324</td>
</tr>
<tr>
<td>Case studies</td>
<td>326</td>
</tr>
<tr>
<td>Conclusion</td>
<td>329</td>
</tr>
<tr>
<td><strong>M. Mayoud</strong></td>
<td></td>
</tr>
<tr>
<td>Implementation and maintenance of the alignment of accelerators</td>
<td>333</td>
</tr>
<tr>
<td>(presented as “sinking and ageing”)</td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>333</td>
</tr>
<tr>
<td>Catalogue of forces acting against man-made constructions and machinery</td>
<td>333</td>
</tr>
<tr>
<td>Basics on positioning, dimensional and physical geodesy</td>
<td>334</td>
</tr>
<tr>
<td>Survey and alignment tolerances for accelerators</td>
<td>335</td>
</tr>
<tr>
<td>Network structures for radial (horizontal) control</td>
<td>335</td>
</tr>
<tr>
<td>Vertical control</td>
<td>338</td>
</tr>
<tr>
<td>Methodological aspects of large networks</td>
<td>339</td>
</tr>
<tr>
<td>Stochastic analysis and comparative surveys</td>
<td>340</td>
</tr>
<tr>
<td>Radial/vertical smoothing in large accelerators</td>
<td>341</td>
</tr>
<tr>
<td>Conclusions</td>
<td>342</td>
</tr>
<tr>
<td><strong>M.J. Price</strong></td>
<td></td>
</tr>
<tr>
<td>Alignment of experiments</td>
<td>343</td>
</tr>
<tr>
<td>Introduction</td>
<td>343</td>
</tr>
<tr>
<td>Physics and detectors</td>
<td>344</td>
</tr>
<tr>
<td>Tools</td>
<td>347</td>
</tr>
<tr>
<td>Systems</td>
<td>352</td>
</tr>
<tr>
<td><strong>M. Ross</strong></td>
<td></td>
</tr>
<tr>
<td>Alignment by feedback</td>
<td>357</td>
</tr>
<tr>
<td>Introduction</td>
<td>357</td>
</tr>
<tr>
<td>Linear collider tolerances</td>
<td>358</td>
</tr>
<tr>
<td>Feedback</td>
<td>362</td>
</tr>
<tr>
<td>Experience at SLC with feedback</td>
<td>365</td>
</tr>
<tr>
<td>Applications – Next Linear Collider</td>
<td>372</td>
</tr>
<tr>
<td>Conclusion</td>
<td>374</td>
</tr>
</tbody>
</table>
S. Zannella

**Biological effects of magnetic fields**

- Introduction
- The magnetic field
- The earth's magnetic field
- Artificial magnetic fields
- Interaction mechanisms between magnetic fields and biologic systems
- Laboratory studies
- Epidemiological studies
- Static and ELF magnetic field exposure guidelines and protective measures
- Conclusions

R. Penco

**Superconducting magnet fabrication and quality control**

- Introduction
- Production risks
- Prototypes
- Quality control during the production
- Analysis of real series production
- Conclusion

M. Poole

**Permanent magnets**

Contribution not received

P. Signoret

**Modern industrial survey techniques**

Contribution not received

List of participants
FOREWORD

The aim of the CERN Accelerator School to collect, preserve and disseminate the knowledge accumulated in the world’s accelerator laboratories applies not only to accelerators and storage rings, but also to the related sub-systems, equipment and technologies. This wider aim is being achieved by means of the specialised courses listed in the Table below. The latest of these was on the topic of Measurement and Alignment of Accelerator and Detector Magnets, Europa Palace Hotel, Anacapri, Italy, 11–17 April 1997, its proceedings forming the present volume.

List of specialised CAS courses and their proceedings

<table>
<thead>
<tr>
<th>Year</th>
<th>Course</th>
<th>Proceedings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CERN 87-01 (1987) also</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lecture Notes in Earth Sciences 12,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Springer Verlag, 1987)</td>
</tr>
<tr>
<td>1989</td>
<td>Synchrotron radiation and free-electron lasers</td>
<td>CERN 90-03 (1990)</td>
</tr>
<tr>
<td>1990</td>
<td>Power converters for particle accelerators</td>
<td>CERN 90-07 (1990)</td>
</tr>
<tr>
<td>1991</td>
<td>RF engineering for particle accelerators</td>
<td>CERN 92-03 (1992)</td>
</tr>
<tr>
<td>1992</td>
<td>Magnetic measurement and alignment</td>
<td>CERN 92-05 (1992)</td>
</tr>
<tr>
<td></td>
<td>(repeat of the 1991 course)</td>
<td></td>
</tr>
<tr>
<td>1993</td>
<td>RF engineering for particle accelerators</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(repeat of the 1991 course)</td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td>Cyclotrons, linacs and their applications</td>
<td>CERN 96-02 (1996)</td>
</tr>
<tr>
<td>1995</td>
<td>Superconductivity in particle accelerators</td>
<td>CERN 96-03 (1996)</td>
</tr>
<tr>
<td>1996</td>
<td>Synchrotron radiation and free-electron lasers</td>
<td>(To be published)</td>
</tr>
<tr>
<td>1997</td>
<td>Measurement and alignment of accelerator and</td>
<td>Present volume</td>
</tr>
<tr>
<td></td>
<td>detector magnets</td>
<td></td>
</tr>
</tbody>
</table>

S. Turner
Editor
BASIC THEORY OF MAGNETS

Animesh K. Jain
RHIC Project, Brookhaven National Laboratory, USA

Abstract
The representation of two-dimensional magnetic field in a current free region is discussed in terms of a harmonic expansion. The expansion is derived for cylindrical components and extended to Cartesian components. The Cartesian components are also described in terms of a complex field. The rules for transformation of the expansion coefficients under various types of coordinate transformation are given. The relationship between a given current distribution and the resulting field harmonics is explored in terms of the vector and complex potentials. Explicit results are presented for some simple geometries. Finally, the harmonics allowed under various symmetries in the magnet current are discussed.

1. MULTIPOLE EXPANSION OF TWO-DIMENSIONAL FIELD

For most practical purposes related to magnetic measurements in an accelerator magnet, one is interested in the magnetic field in the aperture of the magnet, which is in vacuum and carries no current. Also, most accelerator magnets tend to be long compared to their aperture. Thus, a two-dimensional description is valid for most part of the magnet, except at the ends. We shall at first confine ourselves to a description of a purely two-dimensional field in a current free region. The relationship between the field and the currents will be treated later in Sec. 5 onwards.

In free space, with no true currents, the curl of the magnetic field, \( \mathbf{H} \), is zero. Also, the magnetic induction, \( \mathbf{B} \), is given by \( \mu_0 \mathbf{H} \), where \( \mu_0 = 4\pi \times 10^{-7} \) Henry/m is the permeability of free space. Consequently, the magnetic induction, \( \mathbf{B} \), can be expressed as the gradient of a magnetic scalar potential, \( \Phi_m \):

\[
\nabla \times \mathbf{H} = 0 \Rightarrow \nabla \times \mathbf{B} = 0 \Rightarrow \mathbf{B} = -\nabla \Phi_m
\]  

(1)

The magnetic induction, \( \mathbf{B} = \mu_0 \mathbf{H} \) also has a zero divergence. Combining this fact with Eq.(1), we get Laplace's equation for the scalar potential:

\[
\nabla^2 \Phi_m = 0
\]  

(2)

We choose a cylindrical coordinate system with the Z-axis along the length of the magnet and the origin located centrally in the magnet aperture. For a two-dimensional field having no axial component, the scalar potential has no z-dependence. Laplace’s equation can be solved by separation of variables and imposing the boundary conditions that \( \Phi_m \) be periodic in the angular coordinate, \( \theta \), and be finite at \( r = 0 \). The resulting general solution is a series expansion of the scalar potential. The radial and the azimuthal components of \( \mathbf{B} \) are then obtained by taking the gradient of the scalar potential. The components of \( \mathbf{B} \) in cylindrical coordinates can be written in the form:
\[ B_r(r, \theta) = \left( \frac{\partial \Phi_m}{\partial r} \right) = \sum_{n=1}^{\infty} C(n) \left( \frac{r}{R_{\text{ref}}} \right)^{n-1} \sin[n(\theta - \alpha_n)] \]  

(3)

\[ B_\theta(r, \theta) = \left( \frac{1}{r} \frac{\partial \Phi_m}{\partial \theta} \right) = \sum_{n=1}^{\infty} C(n) \left( \frac{r}{R_{\text{ref}}} \right)^{n-1} \cos[n(\theta - \alpha_n)] \]  

(4)

where \( C(n) \) and \( \alpha_n \) are constants and \( R_{\text{ref}} \) is an arbitrary reference radius, typically chosen to be 50-70% of the magnet aperture. For a given \( r \), the \( n \)-th term in \( B_r(r, \theta) \) has \( n \) maxima and \( n \) minima as a function of the azimuthal angle \( \theta \). These angular positions may be regarded as the locations of magnetic poles. Thus, the \( n \)-th terms in Eqs. (3) and (4) correspond to a \( 2n \)-pole field. Accordingly, \( C(n) \) is said to be the amplitude of the \( 2n \)-pole component of the total field. The locations of the south and the north poles in the \( 2n \)-pole field are:

\[ \theta = \frac{\pi}{2n} + \alpha_n; \frac{5\pi}{2n} + \alpha_n; \frac{9\pi}{2n} + \alpha_n; \ldots \text{ SOUTH POLES} \]  

(5)

\[ \theta = \frac{3\pi}{2n} + \alpha_n; \frac{7\pi}{2n} + \alpha_n; \frac{11\pi}{2n} + \alpha_n; \ldots \text{ NORTH POLES} \]  

(6)

The parameter \( \alpha_n \) defines the orientation of the \( 2n \)-pole component of the field with respect to the chosen \( X \)-axis and is called the phase angle. Eqs. (3) and (4) represent the multipole expansion of the components of a two-dimensional field in a current free region.

### 1.1 The Cartesian components

![Cylindrical and Cartesian Components of the magnetic induction vector, B](image)

Fig.1 Cylindrical and Cartesian Components of the magnetic induction vector, \( B \)

Accelerator physicists often prefer to work with the Cartesian components of the field. The relationship between the Cartesian and the cylindrical components is shown in Fig.1. Using the expansions for \( B_r \) and \( B_\theta \) from Eqs. (3) and (4), we get the Cartesian components:
\[ B_x(r, \theta) = B_r \cos \theta - B_\theta \sin \theta = \sum_{n=1}^{\infty} C(n) \left( \frac{r}{R_{\text{ref}}} \right)^{n-1} \sin[(n-1)\theta - n\alpha_n] \]  

(7)

\[ B_y(r, \theta) = B_r \sin \theta + B_\theta \cos \theta = \sum_{n=1}^{\infty} C(n) \left( \frac{r}{R_{\text{ref}}} \right)^{n-1} \cos[(n-1)\theta - n\alpha_n] \]  

(8)

1.2 The complex field

The components of a two-dimensional field can be very conveniently described in terms of a complex field, \( B(z) \), defined as a function of the complex variable \( z = x + iy = r \exp(i\theta) \) [a bold and italic font is used for complex quantities in this paper]. Looking at the expansions in Eqs.(7) and (8), one could define the complex field as:

\[ B(z) = B_y(x, y) + iB_x(x, y) = \sum_{n=1}^{\infty} C(n) \exp(-in\alpha_n) \left( \frac{z}{R_{\text{ref}}} \right)^{n-1} \]  

(9)

The divergence and the curl (in a source free region) of the vector \( \mathbf{B} \) are zero. In two dimensions, we get:

\[ \nabla \cdot \mathbf{B} = 0 \Rightarrow \left( \frac{\partial B_x}{\partial x} \right) + \left( \frac{\partial B_y}{\partial y} \right) = 0; \text{ or } \left( \frac{\partial B_y}{\partial x} \right) = -\left( \frac{\partial B_x}{\partial y} \right) \]  

(10)

\[ \nabla \times \mathbf{B} = 0 \Rightarrow \left( \frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y} \right) = 0; \text{ or } \left( \frac{\partial B_z}{\partial x} \right) = \left( \frac{\partial B_x}{\partial y} \right) \]  

(11)

Eqs.(10) and (11) are nothing but the well known Cauchy-Riemann conditions for the complex field \( B(z) \) to be an analytic function of the complex variable \( z \). The complex field is sometimes defined as \( B_z(x, y) - iB_y(x, y) \), which is also an analytic function of \( z \). It should be noted that the quantity \( B_x + iB_y \) is not an analytic function of \( z \). The analytic property of the complex field has been exploited in solving many two-dimensional problems [1–6].

2. TWO-DIMENSIONAL BEHAVIOUR OF THE INTEGRAL FIELD

In magnets of a finite length, the two-dimensional representation of the field is valid only in the body of the magnet, sufficiently away from the ends. In regions near the ends of the magnet, the field is three dimensional and the usual multipole expansion is no longer valid. Some examples of a three-dimensional treatment may be found in Refs. [7–9]. However, for most practical purposes, one is interested in the integral of the field (or of its derivatives) over the length of the magnet. This is because most magnets are short compared to the wavelength of the betatron oscillations in the machine and the details of axial variation are of little consequence. Also, a typical rotating coil of a finite length only measures integral of the field over its length. It can be shown that the integral field essentially behaves as a two-dimensional field provided the integration is carried out over an appropriate region [10].

In general, for three-dimensional fields in a current free region, the scalar potential satisfies the Laplace’s equation:
\[ \nabla^2 \Phi_m(x, y, z) = \left[ \frac{\partial^2 \Phi_m}{\partial x^2} + \frac{\partial^2 \Phi_m}{\partial y^2} + \frac{\partial^2 \Phi_m}{\partial z^2} \right] = 0 \quad (12) \]

Integrating along the Z-axis from \( Z_i \) to \( Z_f \), we get:

\[ \int_{Z_i}^{Z_f} \left[ \frac{\partial^2 \Phi_m}{\partial x^2} + \frac{\partial^2 \Phi_m}{\partial y^2} + \frac{\partial^2 \Phi_m}{\partial z^2} \right] dz = \left[ \frac{\partial^2 \Phi_m}{\partial x^2} + \frac{\partial^2 \Phi_m}{\partial y^2} \right] Z_f - \left[ \frac{\partial^2 \Phi_m}{\partial x^2} + \frac{\partial^2 \Phi_m}{\partial y^2} \right] Z_i = 0 \quad (13) \]

We define the z-integrated scalar potential as \( \Phi_m(x, y) = \int_{Z_i}^{Z_f} \Phi_m(x, y, z) \, dz \). From Eq.(13):

\[ \left[ \frac{\partial^2 \Phi_m}{\partial x^2} + \frac{\partial^2 \Phi_m}{\partial y^2} \right] = - \left[ \frac{\partial \Phi_m}{\partial z} \right]_{Z_i}^{Z_f} = B_z(x, y, Z_f) - B_z(x, y, Z_i) \quad (14) \]

If the region of integration is so chosen that the Z-component of the field is zero at the boundaries of this region, then the right hand side vanishes and the z-integrated scalar potential satisfies the two-dimensional Laplace's equation. For example, the points \( Z_i \) and \( Z_f \) could both be chosen well outside the magnet on opposite ends to include the integral of the field over the entire magnet. This situation is commonly encountered in the measurement of integral field of relatively short magnets with a long integral coil. Alternatively, one could choose \( Z_i \) well outside the magnet and \( Z_f \) well inside the magnet, where the field is again two-dimensional. Such a situation would apply to the measurement of the end region in a long magnet with a short measuring coil. The measuring coil for this purpose must have sufficient length so that the condition in Eq.(14) can be satisfied.

3. NORMAL AND SKEW COMPONENTS

The multipole expansion of the complex field is given by Eq.(9). The normal and skew components are defined as the real and the imaginary parts of the expansion coefficients:

\[ C(n) \exp(-i n \alpha_n) = (2n\text{-pole NORMAL Term}) + i(2n\text{-pole SKEW Term}) \quad (15) \]

Unfortunately, the index \( n \) in the expansion coefficient is not the same as the corresponding power \((n-1)\) of \( Z \) in Eq.(9). This has led to two different conventions being followed in denoting the normal and the skew terms. The “US Convention” denotes the 2\( n \)-pole normal and skew terms with an index of \((n-1)\), to match the corresponding power of \( Z \) in Eq.(9):

\[ \begin{align*}
2n \text{-pole Normal Term} &= C(n) \cos(n \alpha_n) = B_{n-1} \quad \text{“U.S. CONVENTION”} \\
2n \text{-pole Skew Term} &= - C(n) \sin(n \alpha_n) = A_{n-1}
\end{align*} \quad (16) \]

On the other hand, the “European Convention” denotes the 2\( n \)-pole normal and skew components with an index of \( n \) to retain the simple relationship between the index and the number of poles:

\[ \begin{align*}
2n \text{-pole Normal Term} &= C(n) \cos(n \alpha_n) = B_n \\
2n \text{-pole Skew Term} &= - C(n) \sin(n \alpha_n) = A_n \quad \text{“EUROPEAN CONVENTION”}
\end{align*} \quad (17) \]
Even though the two conventions are commonly referred to as the "US" and the "European" notations, their use is not necessarily restricted to the respective geographic regions. This calls for exercising caution in interpreting the notations. One possible indicator of the notation is the presence or the absence of $B_0$ and $A_0$ terms.

In terms of the normal and the skew components, the expansion of $B(z)$ is:

$$B(z) = B_x + iB_y = \sum_{n=0}^{\infty} \left[ B_n + iA_n \right] \left( \frac{z}{R_{ref}} \right)^n \quad \text{"U.S. CONVENTION"} \quad (18)$$

$$B(z) = B_x + iB_y = \sum_{n=1}^{\infty} \left[ B_n + iA_n \right] \left( \frac{z}{R_{ref}} \right)^{n-1} \quad \text{"EUROPEAN CONVENTION"} \quad (19)$$

The corresponding equations for the radial and the azimuthal components of the field are:

$$B_r(r, \theta) = \sum_{n=0}^{\infty} \left( \frac{r}{R_{ref}} \right)^n \left[ B_n \sin((n+1)\theta) + A_n \cos((n+1)\theta) \right] \quad \text{"U.S. Convention"} \quad (20)$$

$$B_\theta(r, \theta) = \sum_{n=0}^{\infty} \left( \frac{r}{R_{ref}} \right)^n \left[ B_n \cos((n+1)\theta) - A_n \sin((n+1)\theta) \right] \quad \text{"U.S. Convention"} \quad (21)$$

$$B_r(r, \theta) = \sum_{n=1}^{\infty} \left( \frac{r}{R_{ref}} \right)^{n-1} \left[ B_n \sin(n\theta) + A_n \cos(n\theta) \right] \quad \text{"European Convention"} \quad (22)$$

$$B_\theta(r, \theta) = \sum_{n=1}^{\infty} \left( \frac{r}{R_{ref}} \right)^{n-1} \left[ B_n \cos(n\theta) - A_n \sin(n\theta) \right] \quad \text{"European Convention"} \quad (23)$$

It should be noted that sometimes the skew component is defined with a sign opposite to that used here [11]. In view of the different notations in existence, it is important to recognize the convention being used in any particular work.

It is possible to assign a physical significance to the normal and the skew components in terms of the derivatives of the field. Using Eq.(18) or Eq.(19), it is easy to show that

$$B_n(\text{US}) = B_{n+1}(\text{European}) = \frac{R_{ref}^n}{n!} \left( \frac{\partial^n B_y}{\partial x^n} \right)_{x=0; y=0} \quad (24)$$

$$A_n(\text{US}) = A_{n+1}(\text{European}) = \frac{R_{ref}^n}{n!} \left( \frac{\partial^n B_x}{\partial x^n} \right)_{x=0; y=0}$$

A magnet with the $2m$-pole term as the most dominant term is called a normal magnet.
if the $2m$-pole skew term is zero. Similarly, such a magnet is called a skew magnet if the normal term is zero. As per Eqs.(16) and (17), the possible values of the phase angle for the $2m$-pole term are $\alpha_n = 0$ or $\pi/m$ for a normal magnet and $\alpha_n = \pi/(2m)$ or $3\pi/(2m)$ for a skew magnet. The poles in a $2m$-pole magnet [see Eqs.(5) and (6)] are located at $\theta = \pi/(2m), 3\pi/(2m), 5\pi/(2m)$, etc. for a normal magnet and at $\theta = 0, \pi/m, 2\pi/m, 3\pi/m$, etc. for a skew magnet. A $2m$-pole skew magnet is obtained from a $2m$-pole normal magnet by a clockwise rotation of the magnet by an angle $\pi/(2m)$.

3.1 Fractional field coefficients, or “multipoles”

The expansion coefficients $B_n$ and $A_n$ in Eqs.(18)–(23) are related to the actual field strength in the magnet and are dependent on the excitation level. In describing the field quality of a magnet, one is often interested in the shape of the field, rather than its absolute magnitude. This is done by expressing the various harmonic terms in the expansion as a fraction of a reference field, $B_{ref}$. For example, the complex field may be written as:

$$B(z) = B_y(x, y) + iB_x(x, y) = B_{ref} \sum_{n=1}^{\infty} \left[ \frac{C(n)\exp(-in\alpha_n)}{B_{ref}} \right] \left( \frac{z}{R_{ref}} \right)^{n-1}$$

(25)

Generally, this reference field is chosen as the strength of the most dominant term in the expansion. For a $2m$-pole magnet, it is expected that the term for $n = m$ will be the most dominant one. Hence, $B_{ref}$ may be chosen to be equal to $C(m)$. Sometimes the reference field is chosen as the strength of the dipole field used for bending the beam in a circular accelerator.

The Normal and Skew $2n$-pole fractional field coefficients, or “multipoles” are defined as:

$$b_{n-1} = \text{Re} \left[ \frac{C(n)\exp(-in\alpha_n)}{B_{ref}} \right] = \frac{B_{n-1}}{B_{ref}}; \quad a_{n-1} = \text{Im} \left[ \frac{C(n)\exp(-in\alpha_n)}{B_{ref}} \right] = \frac{A_{n-1}}{B_{ref}}$$

(US) (26)

$$b_n = \text{Re} \left[ \frac{C(n)\exp(-in\alpha_n)}{B_{ref}} \right] = \frac{B_n}{B_{ref}}; \quad a_n = \text{Im} \left[ \frac{C(n)\exp(-in\alpha_n)}{B_{ref}} \right] = \frac{A_n}{B_{ref}}$$

(European) (27)

In a typical accelerator magnet, these coefficients are of the order of $10^4$ when calculated at a reference radius comparable to the size of the region occupied by the beam (~50–70% of the coil radius). The coefficients are therefore often quoted after multiplying by a factor of $10^4$. With this multiplicative factor, the values of the multipoles are said to be in “units”.

4. TRANSFORMATION OF FIELD PARAMETERS

The expansion parameters [$C(n)$, $\alpha_n$] or [$B_n$, $A_n$] depend on the choice of the reference frame. For example, the parameter $\alpha_n$ controls the angular locations of the poles for the $2n$-pole component of the field according to Eqs.(5) and (6). If the coordinate axes (or equivalently, the magnet) are rotated, the angular positions of the poles would change, thus necessitating a corresponding change in the parameter $\alpha_n$. In general, the quantities $B_n$ and $A_n$ should be regarded as expansion coefficients according to Eq.(18) or (19), which clearly depend on the choice of coordinate axes. In this section, we shall derive the equations that govern the transformation of these parameters under commonly encountered coordinate transformations.
4.1 Displacement of axes

Let us consider a frame $X'-Y'$ which is displaced from a frame $X-Y$ by $z_0 = x_0 + iy_0$ as shown in Fig. 2. The axes in the two frames are assumed to be parallel. In other words, there is no rotation of the axes. The field parameters are denoted by $[C(n), \alpha_n]$ or $[B_n, A_n]$ in the $X-Y$ frame and by $[C'(n), \alpha'_n]$ or $[B'_n, A'_n]$ in the $X'-Y'$ frame.

We consider a point in space located at $z = x + iy$ in the $X-Y$ frame. The same point is located at $z' = x' + iy' = z - z_0$ in the $X'-Y'$ frame, as shown in Fig. 2. Since the coordinate axes in the two frames are parallel to each other, the Cartesian components of the field are the same in the two frames. In other words, the complex field is the same in the two frames. This fact can be used to derive the expansion of field in the $X'-Y'$ frame as follows:

$$B(z') = B_{x'} + iB_{y'} = B_y + iB_x = B(z)$$

$$= \sum_{k=0}^{\infty} (B_k + iA_k) \left( \frac{z}{R_{\text{ref}}} \right)^k = \sum_{k=0}^{\infty} (B_k + iA_k) \left( \frac{z' + z_0}{R_{\text{ref}}} \right)^k$$

$$= \sum_{k=0}^{\infty} (B_k + iA_k) \frac{k!}{n!(k-n)!} \left( \frac{z'}{R_{\text{ref}}} \right)^n \left( \frac{z_0}{R_{\text{ref}}} \right)^{k-n}$$

Fig. 2 Displacement of Coordinate Axes

In writing the above equations, we have used the US convention as per Eq.(18). With a rearrangement of terms in the double summation, it can be shown that

$$\sum_{k=0}^{\infty} \sum_{n=0}^{k} t_{kn} = \sum_{n=0}^{\infty} \sum_{k=n}^{\infty} t_{kn}$$

Applying this identity to Eq.(28), we may write:

$$B(z') = \sum_{n=0}^{\infty} \left[ \sum_{k=n}^{\infty} \frac{k!}{n!(k-n)!} \left( \frac{z_0}{R_{\text{ref}}} \right)^{k-n} \left( \frac{z'}{R_{\text{ref}}} \right)^n \right] = \sum_{n=0}^{\infty} (B'_n + iA'_n) \left( \frac{z'}{R_{\text{ref}}} \right)^n$$

(30)
The expansion coefficients in the displaced frame are therefore given by:

\[(B_n' + iA_n') = \sum_{k=n}^{\infty} (B_k + iA_k) \left[ \frac{k!}{n!(k-n)!} \right] \left( \frac{x_0 + iy_0}{R_{ref}} \right)^{k-n} \quad ; \quad n \geq 0 \quad \text{(US Notation)} \quad (31)\]

The corresponding equation in the European notation is:

\[(B_n' + iA_n') = \sum_{k=n}^{\infty} (B_k + iA_k) \left[ \frac{(k-1)!}{(n-1)!(k-n)!} \right] \left( \frac{x_0 + iy_0}{R_{ref}} \right)^{k-n} \quad ; \quad n \geq 1 \quad \text{(European Notation)} \quad (32)\]

The transformation for the amplitude and phase of the 2n-pole term is given by:

\[C'(n)\exp(-in\alpha_n') = \sum_{k=n}^{\infty} [C(k)\exp(-ik\alpha_k)] \left[ \frac{(k-1)!}{(n-1)!(k-n)!} \right] \left( \frac{x_0 + iy_0}{R_{ref}} \right)^{k-n} \quad ; \quad n \geq 1 \quad (33)\]

It is seen from Eqs.(31)-(33) that coefficients of any particular order in the displaced frame are given by a combination of all the terms of equal or higher order in the undisplaced frame. In other words, a given harmonic term in the undisplaced frame contributes to all the harmonics of equal or lower order in the displaced frame. For example, a displacement in a pure quadrupole magnet will result in a dipole term in addition to the quadrupole term, a displacement in a pure sextupole field will produce quadrupole and dipole fields and so on. This effect is referred to as the feed down of harmonics.

### 4.2 Rotation of axes

Let us consider a frame $X' - Y'$ which is rotated with respect to a frame $X - Y$ by an angle $\phi$ as shown in Fig.3. The origins of the two frames are assumed to be the same. The field parameters are denoted by $[C(n), \alpha_n]$ or $[B_n, A_n]$ in the $X - Y$ frame and by $[C'(n), \alpha'_n]$ or $[B'_n, A'_n]$ in the $X' - Y'$ frame. The relationship between the two sets of expansion coefficients can be obtained by using the relationship between the complex coordinates $z$ and $z'$ and the relationship between the Cartesian components in the two frames, as was done for the case of a displacement of axes in Sec. 4.1. For the case of a rotation by an angle $\phi$, we have:

![Fig.3 Rotation of Axes](image-url)
\[ B_{x'} = B_x \cos \phi + B_y \sin \phi \quad ; \quad B_{y'} = -B_x \sin \phi + B_y \cos \phi \]  
(34)

\[ z = z' \exp(i\phi) \]  
(35)

Following the US convention, we may write

\[ B(z') = B_{y'} + iB_{x'} = (B_y + iB_x) \exp(i\phi) = \sum_{n=0}^{\infty} [B_n + iA_n] \left( \frac{z'}{R_{ref}} \right)^n \exp(i\phi) \]

\[ = \sum_{n=0}^{\infty} [B_n + iA_n] \left( \frac{z'}{R_{ref}} \right)^n \exp[i(n+1)\phi] = \sum_{n=0}^{\infty} [B'_n + iA'_n] \left( \frac{z'}{R_{ref}} \right)^n \]  
(36)

The transformation of coefficients follows immediately from the above equation:

\[ (B'_n + iA'_n) = (B_n + iA_n) \exp[i(n+1)\phi] ; \quad n \geq 0 \quad \text{(US Notation)} \]  
(37)

\[ (B'_n + iA'_n) = (B_n + iA_n) \exp(in\phi) ; \quad n \geq 1 \quad \text{(European Notation)} \]  
(38)

The transformation for the amplitude and phase of the 2n-pole term is given by:

\[ C'(n) \exp(-in\alpha'_n) = C(n) \exp(-in\alpha_n) \exp(in\phi) \Rightarrow C'(n) = C(n) ; \quad \alpha'_n = \alpha_n - \phi ; \quad n \geq 1 \]  
(39)

As seen from Eqs.(37)–(39), a rotation of axes does not produce any feed down of harmonics, but causes mixing of the normal and the skew components of a given harmonic.

4.3 Reflection of X-axis: Magnet viewed from the opposite end

If a magnet is viewed from an end which is opposite to the end from which the field parameters are measured, then appropriate transformation must be applied to the field parameters. Such a situation may be encountered in practice when all magnets in an accelerator are measured from the same end, but all magnets are not installed with the leads facing the same way. The reference frames X-Y and X'-Y' are shown in Fig.4. The Y'-axis coincides with the Y-axis but the X'-axis points away from the X-axis.

![Reflection of X-axis: Magnet viewed from the opposite end](image)

Fig. 4 Reflection of X-axis: Magnet viewed from the opposite end
To obtain the transformation, we note that
\[ z^* = x - iy = -x' - iy' = -z' \] (40)

\[ B'_{y'} + iB'_{x'} = B_y - iB_x = (B_y + iB_x)^* = \sum_{n=0}^{\infty} (B_{n} + iA_{n})^{*} \left( \frac{z^*}{R_{\text{ref}}} \right)^n \]
\[ = \sum_{n=0}^{\infty} (B_{n} - iA_{n})(-1)^n \left( \frac{z'}{R_{\text{ref}}} \right)^n = \sum_{n=0}^{\infty} (B'_{n} + iA'_{n}) \left( \frac{z'}{R_{\text{ref}}} \right)^n \] (41)

From Eq.(41), it immediately follows that
\[ B'_n = (-1)^n B_n ; \quad A'_n = (-1)^{n+1} A_n ; \quad n \geq 0 \] (US Notation) (42)

\[ B'_n = (-1)^{n+1} B_n ; \quad A'_n = (-1)^n A_n ; \quad n \geq 1 \] (European Notation) (43)

\[ C'(k) = C(k) ; \quad \alpha'_{2k-1} = -\alpha_{2k-1} ; \quad \alpha'_{2k} = \left( \frac{\pi}{2k} \right) - \alpha_{2k} ; \quad k \geq 1 \] (44)

It should be noted that the transformations given in Eqs.(42) and (43) are for the unnormalized normal and skew coefficients. In practice, one often deals with the normalized fractional field coefficients, or multipoles, defined in Sec.3.1. The transformation of the multipoles can be a source of considerable confusion depending on how the reference field is chosen. For example, in a normal quadrupole magnet, the main field component (the normal quadrupole term) will change sign when viewed from the opposite end according to Eq.(42) or Eq.(43). However, one may like to normalize the multipoles in such a way that the main multipole, \( b_1 \) (or \( b_2 \) in the European notation) is always positive. With such a normalization, the normal quadrupole multipole in this case will not change sign under a reflection of axes. Therefore, while deriving the transformation rules for the normalized multipoles, one should also pay attention to the normalization rules. The principal use of the measured multipoles is to provide a series expansion of the field for use in accelerator physics studies. The transformation rules for the multipoles must be derived in such a way that the form of the expansion used in such studies remains valid.

5. RELATIONSHIP BETWEEN CURRENT AND MAGNETIC FIELD

The harmonic expansion of the components of the magnetic induction \( \mathbf{B} \), derived in Sec. 1 made no reference to the current distribution that produced the field. In practice, the distribution of current in an electromagnet determines the shape of the field, and hence the actual values of the normal and skew components, or the amplitudes and phases of various harmonics. From the point of view of magnet design as well as magnetic measurements, it is essential to have an understanding of the relationship between current distribution in a magnet and the field harmonics.

In order to study the relationship between current and the field shape, it is convenient to describe the magnetic induction \( \mathbf{B} \) in terms of the vector potential, \( \mathbf{A} \). From Maxwell’s equations, the divergence of \( \mathbf{B} \) is zero everywhere. So, we may write
\[ \mathbf{B} = \nabla \times \mathbf{A} \quad (\text{ALWAYS HOLDS}) \] (45)
If we restrict ourselves to regions where \( B = \mu_0 H \), then we also have

\[
\nabla \times B = \nabla \times (\nabla \times A) = \nabla (\nabla \cdot A) - \nabla^2 A = \mu_0 (\nabla \times H) = \mu_0 J
\]

(46)

where \( J \) is the current density vector. Since only the curl of \( A \) is important, we may add an arbitrary term \( \nabla \psi \) to \( A \) without affecting the results (Gauge Transformation). We choose this term \( \nabla \psi \) in such a way that the divergence of \( A \) becomes zero. With this choice of gauge, we get the Poisson’s equation for the vector potential,

\[
\nabla^2 A = -\mu_0 J
\]

(47)

Solution of the Poisson’s equation is given by [12]

\[
A(r) = \left( \frac{\mu_0}{4\pi} \right) \int \frac{J(r')}{|r - r'|} dr'
\]

(48)

If the distribution of current is known, one can obtain the vector potential, and hence the magnetic induction, \( B \). In the presence of magnetic materials, the current density in Eq.(48) must also include the current loops that effectively produce the magnetization. This complicates the application of Eq.(48) in practice. We shall derive the expressions for the magnetic induction for several simple geometries of interest in Sec. 7.

6. THE COMPLEX POTENTIAL

The scalar and the vector potentials can be combined into a single complex potential. For a two-dimensional field confined to the \( X-Y \) plane, the current and the vector potential have only a \( Z \)-component. Let us define a function, \( W(z) \), as a function of the complex variable \( z \) by the relation

\[
W(z) = A_z(x,y) + i\Phi_m(x,y)
\]

(49)

where \( A_z \) is \( Z \)-component of the vector potential and \( \Phi_m \) is the magnetic scalar potential. Then,

\[
\left( \frac{dW(z)}{dz} \right) = -\left( \frac{\partial W(z)}{\partial x} \right) = -\left( \frac{\partial A_z}{\partial x} \right) - i\left( \frac{\partial \Phi_m}{\partial x} \right) = B_y(x,y) + iB_z(x,y) = B(z)
\]

(50)

The derivative of \( W(z) \) in the complex plane is the complex field \( B(z) \) defined in Eq.(9). We may think of \( W(z) \) as a complex potential. It should be noted that slightly different versions of the complex potential exist in the literature depending on the definition of the complex field. The complex potential for simple current geometries can be easily calculated, leading to several useful results, as we shall see in the next section. It should be noted that the complex potential involves both the vector and the scalar potential, and, therefore, contains more information than is necessary to describe the field. However, it has the advantage of being an analytic function of the complex variable, \( z \).

7. FIELD DUE TO SOME SIMPLE CURRENT DISTRIBUTIONS

In this section we shall derive the expressions for the field from some very simple types of current distributions. Since we are dealing with only two-dimensional fields, all the current distributions will be assumed to have an infinite extent in the \( Z \)-direction.
7.1 An infinitessimally-thin current filament

The simplest current configuration is an infinitessimally-thin current filament carrying a current, \( I \), located at the origin and extending to infinity in the positive and the negative directions along the \( Z \)-axis, as shown in Fig. 5. Although the vector potential for this configuration can be calculated by explicitly integrating Eq.(48), it is much more convenient in this case to simply utilize Ampère's law to obtain the field, and use Eq.(50) to obtain the complex potential. At any point \( P \) located at \((r, \theta)\), the magnetic field has the same magnitude along a circle of radius \( r \) and is directed along the azimuthal direction, as shown in Fig.5.

From Ampère's law:

\[
B = \mu_0 H = \frac{\mu_0 I}{2\pi r} \hat{\theta}
\]  
(51)

Using Eqs.(7) and (8) for the Cartesian components in terms of the radial and azimuthal components, it is easy to show that the complex field is given by

\[
B(z) = B_r + iB_\theta = (B_\theta + iB_r)\exp(-i\theta) = \frac{\mu_0 I}{2\pi r\exp(i\theta)} = \frac{\mu_0 I}{2\pi z}
\]  
(52)

The complex potential is obtained by integrating the complex field:

\[
W(z) = -\int B(z)dz = -\left(\frac{\mu_0 I}{2\pi}\right)\ln(z) + \text{constant} = -\left(\frac{\mu_0 I}{2\pi}\right)[\ln(r) + i\theta] + \text{constant}
\]  
(53)

The real and the imaginary parts of \( W(z) \) give the vector potential, \( A_z \), and the scalar potential, \( \Phi_n \), respectively [see Eq.(49)].

A more useful result is for a current filament located at an arbitrary point in the complex plane given by \( a = a_x + ia_y \). In a frame \( X'-Y' \) with axes parallel to \( X-Y \) and having origin at the current filament (see Fig.6), the complex field is given by Eq.(52). The complex field and the complex potential in the \( X-Y \) frame can be calculated as follows:

![Fig.5 Field due to an infinitely long, thin current filament located at the origin](image1)

![Fig.6 Calculation of field due to a current filament at an arbitrary location](image2)
\[ B(z) = B_y + iB_x = B_Y' + iB_X' = B(z') = \left( \frac{\mu_0 I}{2\pi z'} \right) = \left( \frac{\mu_0 I}{2\pi(z-a)} \right) \] (54)

\[ W(z) = \int B(z')dz = -\left( \frac{\mu_0 I}{2\pi} \right) \ln(z-a) + \text{const.} = -\left( \frac{\mu_0 I}{2\pi} \right) \ln(r') + i\theta' + \text{const.} \] (55)

Using Eq.(54), we can determine, in principle, the field due to any arbitrary distribution of current by treating it as a superposition of current filaments. We shall see some simple examples in the rest of this section.

7.2 Cylindrical current sheet of uniform density

Let us calculate the complex field due to an infinitely-long, infinitesimally-thin cylindrical conductor of radius \( a \) carrying a current \( I \), as shown in Fig.7. The current density is assumed to be uniform. Also, all currents are assumed to flow parallel to the length of the cylinder, which is along the Z-axis. The complex field from a small element of angular width \( d\phi \), located at azimuthal angle \( \phi \) and carrying a current \( dl \), can be calculated using Eq.(54). The total field can then be obtained by integrating over the entire surface of the cylinder:

\[ B(z) = \left( \frac{\mu_0}{2\pi} \right) \int_0^{2\pi} \frac{dl}{z-a \exp(i\phi)} = \left( \frac{\mu_0}{2\pi} \right) \left( \frac{1}{2\pi i} \right) \int_0^{2\pi} \frac{d\phi}{z-a \exp(i\phi)} = \left( \frac{\mu_0 I}{2\pi} \right) \left( \frac{1}{2\pi i} \right) \int_0^{2\pi} \frac{dz'}{z'(z-z')} \] (56)

where we have used the transformation \( z' = a \exp(i\phi) \) and the contour integral is along the surface of the cylinder. The integrand in Eq.(56) has two simple poles at \( z' = 0 \) and \( z' = z \). If the field point \( z \) is outside the cylinder, then only the pole at \( z' = 0 \) contributes to the integral. Similarly, if the field point \( z \) is inside the cylinder, both the poles contribute to the integral. Using the method of residues, it is easy to show that the total complex field in the regions outside and inside the cylinder is given by

\[ B_{\text{out}}(z) = \left( \frac{\mu_0 I}{2\pi z} \right) \quad B_{\text{in}}(z) = 0 \] (57)

Thus, the field inside the cylinder is zero, while for any point outside, the cylinder behaves as an infinitesimally thin current filament located at the center.

7.3 Solid cylindrical conductor with uniform current density

We now consider an infinitely long solid cylinder of radius \( a \) carrying a total current \( I \) along the Z-axis, as shown in Fig.8. The field at any point can be obtained by dividing the solid cylinder into thin cylindrical shells and using the results of Sec.7.2. For any point outside the cylindrical conductor, all such shells contribute. For any point \( P \) at a radius \( r < a \), as shown in Fig.8, only shells with radii \( \xi < r \) contribute.

The current density \( J \) is given by \( I/(\pi a^2) \). The current carried by a thin shell of radius \( \xi \) and thickness \( d\xi \) is given by \( dl = J \cdot 2\pi \xi \cdot d\xi \). For any point inside the cylinder, the total complex field is:
Fig. 7 A section of a thin, cylindrical sheet carrying current \( I \) along the Z-axis.

\[
B_{\text{in}}(z) = \int_0^r \left( \frac{\mu_0 dl}{2\pi} \right) = \left( \frac{\mu_0 J}{z} \right) \int_0^r \xi d\xi = \left( \frac{\mu_0 J r}{2z} \right) = \left( \frac{\mu_0 J}{2} \right) z^{*} \tag{58}
\]

where we have used the fact that \( r^{2} = a^{2}z^{*} \). For any point outside the solid cylinder, the total complex field is:

\[
B_{\text{out}}(z) = \int_0^a \left( \frac{\mu_0 dl}{2\pi} \right) = \left( \frac{\mu_0 J}{z} \right) \int_0^a \xi d\xi = \left( \frac{\mu_0 Ja^{2}}{2z} \right) = \left( \frac{\mu_0 J}{2} \right) \left( \frac{a^{2}}{z} \right) \tag{59}
\]

It should be noted that \( B_{\in}(z) \) is a function of \( z^{*} \), and hence is not an analytic function of \( z \). This result is not unexpected, because the points inside the cylinder are not in a source free region. The analyticity of the complex field defined as \( B_{\in} + iB_{\text{out}} \) was shown in Sec. 2 to follow from Maxwell's equations in a current free region only [see Eq.(11)].

7.4 Solid conductor of arbitrary cross section

Let us now consider a solid conductor of arbitrary cross section defined by the contour \( C \) in the complex plane, as shown in Fig.9. In principle, one could calculate the field from such a conductor by integrating the result for a thin current filament [see Eq.(54)] over the cross section of the conductor. This approach requires a two-dimensional integral to be evaluated and is not always convenient. A much simplified result has been obtained by Beth[2] by defining a complex function which is analytic everywhere, including the region of the conductor, and then evaluating the field as a contour integral around the path \( C \). The result of Beth, referred to as the integral formula, is briefly summarized here.

Inside the region of the conductor, Maxwell's equations give:

\[
(\nabla \times \mathbf{B})_z = \left( \frac{\partial B_y}{\partial x} \right) - \left( \frac{\partial B_x}{\partial y} \right) = \mu_0 J_z(x, y); \quad \nabla \cdot \mathbf{B} = \left( \frac{\partial B_x}{\partial x} \right) + \left( \frac{\partial B_y}{\partial y} \right) = 0 \tag{60}
\]
where \( J(x, y) \) is the current density at the point \((x, y)\). For a constant current density over the cross section of the conductor, \( J(x, y) = J \), it can be easily shown that the function:

\[
F(z) = B_y(x, y) + iB_x(x, y) - \left( \frac{\mu_0 J}{2} \right) z^* = B(z) - \left( \frac{\mu_0 J}{2} \right) z^*
\]  

(61)

satisfies Cauchy-Riemann conditions and is, therefore, an analytic function of the complex variable, \( z \). For points outside the conductor, \( J = 0 \), and \( F(z) \) becomes identical to \( B(z) \). It has been shown by Beth[2] that this function, \( F(z) \), can be evaluated as a contour integral over the boundary of the conductor. The complex field is then calculated using Eq.(61). The final result is given by:

\[
B_{in}(z) = F(z) + \left( \frac{\mu_0 J}{2} \right) z^* = i \left( \frac{\mu_0 J}{4\pi} \right) \oint_{C} \frac{z'^*}{z' - z} \, dz' + \left( \frac{\mu_0 J}{2} \right) z^*; \text{ for } z = z_{in}
\]  

(62)

\[
B_{out}(z) = F(z) = i \left( \frac{\mu_0 J}{4\pi} \right) \oint_{C} \frac{z'^*}{z' - z} \, dz'; \text{ for } z = z_{out}
\]  

(63)

The expressions for the field are thus reduced to an integral along the perimeter of the conductor cross section, rather than a two-dimensional integral over the area of the conductor cross section. It should be noted that the integrand in Eqs. (62) and (63) contains \( z'^* \), and is not an analytic function of the complex variable. The method of residues is not applicable in general and the integral must be explicitly evaluated.

A quick verification of the integral formula can be made for the case of a solid circular cylinder, for which the expression for field was derived in Sec.7.3. For this special case, \( z'^* = a \exp(-i\phi) = a' / z' \), which is an analytical function with a simple pole at \( z' = 0 \). The integral in Eqs. (62) and (63) becomes identical to the one in Eq. (56) evaluated earlier in Sec. 7.2. Using the results of Sec. 7.2, it is easy to verify that Eqs. (62) and (63) give the same results for a solid cylindrical conductor as Eqs. (58) and (59).

![Fig.9 Section of an infinitely long solid conductor of arbitrary cross section](image)

![Fig. 10 Cross section of an infinitely long elliptical conductor](image)
7.5 Application of the integral formula: A solid elliptical conductor

A good illustration of the application of the integral formula is to calculate the expressions for the field from a solid conductor of an elliptical cross section with semi-axes $a$ and $b$ along the $X$ and $Y$ directions, as shown in Fig. 10. Detailed procedure to carry out the integrations in Eqs. (62) and (63) is described in Ref. [2] for this special case. The final result obtained is:

$$B_{in}(z) = \frac{\mu_0 J}{a+b} [b x - iay] ; \quad B_{out}(z) = \left( \frac{\mu_0 J}{2} \right) \left[ \frac{2ab}{z + \sqrt{z^2 - (a^2 - b^2)}} \right]$$

(64)

For the special case of $b = a$, Eq. (64) reduces to the results in Eqs. (58) and (59) for a solid cylindrical conductor.

8. GENERATING PURE DIPOLE AND QUADRUPOLE FIELDS USING OVERLAPPING CYLINDERS

Magnets that generate a pure dipole or a pure quadrupole field are of utmost importance in accelerators due to their use as bending and focussing elements. Using the result in Eq. (64) for a solid elliptical conductor, it can be shown that, at least in principle, a perfect dipole or a quadrupole field can be produced by overlapping two elliptical conductors carrying equal and opposite current densities, as shown in Figs. 11 and 12.

8.1 Generating a pure dipole field

To generate a pure normal-dipole field, the two elliptical cylinders are displaced with respect to each other by a distance $x_0$ along the $X$-axis, as shown in Fig. 11. To produce a skew-dipole field, the arrangement should be rotated by $90^\circ$. Since the cylinders carry equal and opposite current densities, the region of overlap effectively carries no current, and may be removed, thus producing the “aperture” of the magnet. Any point, $z_{in}$, inside this aperture is also inside both the ellipses. The complex field due to cylinder 1 (See Fig.11) is obtained by substituting $z = z_1 = z_{in} + x_0/2$ in Eq.(64). Similarly, the complex field due to cylinder 2 is obtained by substituting $z = z_2 = z_{in} - x_0/2$. The overall field is given by:

$$B_{in}(z_{in}) = \frac{\mu_0 J}{a+b} \left[ b \left( x + \frac{x_0}{2} \right) - iay - b \left( x - \frac{x_0}{2} \right) + iay \right] = \frac{\mu_0 J b x_0}{(a+b)} = B_y(z_{in}) + i B_x(z_{in})$$

(65)

Fig. 11 Using two overlapping elliptical cylinders to generate a pure normal-dipole field

Fig. 12 Generation of a pure normal-quadrupole field using two overlapping cylinders
It is seen from Eq. (65) that the y-component of the field is the same everywhere in the aperture. Also, the x-component is zero. This is characteristic of a pure, normal-dipole field [see Eq. (18) or (19)]. Similarly, a pure skew-dipole field can be generated by displacing the two ellipses along the y-axis instead of the x-axis. As the separation, \( x_a \), between the two ellipses increases, the aperture size reduces, the total current carried by the conductors increases, and the dipole field increases. It should be noted that a pure dipole field can also be obtained by two overlapping circles of radius \( a \), instead of ellipses. The magnitude of the field in this case is given by Eq. (65), with \( b = a \), to be simply \( (\mu J/2)x_a \). For a fixed separation and current density, larger circles use more total current and provide a larger aperture than smaller circles, but circles of all sizes give the same dipole field strength.

### 8.2 Generating a pure quadrupole field

A scheme for generating a pure normal-quadrupole field is shown in Fig. 12. Two cylinders of elliptical cross section carrying equal and opposite current densities are made to intersect at right angles to each other. The region of overlap carries no current, and can be treated as the aperture of the magnet. Any point inside this "aperture" is also inside both the cylinders. The total complex field at any point \( z_a \) is given by:

\[
B_{in}(z_{in}) = \frac{\mu J}{(a+b)}[-bx + iay + ax - iby] = \frac{\mu J(a-b)}{(a+b)}(x+iy) = \frac{\mu J(a-b)}{(a+b)}z_{in}
\]

(66)

This represents a pure normal-quadrupole field [see Eq. (18) or (19)]. The gradient is given by:

\[
G = \left( \frac{\partial B_y}{\partial x} \right) = \left( \frac{\partial B_x}{\partial y} \right) = \frac{\mu J(a-b)}{(a+b)} = \text{constant.}
\]

(67)

To obtain a skew quadrupole field, the arrangement of Fig. 12 should be rotated by 45 degrees. The gradient is proportional to the difference in the semi-major and semi-minor axes of the ellipse, as seen from Eq. (67). Physically, more oblong ellipses produce smaller apertures and use more total current, thus giving a larger gradient. Clearly, circular cylinders \((b = a)\) cannot be used to produce a quadrupole field in this fashion.

### 9. MULTPOLE EXPANSION OF FIELD DUE TO A CURRENT FILAMENT

In Secs. 7-8, we discussed the field produced by several simple current distributions. The expressions derived so far were for the total complex field. In practice, one is also interested in a harmonic description of the field. Once the total field is known for a given current distribution, one could obtain the various harmonic components simply by a series expansion in the form of Eqs. (18)-(19), or use the differential formulae in Eq. (24). A more convenient approach is to start again with the simplest current distribution given by an infinitely long, infinitesimally thin current filament located at an arbitrary location and express the field produced by this filament [see Eq. (54)] in terms of a multipole expansion. The multipole expansion for any arbitrary current distribution can then be obtained by integrating each harmonic term over the appropriate regions.

The complex field at any point, \( P \), due to a current filament located at \( z = a = a \exp(i\phi) \), as shown in Fig. 13, is given by Eq. (54). This expression is valid for all values of \( z \) and has a singularity at \( z = a \). Various multipole terms in this field can be obtained by a power series expansion. Because of the singularity, such an expansion has a radius of convergence given by \(|z| = a\). In any case, a single power series expansion with non-zero coefficients would tend to
diverge as \(|z| \to \infty\). As a result, we must separate the entire complex plane into two regions — an “inside” region extending to a circle of radius \(a\), and an “outside” region extending beyond the radius \(a\) to infinity, as shown in Fig. 13.

![Diagram showing inside and outside regions](image)

**Fig. 13** Calculation of multipole expansion of the field due to a current filament. A field point, \(P\), is shown in (a) the region \(r < a\), and (b) the region \(r > a\).

### 9.1 The “inside” region

The region occupied by the beam tube in a typical accelerator magnet is enclosed by all the current carrying conductors. The “inside” region, defined by \(r < a\), therefore, is the region of primary importance for accelerator magnets. A multipole expansion of the field in the form of Eq.(18)–(19) is possible in this region. For the case of an infinitely long, thin current filament, we have, from Eq.(54):

\[
B_{in}(z) = \text{Re} \left[ \frac{\mu_0 I}{2\pi a} \frac{1}{z-a} \left( 1 - \frac{r}{a} \exp\{i(\theta - \phi)\} \right) \right]^{-1}
\]

Using the binomial expansion \((1-\xi)^{-1} = \sum_{n=1}^{\infty} \xi^{n-1}\), and introducing a reference radius, we get,

\[
B_{in}(z) = -\text{Re} \left( \frac{\mu_0 I}{2\pi a} \sum_{n=1}^{\infty} \exp(-in\phi) \left( \frac{R_{ref}}{a} \right)^{n-1} \left( \frac{z}{R_{ref}} \right)^{n-1} \right)
\]

which is in the desired form of Eq.(9) or Eqs.(18)–(19). Comparing with Eq.(9), the amplitude and phase of the \(2n\)-pole term are given by:

\[
C(n) = \left( \frac{\mu_0 I}{2\pi a} \right)^{n-1} \frac{R_{ref}}{a} ; \quad \alpha_n = \phi + \frac{\pi}{n} \text{ for } I > 0 ; \quad \alpha_n = \phi \text{ for } I < 0
\]

where a positive value of current corresponds to current flowing along the positive Z-axis. It should be noted that the amplitude is always defined to be a positive quantity and any change in the field direction is absorbed in the phase angle. The **NORMAL** and **SKEW** components of the \(2n\)-pole field in the “US” and “European” notations are:
\[ B_{n-1}(\text{US}) = B_n(\text{European}) = \left( \frac{\mu_0 I}{2\pi a} \right) \left( \frac{R_{ref}}{a} \right)^{n-1} \cos(n\phi) \] 
\[ A_{n-1}(\text{US}) = A_n(\text{European}) = \left( \frac{\mu_0 I}{2\pi a} \right) \left( \frac{R_{ref}}{a} \right)^{n-1} \sin(n\phi) \] 
(71)

9.2 The "outside" region

For the "outside" region, we expand in a power series of \((a/z)\) instead of \((z/a)\) since \(|(a/z)| < 1\) in this case. Starting again with Eq.(54), we write:

\[ B_{out}(z) = \left( \frac{\mu_0 I}{2\pi} \right) \frac{1}{z-a} = \left( \frac{\mu_0 I}{2\pi r} \exp(i\theta) \right) \left[ 1 - \left( \frac{a}{r} \right) \exp\{i(\phi - \theta)\} \right]^{-1} \] 
(72)

Expressing the binomial expansion in the form \((1-\xi)^{-1} = 1 + \sum_{n=1}^{\infty} \xi^n\), we can write,

\[ B_{out}(z) = \left( \frac{\mu_0 I}{2\pi z} \right) \left[ 1 + \sum_{n=1}^{\infty} \left[ \cos(n\phi) + i\sin(n\phi) \right] \left( \frac{a}{z} \right)^n \right] \] 
(73)

This series is not in the usual form of the multipole expansion for the field inside a magnet aperture. However, it can be seen that this expansion gives a field which reduces with distance and converges for \(|z| \to \infty\) to \(B_{out}(z) \to [\mu_0 I/(2\pi z)]\), as expected.

It may seem puzzling at first as to why one is forced to use two different expansions starting from the same expression for the complex field, which is valid everywhere. It should be pointed out that the division of the complex plane into "inner" and "outer" regions is somewhat artificial, and depends solely on the location of the origin, the point around which the series expansion is based. A point that lies in the "outside" region can be brought into the "inside" region by choosing a different origin. Starting from the series expansion for the "inside" region, one can, in fact, obtain the field everywhere, including the points in the "outside" region, by a multistep process. For example, one could use the normal and skew terms at the origin, given by Eq.(71), to calculate the terms at a new origin near the periphery (but still within it) of the "inside" region, using the coordinate transformations described in Sec.4 [see Eqs.(31)–(32)]. This new series expansion will be valid within a circle of a new radius, centered at the new origin. Such a circle will encompass some regions of the complex plane that were earlier in the "outer" region. This is essentially the technique of analytic continuation.

9.3 Effect of an iron yoke

In accelerator magnets, an iron yoke is almost invariably used as part of the mechanical and magnetic design. In the design of practical magnets, one must use sophisticated numerical calculations to include the detailed geometry as well as non-linear magnetic properties of the iron. However, analytical results can be obtained under the assumption of a constant relative permeability, \(\mu_r\), and a circular aperture in an infinitely large yoke, as shown in Fig.14.
Fig. 14 A current filament inside a circular aperture in an infinite iron yoke

We consider an infinitely long current filament located at $a = a \exp(i\phi)$, inside a circular aperture of radius $R_{\text{yoke}}$, in the iron. The effect of the iron is obtained by applying the appropriate boundary conditions on the magnetic induction, $B$, and the magnetic field, $H$, at the inner surface of the yoke. For calculating the field inside the aperture of the yoke, this effect can be described by replacing the iron with an image current of magnitude $I'$ located at $a' = a \exp(i\phi')$ where

$$a' = R_{\text{yoke}}^2 / a; \quad I' = \left[\frac{\mu_r - 1}{\mu_r + 1}\right] I; \quad \phi' = \phi$$  \hspace{1cm} (74)

Using Eq.(70), the amplitude and phase of the $2n$-pole term are given by:

$$C(n) \exp(-in\alpha_n) = -\left(\frac{\mu_0 I}{2\pi a}\right)^{n-1} \left[ 1 + \left(\frac{\mu_r - 1}{\mu_r + 1}\right) \left(\frac{a}{R_{\text{yoke}}}\right)^{2n} \right] \exp(-in\phi)$$  \hspace{1cm} (75)

A comparison of Eqs.(75) and (70) shows that the presence of the iron yoke results in an increase in the field strength. Typically, for a superconducting dipole magnet, the contribution of the iron yoke to the total field is between 10–35%. This can result in substantial savings in the superconductor. A drawback of having a large contribution from the iron is the non-linear behaviour at higher fields, often leading to undesirable higher harmonics due to saturation. Such effects can be minimized to tolerable levels by a careful placement of holes and other features in the iron yoke [13, 14].

10. GENERATING A PURE $2m$-POLE FIELD

We had seen in Sec. 8 how simple-looking current distributions obtained by overlapping cylinders could be used to generate a pure dipole or a pure quadrupole field. Another approach towards generating a field of any given multipolarity can be arrived at by looking at the multipole expansion of the field from a thin current filament [see Sec. 9].

To generate a pure $2m$-pole field, the goal is to have all the terms in the expansion vanish, except for $n = m$. Let us assume that such a field is generated by using a thin, cylindrical current shell of radius $a$, similar to that shown in Fig.7, except that the current is not uniformly
distributed. Instead, the current is a function of the azimuthal angle, $\phi$. The complex field in the region inside the current shell is obtained by integrating Eq.(69):

$$B_{in}(z) = -\left(\frac{\mu_0}{2\pi a}\right)^{n-1} \sum_{n=1}^{\infty} \left(\frac{z}{a}\right)^{n-1} \exp(-in\phi)I(\phi)d\phi$$

(76)

In order to make all the terms except $n = m$ vanish, it is clear that the current distribution $I(\phi)$ must be orthogonal to the function $\exp(-in\phi)$ for all $n$, except $n = m$. Clearly, such functions are $\sin(n\phi)$ and $\cos(n\phi)$, since

$$\int_0^{2\pi} \cos(n\phi)\cos(m\phi)d\phi = \pi \delta_{mn}; \quad \int_0^{2\pi} \sin(n\phi)\cos(m\phi)d\phi = 0$$

(77)

Therefore, for a “cosine-theta” current distribution given by

$$I(\phi) = I_0 \cos(m\phi)$$

(78)

the complex field is given by

$$B_{in}(z) = -\left(\frac{\mu_0 I_0}{2a}\right) \left(\frac{z}{a}\right)^{m-1}$$

(79)

which represents a pure $2m$-pole (normal) field. Similarly, for a “sine-theta” current distribution given by $I(\phi) = I_0 \sin(m\phi)$, the field is a skew $2m$-pole field.

Unfortunately, neither the intersecting cylinders, nor the “cosine-theta” distributions can be accurately reproduced in practice. For example, the intersecting cylinders have sharp edges that must be rounded off in practice, leading to unacceptable harmonic distortions. In the design of conductor-dominated magnets (such as superconducting magnets), one has to approximate these distributions with several turns of a finite-sized conductor carrying a given current. This has led to the concept of “current blocks” approximation to the “cosine-theta” distribution. In this approach, a nearly pure multipole field is generated by employing one or more “current blocks”. Each of the blocks is formed by one or more turns of a conductor with a fixed cross section. The current is usually kept the same in each of the blocks for practical reasons. A description of this approach can be found in reference [15]. As an example, the coils for the arc dipole magnets for the Relativistic Heavy Ion Collider (RHIC), being built at the Brookhaven National Laboratory, consist of four current blocks with 9, 11, 8 and 4 turns [16]. Such a design attempts to zero out several of the lower order harmonics and also reduce the remaining higher order terms to negligible levels.

11. CURRENT SYMMETRIES AND ALLOWED HARMONICS

For a general current distribution, all harmonic terms in the multipole expansion of the field would be present. The current distributions in accelerator magnets have various symmetry properties depending on the type of the magnet. These symmetries imply that the normal and/or skew components of certain harmonics must vanish [17]. Such harmonic terms are referred to as the unallowed terms. The harmonics that can possibly be non-zero are called the allowed terms. It should be noted that a particular term may be allowed, but absent due to a careful
design of the magnet. On the other hand, a particular term may be \textit{unallowed}, but \textit{present} in a magnet due to violation of the relevant symmetry as a result of construction errors. In this section, we shall discuss the various symmetries and derive the corresponding allowed terms.

We shall assume that the current density, \( J \), is a function of the azimuthal angle only. This assumption is valid for the current blocks formed with finite sized conductors. Based on Eq.(69), the amplitude and phase of the \( 2n \)-pole term are given by:

\[
C(n) \exp(-in\alpha_n) \approx \int_0^{2\pi} J(\phi) e^{-in\phi} d\phi
\]  

(80)

The normal and the skew components are proportional to the real and the imaginary parts of the integral in Eq.(80). The fundamental symmetry in a \( 2m \)-pole magnet is a \( 2m \)-fold rotational antisymmetry, since the north and south poles of such a magnet are interchanged under a rotation by \( (\pi/m) \) radians. In addition, a magnet may have left-right, or top-bottom symmetry (or antisymmetry) if the \( X-Y \) axes are properly oriented.

11.1 \textbf{Allowed harmonics in a \( 2m \)-pole magnet}

It was shown in Sec. 10 that a pure \( 2m \)-pole field is produced by a \( \cos(m\theta) \) or a \( \sin(m\theta) \) distribution of current. Such a current distribution is antisymmetric under a rotation by \( (\pi/m) \) radians and is symmetric under a rotation by \( (2\pi/m) \) radians. Even when the \( \cos(m\theta) \) distribution is approximated by current blocks, this rotational symmetry is still preserved for a \( 2m \)-pole magnet. The current density at any azimuthal angle \( \phi + \pi/m \) is opposite that at \( \phi \), and so on, as shown in Fig.15. Using this fact, the integral in Eq.(80) can be written as a sum of \( 2m \) terms:

\[
C(n) \exp(in\alpha_n) \approx \int_0^{\pi/m} J(\phi) e^{in\phi} [1 - e^{in\pi/m} + e^{2in\pi/m} + \ldots + e^{i(2m-1)n\pi/m}] d\phi
\]  

(81)

or, \( C(n) \exp(in\alpha_n) \approx \left[ J(\phi) e^{in\phi} d\phi \right] \left[ \frac{1-e^{2in\pi}}{1+e^{2in\pi/m}} \right] \)  

(82)

Since \( n \) is always an integer, \( \exp(2in\pi) = 1 \) and the right hand side in the above integral vanishes, unless the denominator also vanishes. This requires that \( n \) must be an odd multiple of \( m \). Thus, for a dipole magnet (\( m = 1 \)), only terms with \( n = 1, 3, 5, \ldots \) are allowed. For a quadrupole magnet (\( m = 2 \)), terms with \( n = 2, 6, 10, \ldots \) are allowed, and so on.

11.2 \textbf{“Top-bottom” symmetry or anti-symmetry}

A “top-bottom” symmetry about the \( X \)-axis implies \( J(2\pi - \phi) = J(\phi) \), as shown in Fig. 16. In this case, Eq. (80) can be written as:

\[
C(n) \exp(in\alpha_n) \approx \int_0^{2\pi} J(\phi) e^{in\phi} d\phi = \int_0^{\pi} J(\phi) \left[ e^{in\phi} + e^{i(2\pi - \phi)} \right] d\phi = \int_0^{\pi} J(\phi) \cos(n\phi) d\phi
\]  

(83)

The result of integration has no imaginary part in this case. This implies that all the skew terms vanish and only the normal terms are allowed as a result of the “top-bottom” symmetry in the current distribution. It is obvious that the current distribution in all normal magnets, such as a normal dipole, a normal quadrupole, etc. must have top-bottom symmetry.
Fig. 15 Rotational symmetry of current density in a 2m-pole magnet

The case of the "top-bottom" antisymmetry is similar, except that the current density satisfies $J(2\pi - \phi) = -J(\phi)$. The integral in Eq.(80) becomes,

$$C(n) \exp(in\alpha_n) \propto \int_{0}^{2\pi} J(\phi) e^{i n\phi} d\phi = \int_{0}^{\pi} J(\phi) [e^{i n\phi} - e^{i n(2\pi - \phi)}] d\phi = i \int_{0}^{\pi} J(\phi) \sin(n\phi) d\phi \quad (84)$$

The result of integration has no real part in this case. This implies that all the normal terms vanish and only the skew terms are allowed as a result of the "top-bottom" antisymmetry in the current distribution. All skew magnets, such as a skew dipole, a skew quadrupole, etc. have a "top-bottom" antisymmetry.

11.3 "Left-right" symmetry or anti-symmetry

A "left-right" symmetry about the $Y$-axis implies $J(\pi - \phi) = J(\phi)$, as shown in Fig. 17. In this case, Eq. (80) can be written as:

$$C(n) \exp(in\alpha_n) \propto \int_{-\pi/2}^{\pi/2} J(\phi) [e^{i n\phi} + e^{i n(\pi - \phi)}] d\phi = \int_{-\pi/2}^{\pi/2} J(\phi) [e^{i n\phi} + (-1)^n e^{-i n\phi}] d\phi \quad (85)$$

$$\begin{align*}
C(n) \exp(in\alpha_n) &\propto i \int_{-\pi/2}^{\pi/2} J(\phi) \sin(n\phi) d\phi \quad \text{for ODD } n \\
&\quad \text{Left-Right Symmetry} \\
C(n) \exp(in\alpha_n) &\propto \int_{-\pi/2}^{\pi/2} J(\phi) \cos(n\phi) d\phi \quad \text{for EVEN } n
\end{align*} \quad (86)$$

The result of integration has no real part for odd multipoles and has no imaginary part for even multipoles. This implies that all the odd normal terms (such as the normal dipole, the normal sextupole, etc.) and all the even skew terms (such as the skew quadrupole, the skew octupole, etc.) vanish as a result of the "left-right" symmetry in the current distribution. All the normal
magnets of even order, such as a normal quadrupole, and all the skew magnets of odd order, such as the skew dipole, have this type of current symmetry.

Similarly, a "left-right" antisymmetry implies \( J(\pi - \phi) = -J(\phi) \). In this case, Eq.(80) can be written as:

\[
C(n)\exp(in\alpha_n) \propto \int_{-\pi/2}^{\pi/2} J(\phi) e^{in\phi} = \int_{-\pi/2}^{\pi/2} J(\phi) e^{i(n-1)\phi} d\phi \quad (87)
\]

\[
C(n)\exp(in\alpha_n) \propto i \int_{-\pi/2}^{\pi/2} J(\phi) \sin(n\phi) d\phi \quad \text{for EVEN } n
\]

\[
C(n)\exp(in\alpha_n) \propto \int_{-\pi/2}^{\pi/2} J(\phi) \cos(n\phi) d\phi \quad \text{for ODD } n
\]

Left-Right Antisymmetry \quad (88)

The result of integration has no real part for even multipoles and has no imaginary part for odd multipoles. This implies that all the even normal terms (such as the normal quadrupole, the normal octupole, etc.) and all the odd skew terms (such as the skew dipole, the skew sextupole, etc.) vanish as a result of the "left-right" antisymmetry in the current distribution. All the normal magnets of odd order, such as a normal dipole, and all the skew magnets of even order, such as the skew quadrupole, have this type of current symmetry.

The unallowed harmonics under various symmetries are summarized in Table 1. The entries for various normal and skew terms in this table follow the "European notation". To obtain a table in the US notation, the indices should be reduced by one. It should be noted that these symmetries refer to the \textit{current distribution} in the magnet, and not to the symmetry in the shapes of the pole-tips. One can obtain similar results for symmetries in the shapes of the pole-tips also [10].
Table 1
Allowed harmonics under various symmetries in the current distribution

<table>
<thead>
<tr>
<th>Type of Symmetry in the Current Distribution</th>
<th>Example</th>
<th>Normal Terms*</th>
<th>Skew Terms*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left-right symmetry</td>
<td>Normal quadrupole</td>
<td>$B_{2k+1} = 0$</td>
<td>$A_{2k} = 0$</td>
</tr>
<tr>
<td>Left-right anti-symmetry</td>
<td>Normal dipole</td>
<td>$B_{2k} = 0$</td>
<td>$A_{2k+1} = 0$</td>
</tr>
<tr>
<td>Top-bottom symmetry</td>
<td>Normal dipole</td>
<td>---</td>
<td>$A_{1} = 0$</td>
</tr>
<tr>
<td>Top-bottom anti-symmetry</td>
<td>Skew dipole</td>
<td>$B_{k} = 0$</td>
<td>---</td>
</tr>
</tbody>
</table>

[* The "European Notation" is used in this table to denote the normal and skew terms.]

ACKNOWLEDGEMENTS

The author would like to thank Peter Wanderer for a critical review of the manuscript and for many useful suggestions. Many valuable discussions with J. Herrera are also appreciated. This work was supported by the U.S. Department of Energy under contract no. DE-AC02-76CH00016.

REFERENCES


THE MOTIVATION FOR MAGNET MEASUREMENTS

E.J.N. Wilson
CERN, Geneva, Switzerland

Abstract
Construction of magnets to precise tolerances and the subsequent checking of their manufacture by even more precise measurements is the key to the construction of reliable synchrotrons which outperform their specification. It is important to have a working knowledge of the transverse beam dynamics determining these tolerances in order to intelligently analyse measurements and design more efficient procedures for assembly and measurement.

1. INTRODUCTION

The first step in designing an accelerator or storage ring is to choose an optimum pattern of focusing and bending magnets, the lattice. At this stage, non-linearities in the guide field are ignored. It is assumed that the bending magnets are identical and have a pure dipole field. Gradient magnets or quadrupoles have radial field shapes which have a constant slope, unperturbed by higher-order multipole terms.

Quadrupole magnets provide the restoring fields and are usually embedded in the lattice of bending magnets in an alternating pattern. Half focus the beam, while the other half defocus the beam. The lattice has a strong influence on the aperture of these magnets which are usually the most expensive single system in the accelerator. We see in Fig. 1 the pattern of one cell of the CERN SPS which is repeated 108 times around the circumference. Focusing and defocusing quadrupoles are labelled F and D and for obvious reasons this focusing structure is called FODO.

![Diagram of FODO lattice](image1)

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

![Diagram of particle trajectories](image2)

Fig. 2 The paths of particles within a FODO lattice are within the envelope of betatron motion are always closer in the D quadrupoles so they receive a net focusing effect. The phase space ellipse is tall and narrow at the D lens where the beam has a large divergence spread.
Particles make betatron oscillations within the envelopes indicated in Fig. 2. The envelope of these oscillations follows the function $\beta(s)$ which has waists near each defocusing magnet and has a maximum at the centres of F quadrupoles. Since F quadrupoles in the horizontal plane are D quadrupoles vertically, and vice versa, the two functions $\beta_x(s)$ and $\beta_y(s)$ are one half-cell out of register in the two transverse planes. For formal reasons $\beta$ has the dimensions of length but the units bear no relation at this stage to physical beam size.

Imperfections in the field of bending magnets can cause vertical and horizontal distortions to the central closed orbit of the machine – the horizontal axis of Fig. 2 – and errors in the gradient of the quadrupoles can cause distortion of the shape of the envelope of the betatron oscillations shown in Fig. 1. Both these effects can dominate the aperture required in the magnets unless magnet field errors are kept strictly under control in the manufacturing process. Hence, before going too far in fixing parameters, the practical difficulties in designing the magnets must be considered and the tolerances which can be reasonably written into the engineering specification determined. Estimates must be made of the non-linear departures from pure dipole or gradient field shape, and of the statistical fluctuation of these errors around the ring at each field level.

We must take into consideration that the remanent field of a magnet may have quite a different shape from that defined by the pole geometry; that steel properties may vary during the production of laminations; that eddy currents in vacuum chamber and coils may perturb the linear field shape. Mechanical tolerances must ensure that asymmetries do not creep in. At high field the linearity may deteriorate owing to saturation and variations in packing factor can become important. Superconducting magnets will have strong error fields due to persistent currents in their coils.

When these effects have been reviewed, tolerances and assembly errors may have to be revised and measures taken to mix or match batches of laminations with different steel properties or coils made from different batches of superconductor. It may be necessary to place magnets in a particular order in the ring in the light of production measurements of field uniformity or to shim some magnets at the edge of the statistical distribution. Even when all these precautions have been taken, non-linear errors may remain whose effect it is simpler to compensate with auxiliary multipole magnets.

All these effects tend to become more serious in rings of larger radius and in machines with superconducting magnets the conductors whose position defines the field shape must be located with very tight tolerances.

2. MAGNETIC RIGIDITY

When calculating the effect of errors it is conventional to consider the small angular deflection produced by a magnet of length, $\ell$, and strength, $\Delta B$.

$$\theta = \frac{\ell \Delta B}{(B \rho)}$$

where the denominator is the “magnetic rigidity” of the particle and related to its relativistic momentum, $p$, and the particles charge, $e$, by

$$(B \rho) = \frac{p}{e}.$$ 

The common convention in charged particle dynamics is to quote $p$ in units of GeV and $B \rho$ in Tesla.meters. Whereupon:

$$(B \rho)[T.m] = 3.3356p[GeV]$$
Figure 3 shows the trajectory of a particle in a bending magnet or dipole of length \( \ell \). Usually the magnet is placed symmetrically about the arc which is the particle's path. One may see immediately that:

\[
\sin(\theta/2) = \frac{\ell}{2\rho} = \frac{\ell B}{2(B\rho)} ,
\]

and if \( \theta \ll \pi/2 \)

\[
\theta = \frac{\ell B}{(B\rho)} .
\]

The ends of bending magnets are often parallel but in some machines are designed to be normal to the beam. There is a focusing effect at the end which depends on the angle of these faces.

3. FOCUSING

The principal focusing elements in a modern synchrotron are quadrupole magnets. The poles are truncated rectangular hyperbolae and alternate in polarity. The field shape (Fig. 4) is such that it is zero on the axis of the device but its strength rises linearly with distance from the axis. This can be seen from a superficial examination of Fig. 4, remembering the product of field and length of a field line joining the poles is a constant. Symmetry tells us that the field is vertical in the median plane (and purely horizontal in the vertical plane of asymmetry). The field must be downwards on the left of the axis if it is upwards on the right.

![Fig. 4 Components of field and force in a magnetic quadrupole. Positive ions approach the reader on paths parallel to the y axis.](image)

This last observation ensures that the horizontal focusing force, \( evB_z \), has an inward direction on both sides and, like the restoring force of a spring, rises linearly with displacement, \( x \). The strength of the quadrupole is characterised by its gradient \( dB_z/dx \) normalised with respect to magnetic rigidity:

\[
k = \frac{1}{(B\rho)} \frac{dB_z}{dx} .
\]

The angular deflection given to a particle passing through a short quadrupole of length \( Q \) and strength \( k \) at a displacement \( x \) is therefore:

\[
\Delta \chi = \ell kx .
\]
Compare this with a converging lens:

$$\Delta x' = -x / f$$

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

$$f = -1/(k\ell) .$$

Note that conventionally $k$ is negative for a horizontally focusing quadrupole.

The particular quadrupole shown in Fig. 4 would focus positive particles coming out of the paper or negative particles going into the paper in the horizontal plane. A closer examination reveals that such a quadrupole is defocusing in the vertical plane and deflects particles with a vertical displacement away from the axis – vertical displacements are defocused. You can immediately see this if you turn the figure $90^\circ$. In spite of this seemingly damning characteristic, a FODO pattern of alternating polarity quadrupoles has a net focusing effect in both planes.

4. THE GUTTER ANALOGY

In order to understand focusing we ignore vertical defocusing for a moment and consider an infinitely long quadrupole. A particle oscillates within it exactly like a small sphere rolling down a slightly inclined gutter with constant speed. Figure 5 shows two views of this motion and from the right hand view we recognise the motion as a sine wave: note too that the sphere makes four complete oscillations along the gutter. In accelerator terms its motion has a wave number $Q = 4$. In a ring $Q$ is the number of oscillations per turn.

Now extend this analogy by bending the gutter into a circle rather like the brim of a hat. Suppose we provide the necessary instrumentation to measure the displacement of the sphere from the centre of the gutter each time it passes a given mark on the brim. We also suppose we have a means to measure its transverse velocity which, with the aid of a computer, we might convert into its divergence angle defined in Fig. 6:

$$x' = \frac{dx}{dt} = \frac{v_{\perp}}{v_{\parallel}}$$

Suppose we make the hat out of a slightly different length of gutter than shown so that $Q$ is not an integer. If we plot a diagram of $x'$ against $x$ we can plot a point for each arrival of the sphere. This is called a phase space diagram. The sphere may have a large transverse velocity as it crosses the axis of the gutter or it might have almost zero transverse velocity as it reaches its maximum displacement.

The locus will be an ellipse (Fig. 6) and the phase will advance by $Q$ revolutions each time the particle returns.
Fig. 6 The elliptical locus of a particle's history in phase space as it circulates in a synchrotron

The time has come to define some of the beam dynamical quantities more rigorously. The area of the ellipse is a measure of how much the particle departs from the ideal trajectory represented by the origin:

\[
\text{Area} = \pi \varepsilon \text{ [mm.mrad]}.
\]

The maximum amplitude, or major axis, of the ellipse is defined:

\[
\hat{x} = \sqrt{\pi \beta},
\]

so that to satisfy Eq. (15)

\[
\hat{x} = \sqrt{\pi \beta}.
\]

The quantity \( \beta \) is a property of the focusing system, not the beam, it varies around the ring and is the same function we have plotted in Fig. 1. In an alternating gradient focusing system such as Fig. 1, the brim of the hat will vary its width and curvature around the crown and \( \beta \) will follow this variation in some way. Note that the aspect ratio of the ellipse is just \( \beta \).

5. TRANSVERSE EQUATION OF MOTION

In Section 3 we derived an expression for the change in divergence of a particle passing through the quadrupole. The angular deflection given to a particle passing through a short quadrupole of length \( ds \) and strength \( k \) at a displacement \( x \):

\[
dx' = k x ds
\]

We can immediately rearrange this to form the differential equation for the motion

\[
x'' - k(s)x = 0
\]

This is the famous Hill's Equation, a second-order linear equation with periodic coefficient, \( k(s) \) which is the distribution of focusing strength around the ring. In the horizontal plane we must strictly include an extra focusing term for the curvature which can be significant in small rings:

\[
x'' + \frac{1}{\rho(s)^2} - k(s)x = 0
\]
In the vertical plane the equation has the opposite sign of \( k \)
\[ z'' + k(s)z = 0 \]

The equation is reminiscent of simple harmonic motion but with a restoring constant \( k(s) \) which varies around the accelerator. In order to arrive at a solution we first assume that \( k(s) \) is periodic on the scale of one turn of the ring or some smaller unit, a cell or period, from which the ring is built. The solution of Hill's Equation is not unlike simple harmonic motion:
\[ x = \sqrt{\beta(s)c} \sin[\phi(s) + \phi_0] \]

In simple harmonic motion the amplitude is a constant but now we see that in addition to \( \sqrt{c} \), a property of the beam, there is another amplitude component, a function, \( \sqrt{\beta(s)} \) which is defined by the lattice pattern. Another difference with harmonic motion is that phase does not advance linearly with time and with distance \( s \) around the ring but is a seemingly arbitrary function of \( s \). It will perhaps not be difficult to believe that these functions of \( s \) must have the same periodicity as the lattice. (All this is to be seen in Fig. 1.)

6. CIRCLE DIAGRAM

When we plot the locus of the motion of a particle obeying Hill's equation we obtain ellipses in phase space similar to Fig. 6. The aspect ratio of the ellipse depends on the position of the observer in the ring. Close to an F quadrupole where \( \beta(s) \) is large the horizontal he ellipse will be very wide and not very high in divergence angle and it is in such positions that a small angular kick due to an imperfection has the greatest effect on the beam. Imagine, for example, how little angular displacement is needed to move the ellipse by its own height and increase the beam dimensions by a factor 2. The effect of errors is most serious near F quadrupoles.

![Fig. 7 Phase-space diagram at \( \beta \) minimum and \( \beta \) maximum](image)

So predominant is the effect of perturbations near \( \hat{\beta} \) positions that you can often do quite good "back of the envelope" calculations by closing your eyes to what happens to the protons in between F quadrupoles. At F quadrupoles the ellipse always looks the same, i.e., upright, with semi-axes in displacement and divergence
\[ \sqrt{(\beta c)}, \sqrt{(e/\beta)} \]

This can be reduced to a circle radius by using the new coordinates
\[ y = y \]
\[ p = \beta y' \]
If the machine has 108 periods and a $Q$ of 27.6, the proton advances in phase by $2\pi Q/108$ from one period to the next; this is just the angle subtended at the centre of the circle multiplied by $Q$. After one turn of the machine, it has made 27 revolutions of the circle plus an angle of $2\pi$ multiplied by the fractional part of $Q$, see Fig. 8.

7. CLOSED ORBIT DISTORTION

In designing a synchrotron, the bending field is matched to some ideal synchronous momentum, $p_0$. A particle of this momentum and of zero emittance will pass down the centre of each quadrupole, be bent by exactly $2\pi$ by the bending magnets in one turn of the ring and remain synchronous with the r.f. frequency. Its path is called the central (or synchronous) momentum closed orbit, (Fig. 2). This orbit closes on itself so that $x$ and $x'$ remain zero. Imperfections in the guide field can distort this orbit as in Fig. 9 where a dipole error is progressively increased.

![Dipole](image)

Fig. 9 Closed orbit distortion as a dipole is slowly switched on

Even the best synchrotron magnets cannot be made absolutely identical. Each magnet differs from the mean by some small error in integrated strength:

$$\delta(Bt) = \int B\, dt - \left( \int B\, dt \right)_{\text{ideal}} .$$

These and other machine imperfections, such as survey errors which can be expressed as equivalent to field errors, are randomly spread around the ring.

One of the most important considerations in designing a machine is to keep this closed orbit distortion to a minimum because it eats up available machine aperture. Also, once we have succeeded in getting a few turns round the machine, we want to reduce this distortion with correcting dipole magnets whose strength must be estimated before construction. As a first step let us consider the effect on the orbit of such a correcting dipole located at a position where $\beta = \beta_K$ and observed at another position.

We turn on a short dipole (we shall assume it is a delta function in $s$) which makes a constant angular kick in divergence

$$\delta y' = \delta(Bt) / (Bp) ,$$
8. CLOSED ORBIT IN THE CIRCLE DIAGRAM

In the unperturbed part of the lattice the distorted closed orbit must of necessity follow the same equations of motion as a particle. Its locus in phase space will therefore be an ellipse. To illustrate the physics of closed orbit distortion let us examine this ellipse only at all points of equal \( \beta(s) \), for example at the middle of the F quadrupoles. By multiplying the divergence by \( \beta(s) \) the locus becomes a circle. Consider the special case where dipole error and observation point are at the same value of \( \beta(s) \). We see quite from Fig. 10 how the equation for the amplitude of the distortion appears. The “kick” due to the dipole just completes the circular trajectory after a non-integer number of turns.

9. SOURCES OF DISTORTION

<table>
<thead>
<tr>
<th>Type of element</th>
<th>Source of kick</th>
<th>r.m.s. value</th>
<th>( \langle \Delta B \ell/(B \rho) \rangle )\text{rms}</th>
<th>Plane</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gradient magnet</td>
<td>Displacement</td>
<td>( &lt;\Delta y&gt; )</td>
<td>( k_d &lt;\Delta y&gt; )</td>
<td>( x,z )</td>
</tr>
<tr>
<td>Bending magnet (bending angle = ( \theta_l ))</td>
<td>Tilt</td>
<td>( &lt;\Delta&gt; )</td>
<td>( \theta_l &lt;\Delta&gt; )</td>
<td>( z )</td>
</tr>
<tr>
<td>Bending magnet (length = ( d_l ))</td>
<td>Field error</td>
<td>( &lt;\Delta B/B&gt; )</td>
<td>( \theta_l &lt;\Delta B/B&gt; )</td>
<td>( x )</td>
</tr>
<tr>
<td>Straight sections (length = ( d_l ))</td>
<td>Stray field</td>
<td>( &lt;\Delta B_s&gt; )</td>
<td>( d_l \langle \Delta B_s/(B \rho) \rangle_{inj} )</td>
<td>( x,z )</td>
</tr>
</tbody>
</table>

The principal imperfections in a synchrotron causing orbit distortion are shown in Table 1. These include errors in magnetic field and in the alignment of the ring. The first line in the table represents the random variations in the position of quadrupole magnets with respect to their ideal location. A small displacement of a quadrupole gives an effective dipole perturbation, \( k_l \Delta y \). The tilt of bending magnets causes a small resultant dipole in the horizontal direction which deflects vertically. Obviously there may also be random errors in magnet gap, length or in the coercivity of the steel yoke which determines remanent field contributing to the third line. Both remanent and stray fields in straight sections tend to be constant and their effect scales as \( 1/B \) as the machine pulses. Their effect should therefore be evaluated where it is worst, at injection. In a modern superconducting machine the persistent current fields play the role of remanent effects.

10. UNCORRELATED ERRORS

In estimating the effect of a random distribution of dipole errors we must take the r.m.s. average, weighted according to the \( \beta_i \) values over all of the kicks \( \delta y_i \) from the \( N \) magnets in
the ring. If we observe the distortion at $\beta(s)$ and its source is at $\beta$, the effect is scaled as $\sqrt{\beta(s)/\beta}$, and the expectation value of the amplitude is:

$$\langle y(s) \rangle = \frac{\sqrt{\beta(s)}}{2\sqrt{2} \sin \pi Q} \sqrt{\sum_i \beta_i \delta y_i^2}$$

$$= \frac{\sqrt{\beta(s)\beta}}{2\sqrt{2} \sin \pi Q} \sqrt{N} \frac{(\Delta B \rho)_{\text{rms}}}{\rho}.$$

The factor $\sqrt{2}$ comes from averaging over all the phases of distortion produced. In designing a machine it used to be conventional wisdom to make sure that the vacuum chamber will accommodate twice this expectation value. The probability of no particles making the first turn is thus reduced to a mere 2%. More modern designs rely on closed-orbit steering to thread the first turn and thereafter assume that orbit correction to a millimetre or so will be feasible.

11. MAGNIFICATION OF ERRORS

A more rigorous renormalisation of phase space which does not imply any approximation but which simplifies the problem is the $(\eta, \psi)$ transformation to convert Hill's equation into that of a harmonic oscillator:

$$\frac{d^2 \eta}{d \psi^2} + Q^2 \eta = g(\psi)$$

where $g(\psi)$ is the azimuthal pattern of some perturbation of the guide field related to

$$F(s) = \frac{\Delta B(s)}{\rho}.$$

In the ideal case $g(\psi)$ is everywhere zero.

I will not bother you with how this transformation is found, but just state it. The new coordinates are related to the old:

$$\eta = \beta^{-1/2} y$$

$$\psi = \int \frac{ds}{QB}, \quad g(\psi) = Q^2 \beta^{3/2} F(s),$$

where $\psi$ advances by $2\pi$ every revolution. It coincides with $\theta$ at each location and does not depart very much from $\theta$ in between.

One of the advantages of reducing the problem to that of a harmonic oscillator in $(\eta, \psi)$ coordinates is that perturbations can be treated as the driving term of the oscillator. They may be broken down into their Fourier components, and the whole problem solved like the forced oscillations of a pendulum. The driving term is put on the right hand side of Hill's equation:

$$\frac{d^2 \eta}{d \psi^2} + Q^2 \eta = Q^2 \sum_{n=1}^{\infty} f_n e^{i n \psi} = Q^2 \beta^{3/2} F(s),$$

where $F(s)$ is the azimuthal pattern of the perturbation $\Delta B/(B\rho)$; and $Q^2 \beta^{3/2}$ comes from the transformation from physical coordinates to $(\eta, \psi)$.

The Fourier amplitudes are defined:
\[ f(\psi) = \beta^{3/2} F(s) = \sum_k f_k e^{ik\psi}, \]

where

\[ f_k = \frac{1}{2\pi} \int_0^{2\pi} f(\psi) e^{-ik\psi} d\psi = \frac{1}{2\pi\beta^{1/2}} \int f(\psi) e^{-ik\psi} ds. \]

We can then solve Hill's equation as

\[ \eta = \sum_{n=1}^{\infty} \frac{Q^2 f_k}{Q^2 - k^2} e^{ik\psi} \text{ (or its real part).} \]

In fact this solution is a closed orbit, a particular solution of Hill's differential equation, to which we must add the general solutions which describe betatron oscillations about this orbit.

The function \( Q^2/(Q^2 - k^2) \) is sometimes called the magnification factor for a particular Fourier component of \( \Delta B \). It rises steeply when the wave number \( k \) is close to \( Q \), and the effect of the two Fourier components in the random error pattern with \( k \) values adjacent to \( Q \) accounts for about 60% of the total distortion due to all random errors. Figure 11 shows a closed orbit pattern from electrostatic pick-ups in the FNAL ring, whose \( Q \) is between 19 and 20. The pattern shows strong components with these wave numbers.

![Fig. 11 FNAL main ring electrostatic pick-ups show closed orbit around the ring \((Q = 19.2)\)](image)

12. MULTIPOLE FIELD EXPANSION

Before we come to discuss the non-linear terms in the dynamics, we shall need to describe the field errors which drive them. The magnetic vector potential of a magnet with \( 2n \) poles in Cartesian coordinates is:

\[ A = \sum_n A_n f_n(x,z), \]

where \( f_n \) is a homogeneous function in \( x \) and \( z \) of order \( n \).

### Table 2

<table>
<thead>
<tr>
<th>Multipole</th>
<th>( n )</th>
<th>Regular ( f_n )</th>
<th>Skew ( f_n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quadrupole</td>
<td>2</td>
<td>( x^2 - z^2 )</td>
<td>2xz</td>
</tr>
<tr>
<td>Sextupole</td>
<td>3</td>
<td>( x^3 - 3xz^2 )</td>
<td>3x^2z - z^3</td>
</tr>
<tr>
<td>Octupole</td>
<td>4</td>
<td>( x^4 - 6x^2z^2 + z^4 )</td>
<td>4x^3z - 4xz^3</td>
</tr>
<tr>
<td>Decapole</td>
<td>5</td>
<td>( x^5 - 10x^3z^2 + 5xz^4 )</td>
<td>5x^4z - 10x^2z^3 + z^5</td>
</tr>
</tbody>
</table>
Table 2 gives $f_n(x, z)$ for low-order multipoles, both regular and skew. Figure 12 shows the distinction. We can obtain the function for other multipoles from the binomial expansion of

$$f_n(x, z) = (x + iz)^n.$$  

The real terms correspond to regular multipoles, the imaginary ones to skew multipoles. The difference between regular and skew is illustrated in Fig. 12.

![Fig. 12 Pole configurations for a regular sextupole and a skew sextupole](image)

For numerical calculations it is useful to relate $A_n$ and the corresponding derivative of field and compare the resulting series with a Taylor expansion of the field expressed as a series of derivatives. For regular magnets:

$$B_z(z = 0) = \frac{\partial A_z}{\partial x} = \sum_{n=1}^{\infty} nA_n x^{(n-1)} = \sum_{n=1}^{\infty} \frac{1}{(n-1)!} \left( \frac{d^{(n-1)}B_z}{dx^{(n-1)}} \right)_0 x^{n-1}$$

so that by inspection

$$A_n = \frac{1}{n!} \left( \frac{d^{(n-1)}B_z}{dx^{(n-1)}} \right)_0.$$

A more modern convention (in Europe) is to speak of multipole coefficients, $b_n$, for normal components and $a_n$ for skew components, where $R_z$ is some reference radius (10 mm for the LHC), $B_z$ is the magnitude of the nominal dipole field, $B_y$, and $Z = x + iy$.

$$B_y + i B_z = B_z \sum_n (b_n + i a_n) (Z/R_z)^{n-1}.$$

The suffix, $n = 1$ for the dipole, 2 for the quadrupole and 3 for sextupole etc. Note that US notation may start with $n = 0$ for the dipole and different laboratories use other reference radius. In spite of the possibilities for confusion this has the advantage that the coefficients are a measure of the tolerated fraction of field error inside the reference circle, where beam is supposed to be stable.

As a practical example of how one may identify the multipole components of a magnet by inspecting its symmetry, we digress a little to discuss the sextupole errors in the main dipoles of a large synchrotron.

Let us look at a simple dipole (Fig. 13). It is symmetric about the vertical axis and its field distribution will contain mainly even exponents of $x$, corresponding to odd $n$ values: dipole, sextupole, decapole, etc. We can see, too, that cutting off the poles at a finite width can produce a virtual sextupole. Moreover, the remanent field pattern is frozen in at high field where the flux lines leading to the pole edges are shorter than those leading to the centre. The remanent magneto-motive force $\int H_c d\ell$ is weaker at the pole edges, and the field tends to sag into a parabolic or sextupole configuration. This too produces a sextupole.
Fig. 13 The field in a simple dipole. The $\delta N$ and $\delta S$ poles superimposed on the magnet poles give the effect of cutting off the poles to a finite width.

These three sources of sextupole error are the principle non-linearities in a large machine like the SPS. Note that these sextupole fields have no skew component. Similar sextupole fields which are even stronger and which vary with time are a principle error component in the injection field shape of a modern superconducting ring such as LHC. These are caused by persistent circulating currents set up in the individual turns of the superconducting coils rather as eddy currents are set up during magnet pulsing in the vacuum chamber of a warm machine. In the LHC dipoles the field shape is determined by the positional tolerances on the position of the windings which are very difficult tolerances to achieve. However, before launching into non-linearities let us examine a simple linear resonance.

13. WORKING DIAGRAM

Apart from the obvious need to minimise closed orbit distortion, measures must be taken to reduce the influence of non-linear resonances on the beam. A glance at the working diagram (Fig. 14) shows why this is so. The $Q_H, Q_V$ plot is traversed by a mesh of non-linear resonance lines or stopbands of first, second, third, and fourth order. The order, $n$, determines the spacing in the $Q$ diagram; third-order stopbands, for instance, converge on a point which occurs at every $1/3$ integer $Q$-value (including the integer itself). The order, $n$, is related to the order of the multipole which drives the resonance. For example, fourth-order resonances are driven by multipoles with $2n$ poles, i.e. octupoles. Multipoles can drive resonances of lower-order; octupoles drive fourth- and second-order; sextupoles third- and first-order, etc., but here we simply consider the highest order driven.

The non-linear resonances are those of third-order and above, driven by non-linear multipoles. Their strength is amplitude-dependent so that they become more important as we seek to use more and more of the machine aperture. Theory used to discount resonances of fifth- and higher-order as harmless (self-stabilised), but experience in the ISR, FNAL and SPS suggests this is not to be relied upon when we want beams to be stored for more than a second or so.

Each resonance line is driven by a particular pattern of multipole field error which can be present in the guide field. The lines have a finite width depending directly on the strength of the error. In the case of those driven by non-linear fields, the width increases as we seek to exploit a larger fraction of the magnet aperture. We must ensure that the errors are small enough to leave some clear space between the stopbands to tune the machine, otherwise particles will fall within the stopbands and be rapidly ejected before they have even been accelerated. In general, the line width is influenced by the random fluctuations in multipole error around the ring rather than the mean multipole strength.

Systematic or average non-linear field errors also make life difficult. They cause $Q$ to be different for the different particles in the beam depending on their betatron amplitude or momentum defect. Such a $Q$-spread implies that the beam will need a large resonance-free window in the $Q$ diagram.
Fig. 14 Working diagram or $Q_H, Q_V$ plot showing the non-linear resonances in the operating region of the CERN SPS

14. PHASE SPACE TRAJECTORY FOR 1/3 RESONANCE

The third-integer stopbands are driven by sextupole field errors and are therefore non-linear. First take the phase space of Fig. 8 and imagine the perturbation is a single short sextupole of length $\ell$, near a horizontal maximum beta location. Its field is

$$\Delta B = \frac{d^2 B}{dx^2} x^2 = \frac{B''}{2} x^2,$$

and it kicks a particle with betatron phase $Q \theta$ by

$$\Delta p = \frac{\beta \ell B''}{2B \rho} x^2 = \frac{\beta \ell B'' a^2}{2B \rho} \cos^2 Q \theta$$

inducing increments in phase and amplitude,

$$\frac{\Delta a}{a} = \frac{\Delta p}{a} \sin Q \theta = \frac{\beta \ell B'' a}{2B \rho} \cos^2 Q \theta \sin Q \theta$$

$$\Delta \phi = \frac{\Delta p}{a} \cos Q \theta = \frac{\beta \ell B'' a}{2B \rho} \cos^3 Q \theta$$

$$= \frac{\beta \ell B'' a}{8B \rho} (\cos 3Q \theta + 3 \cos Q \theta).$$
Suppose $Q$ is close to a third integer, then the second term in the above equation averages to zero over three turns and we are left with a phase shift:

$$2\pi\Delta Q = \Delta \phi = \frac{\beta \ell \mathcal{B}''(s)}{8Bp} \cos 3Q\theta.$$ 

Suppose that Fourier analysis of $\beta(s)B''(s)$ results in a term with an azimuthal dependence $\cos p\theta$. Together with $3Q\theta$ in the above equation this produces a slowly varying term, $\cos(3Q - p)\theta$. Hence, close to $Q = p/3$, where $p$ is an integer, $\cos(3Q - p)\theta$ varies slowly, wandering within a band about the unperturbed $Q_0$ within the limits

$$Q_0 - \frac{\beta \ell \mathcal{B}''(s)}{16\pi Bp} < Q < Q_0 + \frac{\beta \ell \mathcal{B}''(s)}{16\pi Bp}.$$ 

This is the stopband width but in reality is a perturbation in the motion of the particle itself.

We can write the expression for amplitude perturbation

$$\frac{\Delta a}{a} = \frac{\beta \ell \mathcal{B}''(s)}{8Bp} \sin 3Q\theta.$$ 

Suppose the third integer $Q$-value is somewhere in the band. Then, after a sufficient number of turns, the perturbed $Q$ of the machine will be modulated to coincide with $3p$. On each subsequent revolution this increment in amplitude builds up until the particle is lost. Growth is rapid and the modulation of $Q$ away from the resonant line is comparatively slow.

Before leaving these expressions, we should note that the resonant condition, $3Q = \text{integer}$, arises because of the $\cos^3 Q\theta$ which occurs early in the above analysis, which in turn stems from the $x^2$ dependence of the sextupole field. This reveals the link between the order of the multipole and that of the resonance. We also see that the $a^2$ in leads to a linear dependence of width upon amplitude. The reader may care to speculate how other higher-order multipoles link to other lines in the working diagram and their degree of non-linearity expressed as the exponent of $a$.

But returning to the subject of stopbands, if $Q_0$ is a distance $\Delta Q$ from the third integer resonance, particles with amplitudes less than

$$a < \frac{16\pi(Bp)\Delta Q}{\beta \ell \mathcal{B}''}$$

will never reach a one third integer $Q$ and are in a central region of stability. All of this is perhaps best visualised in transverse phase space shown in Fig. 15 where the circular orbit of small amplitude becomes triangular for particles of larger amplitude experiencing more perturbation from the non-linear field. At a certain amplitude, obtained by replacing the inequality above by an equality, the effect of the non-linear force just brings the $Q$ to the $1/3$ integer value and we find metastable fixed points in phase space where there is resonant condition but infinitely slow growth.

Fig. 15 Third-order separatrix
For a one third integer resonance there are three fixed points at \( \theta = 0, 2\pi/3, \) and \( 4\pi/3. \) For a resonance of order, \( n, \) there will be \( n \) such points. The fixed points are joined by a separatrix, which is the bound of stable motion. This stable area is a rudimentary example of dynamic aperture. A more rigorous theory, which takes into account the perturbation in amplitude, would tell us that the separatrix is triangular in shape with three arms to which particles cling on their way out of the machine.

15. **INJECTION STUDIES AT FNAL**

As a cautionary tale to complete this motivation towards precision in magnet measurement, we should examine the contour model of beam survival during the one second it takes to inject into the FNAL main ring. The peaks of the mountains in Fig. 17 show the small regions in the \( Q (v \text{ in US parlance}) \) diagram where only 20 or 30% of the beam is lost due to the steep valleys of non-linear resonances in the working diagram. The driving term was a remanent sextupole randomly distributed among the dipoles of the ring. This machine broke new ground while its predecessors, much smaller machines without an injection dwell time, had hardly experienced the problem of non-linear resonances. Its constructors may therefore be forgiven for not taking stringent precautions against non-linear resonances. Measurements which might have predicted such a driving term and prompted a remedy were preserved for only one sample magnet and there was no attempt to smooth out the effect by

![Fig. 17 Beam survival impaired by resonances in the FNAL main ring](image)

shuffling laminations. Magnet ripple, another effect which scales adversely with ring size, ensured that the resonances were broadened to make it virtually impossible to obtain full survival. Fortunately the resonances were subsequently compensated and the machine went on to break all records, at the cost of a year's delay in understanding and correcting the problem. Naturally the machines which followed took good care to learn from this experience — a practice we should not forget to emulate!
1. TYPES OF STORAGE-RING MAGNETS

1.1 What is a storage ring?

A storage ring is the last stage in a chain of accelerators designed to produce beams of charged particles for experiments in nuclear or high energy physics [1]. The beam is prepared in various pre-accelerators before being injected at low energy into the main storage ring. At the end of injection, the beam is accelerated to the desired energy. Once the nominal energy is reached, the beam is circulated in the storage ring for as long as possible (typically up to 24 hours) and the physics experiments are performed. There are two types of experiments: 1) fixed-target experiments, for which the beam is extracted from the storage ring to be blasted against a fixed target, and 2) colliding-beam experiments, for which two counter-rotating beams are blasted at each other. The collision products are analyzed in large detector arrays which surround the targets or the collision points.

A storage ring is designed as a synchrotron-type accelerator and the beam is circulated on an ideally circular orbit which remains the same throughout injection, acceleration and storage. The charged particles are accelerated by means of electrical fields and are guided and focused by means of magnetic fields. The electrical fields are provided by RF cavities. In large particle accelerators, the bending and focusing functions are separated: the former is provided by dipole magnets while the latter is provided by pairs of focusing/defocusing quadrupole magnets (see below). The magnets are arranged around the ring in a regular lattice of cells, constituted of a focusing quadrupole, a set of bending dipoles, a defocusing quadrupole and another set of bending dipoles. During acceleration, the field and field gradient of the magnets are raised in proportion to particle momentum to maintain the beam on the design orbit and to preserve its size and intensity.

1.2 Bending and focusing magnets

1.2.1 Coordinate system definitions

Let \((O, \hat{u}, \hat{v}, \hat{w})\) designate a rectangular coordinate system and let \((C)\) be a circle of center \(O\), located in the \((\hat{u}, \hat{v})\) plane and representing the design orbit of a storage ring. Furthermore, let \(P\) be a given point of \((C)\) and let \((P, \hat{x}, \hat{y}, \hat{z})\) designate a rectangular coordinate system associated with \(P\), such that \(\hat{x}\) is a unit vector parallel to (OP), \(\hat{y}\) and \(\hat{w}\) are one and the same and \(\hat{z}\) is tangent to \((C)\) at \(P\). The \(x\)-axis defines the horizontal direction, the \(y\)-axis defines the vertical direction and the \(z\)-axis corresponds to the main direction of particle motions.

1.2.2 Normal dipole magnet

A normal dipole magnet is a magnet, which, when positioned in \(P\), produces within its aperture a magnetic flux density parallel to the \((\hat{x}, \hat{y})\) plane and such that

\[
B_x = 0 \quad \text{and} \quad B_y = B_1
\]
where $B_x$ and $B_y$ are the $x$- and $y$-components of the flux density and $B_1$ is a constant.

According to Lorentz' law, a charged particle traveling along the direction of the $z$-axis through the aperture of such a magnet is deflected on a circular trajectory parallel to the horizontal ($\bar{x}, \bar{z}$) plane. The trajectory radius of curvature, $\chi$, is given by

$$\chi = \frac{p}{0.3 \, q \, B_1} \quad (2)$$

Here, $\chi$ is in meters, $B_1$ is in teslas, $q$ is the particle charge in units of electron charge, and $p$ is the particle momentum in GeV/c, where $c$ is the speed of light ($2.998 \times 10^8$ m/s). The effect of a dipole magnet on a beam of charged particles is similar to that of a prism on a light ray.

Equation (2) shows that, to maintain a constant radius of curvature as the particle is accelerated, the dipole field must be ramped up in proportion to particle momentum.

1.2.3 Normal quadrupole magnet

A normal quadrupole magnet is a magnet, which, when positioned in $P$, produces within its aperture a magnetic flux density parallel to the ($\bar{x}, \bar{y}$) plane and such that

$$B_x = g \, y \quad \text{and} \quad B_y = g \, x \quad (3)$$

where $g$ is a constant referred to as the quadrupole field gradient (in T/m).

According to the Lorentz law, a beam of positively charged particles traveling along the direction of the $z$-axis through the aperture of such a magnet is horizontally focused and vertically defocused when $g$ is positive, and vertically focused and horizontally defocused when $g$ is negative. In reference to its action along the $x$-axis on a beam of positively charged particles traveling in the $z$-direction, a magnet with a positive gradient is called a focusing quadrupole, while a magnet with a negative gradient is called a defocusing quadrupole. To obtain a net focusing effect along both $x$- and $y$-axes, a pair of focusing/defocusing quadrupoles must be used. For both types of quadrupole, the focal length, $f$, is given by

$$f = \frac{p}{0.3 \, q \, g \, l_q} \quad (4)$$

Here, $f$ is in meters, $p$ is in GeV/c, $q$ is in units of electron charge, $g$ is in T/m and $l_q$ is the quadrupole magnetic length in meters. The effect of focusing/defocusing quadrupoles on a beam of charged particles is similar to that of convex/concave lenses on a light ray.

Equation (4) shows that to maintain $f$ constant as the particle beam is accelerated, the quadrupole field gradient must be ramped up in proportion to beam energy.

2. STORAGE RINGS AND SUPERCONDUCTIVITY

2.1 Why superconductivity?

Throughout the years, the quest for elementary particles has promoted the development of accelerator complexes producing beams of increasingly higher energies. Equation (2) shows that, for a synchrotron, the particle momentum is directly related to the product ($\chi B_1$). Hence, to reach higher energies, one must increase either the accelerator radius or the dipole field (or both). Increasing the accelerator radius means a bigger tunnel. Increasing the dipole field above 2 T implies the use of superconducting magnets. The trade-off between tunneling costs, magnet development costs and accelerator operating costs is, since the late 1970's, in
favor of using superconducting magnets generating the highest possible field and field
gradient [2].

Superconductivity is a unique property exhibited by some materials at low temperatures
where the resistivity drops to zero. As a result, materials in the superconducting state can
transport current without power dissipation by the Joule effect. This offers at least two
advantages for large magnet systems such as those needed in storage rings: (1) significant
reduction in electrical power consumption and (2) the possibility of relying on much higher
overall current densities in the magnets coils. There are, however, at least three drawbacks in
using superconducting magnets: (1) the superconductor generates magnetization effects which
result in field distortions that have to be corrected (see section on field quality), (2) the
magnets must be cooled down and maintained at low temperatures, which requires large
cryogenic systems (see section on magnet cooling) and (3) it may happen that an energized
magnet, initially in the superconducting state, abruptly and irreversibly switches back to the
normal resistive state in a phenomenon referred to as a quench (see section on quench
performance).

The occurrence of a quench causes an instantaneous beam loss and requires that all or
part of the magnet ring be rapidly ramped down to limit conductor heating in the quenching
magnet (see section on quench protection). Once the quenching magnet is discharged, it can
be cooled down again and restored into the superconducting state, and the machine operations
can resume. A quench is seldom fatal but is always a serious disturbance. Everything must
be done to prevent it from happening and all precautions must be taken to ensure the safety of
the installation when it does happen.

2.2 Review of superconducting storage rings

2.2.1 Tevatron

The first large scale application of superconductivity was the Tevatron, a proton
synchrotron with a circumference of 6.3 km built at Fermi National Accelerator Laboratory
(FNAL) near Chicago, Illinois and commissioned in 1983 [3]. The Tevatron now operates as
a proton/antiproton collider with a maximum energy of 900 GeV per beam. It relies on about
1000 superconducting dipole and quadrupole magnets, with a maximum operating dipole field
of 4 T [4].

2.2.2 HERA

The next large particle accelerator to rely massively on superconducting magnet
technology was HERA (Hadron Elektron Ring Anlage) built at DESY (Deutsches
Elektronen–SYnchrotron) near Hamburg, Germany and commissioned in 1990 [5]. HERA is
an electron/proton collider with a circumference of 6.3 km. It is composed of two storage
rings: (1) an electron ring, relying on conventional magnets (maximum energy: 30 GeV) and
(2) a proton ring, relying on superconducting magnets (maximum energy: 820 GeV).
The maximum operating field of the superconducting dipole magnets is 4.7 T [6]. The dipole
magnets of the proton ring were developed at DESY, while the quadrupole magnets were
developed at CEA/Saclay (Commissariat à l'Energie Atomique at Saclay near Paris, France).

2.2.3 UNK

Since the early 1980's, the Institute for High Energy Physics (IHEP) located in Protvino,

near Moscow, Russia is working on the project of a proton accelerator named UNK
(Uскорительный-Накопительный Комплекс). The circumference of UNK is 21 km for a maximum
energy of 3 TeV in a fixed target mode [7]. The maximum operating dipole field is 5 T [8]. A
number of superconducting dipole and quadrupole magnet prototypes have been built and
cold-tested and the tunnel is almost completed, but, given the economical situation in Russia, the future of the machine is undecided.

2.2.4 SSC

In the mid 1980’s, the USA started the Superconducting Super Collider (SSC) project, a giant proton/proton collider with a maximum energy of 20 TeV per beam [9]. The SSC would have been constituted of two identical storage rings of superconducting magnets installed on top of each other in a tunnel with a circumference of 87 km. The maximum operating dipole field was 6.8 T. The project was eventually canceled in October 1993 by decision of the United States Congress, after 12 miles of tunnel had been dug near Dallas, Texas, and a successful superconducting magnet R&D program had been carried out [10].

2.2.5 RHIC

Brookhaven National Laboratory (BNL), located on Long Island, New York, will complete in 1999 the construction on its site of the Relativistic Heavy Ion Collider (RHIC). RHIC is designed to collide beams of nuclei as heavy as gold, accelerated in two identical storage rings to energies between 7 and 100 GeV per beam and per unit of atomic mass [11]. Each ring has a circumference of 3.8 km; the maximum operating dipole field is 3.4 T [12].

2.2.6 LHC

Finally, in December 1994, the European Laboratory for Particle Physics (CERN) approved the construction of the Large Hadron Collider (LHC) in its existing 27-km-circumference tunnel located at the Swiss/French border, near Geneva, Switzerland [13]. LHC will be a proton/proton collider with a maximum energy of 7 TeV per beam. It will consist of a ring of so-called twin-aperture superconducting magnets, housing within the same mechanical structure, the pipes for two counter-rotating proton beams [14]. The maximum operating dipole field is 8.36 T. The dipole magnets are developed at CERN while the quadrupole magnets are developed at CEA/Saclay. Commissioning is planned for 2005.

2.2.7 Prominent features of superconducting storage ring magnets

Selected parameters of the major projects of superconducting storage rings are summarized in Table 1, while Figs. 1(a) through 1(e) present cross-sectional views of the Tevatron, HERA, SSC, RHIC and LHC dipole magnets [15].

<table>
<thead>
<tr>
<th>Laboratory</th>
<th>FNAL</th>
<th>DESY</th>
<th>IHEP</th>
<th>SSCL</th>
<th>BNL</th>
<th>CERN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Tevatron</td>
<td>HERA</td>
<td>UNK</td>
<td>SSC</td>
<td>RHIC</td>
<td>LHC</td>
</tr>
<tr>
<td>Circumference (km)</td>
<td>6.3</td>
<td>6.3</td>
<td>21</td>
<td>87</td>
<td>3.8</td>
<td>27</td>
</tr>
<tr>
<td>Particle type</td>
<td>p\overline{p}</td>
<td>ep</td>
<td>pp</td>
<td>pp Ions</td>
<td>Heavy</td>
<td>PP</td>
</tr>
<tr>
<td>Energy/beam (TeV)</td>
<td>0.9</td>
<td>0.82</td>
<td>3</td>
<td>20</td>
<td>up to 0.1\textsuperscript{a)}</td>
<td>7</td>
</tr>
<tr>
<td>Number of dipoles</td>
<td>774</td>
<td>416</td>
<td>2168</td>
<td>7944</td>
<td>264</td>
<td>1232\textsuperscript{b)}</td>
</tr>
<tr>
<td>Aperture (mm)</td>
<td>76.2</td>
<td>75</td>
<td>70</td>
<td>50</td>
<td>80</td>
<td>56</td>
</tr>
<tr>
<td>---------------</td>
<td>------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Magnetic length (m)</td>
<td>6.1</td>
<td>8.8</td>
<td>5.8</td>
<td>15</td>
<td>9.7</td>
<td>14.2</td>
</tr>
<tr>
<td>Field (T)</td>
<td>4</td>
<td>4.68</td>
<td>5.0</td>
<td>6.79</td>
<td>3.4</td>
<td>8.36</td>
</tr>
<tr>
<td>Number of quads.</td>
<td>216</td>
<td>256</td>
<td>322</td>
<td>1696</td>
<td>276</td>
<td>386</td>
</tr>
<tr>
<td>Aperture (mm)</td>
<td>88.9</td>
<td>75</td>
<td>70</td>
<td>50</td>
<td>80</td>
<td>56</td>
</tr>
<tr>
<td>Mag. length&lt;sup&gt;c)&lt;/sup&gt; (m)</td>
<td>1.7</td>
<td>1.9</td>
<td>3.0</td>
<td>5.7</td>
<td>1.1</td>
<td>3.1</td>
</tr>
<tr>
<td>Gradient T/m</td>
<td>76</td>
<td>91.2</td>
<td>97</td>
<td>194</td>
<td>71</td>
<td>223</td>
</tr>
<tr>
<td>Commissioning</td>
<td>1983</td>
<td>1990</td>
<td>undecided</td>
<td>cancelled</td>
<td>1999</td>
<td>2005</td>
</tr>
</tbody>
</table>

<sup>a)</sup> per unit of atomic mass,  
<sup>b)</sup> two-in-one magnets,  
<sup>c)</sup> quadrupoles come in several lengths

Fig. 1 Cross-sectional views of superconducting dipole magnets for large particle accelerator magnets [15]: (a) Tevatron, (b) HERA, (c) SSC, (d), RHIC and (e) LHC.

The magnets rely on similar design principles which are detailed in the oncoming sections. The field is produced by saddle-shape coils that, in their long straight sections, approximate \( \cos \theta \) conductor distributions for dipole magnets and \( \cos 2 \theta \) conductor distributions for quadrupole magnets. The coils are wound from Rutherford-type cables made of NbTi multifilamentary strands and are mechanically restrained by means of laminated collars. The collared-coil assembly is placed within an iron yoke providing a return path for the magnetic flux. In the case of Tevatron, the collared-coil assembly is cold while the iron yoke is warm. Starting with HERA, the iron yoke is included in the magnet cryostat and the cold mass is completed by an outer shell delimiting the region of helium circulation. The cold mass of the LHC magnets include two collared-coil assemblies within a common yoke. Tevatron, HERA, UNK, SSC and RHIC magnets are cooled by boiling helium at
1 atmosphere (4.2 K) or supercritical helium at 3 to 5 atmosphere (between 4.5 and 5 K) while LHC magnets are cooled by superfluid helium at 1.9 K.

2.3 Superconducting storage-ring magnet R&D

A number of laboratories are presently involved in R&D work on high field or high field gradient accelerator magnets. Among them is Twente University, located near Enschede in the Netherlands, which, in 1995, cold-tested at CERN a short model dipole (made with Nb$_3$Sn cable) which reached 11 T on its first quench at 4.4 K [16]. Soon after, in 1996, Lawrence Berkeley National Laboratory (LBNL), located in Berkeley, California cold-tested a short model dipole (also made with Nb$_3$Sn cable), referred to as D20, which, after a number of training quenches, reached a record dipole field of 13.5 T at 1.8 K [17].

3. CONDUCTOR AND CONDUCTOR INSULATION

3.1 Superconducting material

3.1.1 NbTi

The most widely used superconducting material is a metallic alloy of niobium and titanium (NbTi), with a Ti content between 45 and 50% in weight [18]. NbTi is easy to mass-produce and has good mechanical properties. It is a type-II superconductor, with a coherence length, $\xi$, of 5 nm, and a London penetration depth, $\lambda$, of 300 nm (Chapter 2 of Reference 2).

The upper critical magnetic flux density, $B_{c2}$, can be estimated as a function of temperature, $T$, using

$$B_{c2}(T) = B_{c20} \left[ 1 - \left( \frac{T}{T_{c0}} \right)^{1.7} \right]$$

(5)

where $B_{c20}$ is the upper critical magnetic flux density at zero temperature (about 14.5 T) and $T_{c0}$ is the critical temperature at zero magnetic flux density (about 9.2 K).

The critical current density, $J_C$, can be parametrized as a function of temperature, magnetic flux density, $B$, and critical current density at 4.2 K and 5 T, $J_{C_{\text{ref}}}$, using [19]

$$\frac{J_C(B,T)}{J_{C_{\text{ref}}}} = C_0 \frac{B}{B_{c2}(T)} \left[ 1 - \frac{B}{B_{c2}(T)} \right]^{-\alpha} \left[ 1 - \left( \frac{T}{T_{c0}} \right)^{1.7} \right]^{-\beta}$$

(6)

where $C_0$, $\alpha$, $\beta$ and $\gamma$ are fitting parameters.

Since the time of the Tevatron, a factor of about 2 has been gained on the critical current density at 4.2 K and 5 T and values in excess of 3000 A/mm$^2$ are now obtained in industrial production [20]. Typical fitting parameters values for LHC strands are: $C_0 = 30$ T, $\alpha = 0.6$, $\beta = 1.0$ and $\gamma = 2.0$.

The highest dipole field reached on a NbTi magnet is 10.53 T at 1.77 K [21].

3.1.2 Nb$_3$Sn

Magnet designers consider that 10 to 11 T is about the limit for NbTi and that to produce higher fields, it is necessary to change the material. The only other material that is readily available on an industrial scale is an intermetallic compound of niobium and tin (Nb$_3$Sn) belonging to the A15 crystallographic family [18]. Nb$_3$Sn presents interesting superconducting properties, which may be enhanced by a small addition of titanium or tantalum in the niobium. However, the compound formation requires a heat treatment at temperatures up to 700 °C for times up to 300 hours in a vacuum or in inert atmosphere such
as argon. Furthermore, once reacted, the compound becomes brittle and its properties are strain sensitive.

The upper critical magnetic flux density, $B_{C2}$, can be estimated as a function of temperature, $T$, and strain, $\varepsilon$, using [22]

$$\frac{B_{C2}(T, \varepsilon)}{B_{C20}(\varepsilon)} = \left[ 1 - \frac{T}{T_{C0}(\varepsilon)} \right] \left[ 1 - 0.31 \left( \frac{T}{T_{C0}(\varepsilon)} \right)^2 \right] \left[ 1 - 1.77 \ln \left( \frac{T}{T_{C0}(\varepsilon)} \right) \right]$$

(7)

where $B_{C20}$ is the upper critical magnetic flux density at zero temperature

$$B_{C20}(\varepsilon) = B_{C20m} (1 - a |\varepsilon|^{1.7})$$

(8)

and $T_{C0}$ is the critical temperature at zero magnetic flux density

$$T_{C0}(\varepsilon) = T_{C0m} (1 - a |\varepsilon|^{1.7})^{1/3}$$

(9)

Here, $a$ is a parameter equal to 900 for compressive strain ($\varepsilon < 0$) and to 1250 for tensile strain ($\varepsilon > 0$). $B_{C20m}$ is the upper critical magnetic flux density at zero temperature and zero strain and $T_{C0m}$ is the critical temperature at zero magnetic flux density and zero strain. For binary compounds, $T_{C0m}$ and $B_{C20m}$ can be taken equal to 16 K and 24 T, while for ternary compounds, they can be taken equal to 18 K and 28 T.

The critical current density can be parametrized as a function of temperature, $T$, magnetic flux density, $B$, and strain, $\varepsilon$, using [22]

$$J_C(B, T, \varepsilon) = \frac{C(\varepsilon)}{\sqrt{B}} \left[ 1 - \frac{B}{B_{C2}(T, \varepsilon)} \right]^2 \left[ 1 - \left( \frac{T}{T_{C0}(\varepsilon)} \right)^2 \right]^2$$

(10)

where

$$C(\varepsilon) = C_0 (1 - a |\varepsilon|^{1.7})^{1/2}$$

(11)

Here $C_0$ is a fitting parameter.

In recent years, significant R&D work has been carried out to improve the performance of Nb$_3$Sn multifilamentary wires, thanks to the International Thermonuclear Experimental Reactor (ITER) program [23]. Critical current density values of 750 A/mm$^2$ at 4.2 K and 12 T with effective filament diameters of 15 to 20 $\mu$m are now reached in industrial production [24]. Such values correspond to a $C_0$ of the order 12000 AT$^{1/2}$-mm$^{-2}$. Note that the strain in a free standing Nb$_3$Sn multifilamentary wire is estimated at about -0.25%.

Given that reacted Nb$_3$Sn conductors are very fragile and cannot be bent on small radii, the manufacturing of Nb$_3$Sn coils calls for special fabrication processes which are risky and onerous and which, so far, have limited the use of this material. In the case of accelerator magnet coils, the conductor is wound un-reacted, and the whole coil is subjected to heat-treatment, according to the so-called wind-and-react technique.

As already mentioned, the highest dipole field reached on a Nb$_3$Sn magnet is 13.5 T at 1.8 K [17].

3.1.3 High-temperature superconductors

Although great progress has been made in the development of so-called high temperature superconductors (HTS), such as bismuth copper oxides, Bi$_2$Sr$_2$CaCu$_2$O$_x$ and (Bi,Pb)$_2$Sr$_2$Ca$_2$Cu$_3$O$_x$, and yttrium copper oxides, YBa$_2$Cu$_3$O$_7$, these materials are not yet
ready for applications requiring low cost, mass-production and high critical current density [25].

3.2 Rutherford-type cable

Superconducting particle accelerator magnet coils are wound from so-called Rutherford-type cables. As illustrated in Fig. 2(a), a Rutherford-type cable consists of a few tens of strands, twisted together, and shaped into a flat, two-layer, slightly keystoned cable [26]. The strands themselves consist of thousands of superconducting filaments, twisted together and embedded in a matrix of normal metal [18]. Except for the cables used in a few R&D model magnets, the filaments are made of NbTi and the matrix is high-purity copper. Filament diameters range from 5 to 15 μm. Fig. 2(b) presents a cross-sectional view of a typical SSC strand.

![Fig. 2 Rutherford-type cable for storage ring magnet: (a) cable sketch and (b) cross-sectional view of a cable strand.](image)

The small radii of curvature of the coil ends preclude the use of a monolithic conductor because it would be too hard to bend. A multi-strand cable is preferred to a single wire for at least four reasons: (1) it limits piece length requirement for wire manufacturing (a coil wound with a N-strand cable requires piece lengths which are 1/N shorter than for a similar coil wound with a single wire), (2) it allows strand-to-strand current redistribution in the case of a localized defect or when a quench originates in one strand [27, 28], (3) it limits the number of turns and facilitates coil winding, and (4) it limits coil inductance (the inductance of a coil wound with a N-strand cable is 1/N² smaller than that of a similar coil wound with a single wire). A smaller inductance reduces the voltage requirement on the power supply to ramp-up the magnets to their operating current in a given time and limits the maximum voltage to ground in case of a quench (see quench protection section). The main disadvantage of using a cable is the high operating current (over a few thousand amperes) which requires large current supplies and large current leads.

The main issues for NbTi strand design and manufacturing are: (1) copper-to-superconductor ratio, which should not be too small to limit conductor heating in case of a quench while achieving a high overall critical current, (2) filament size, to limit field distortions resulting from superconductor magnetization at low field (see field quality section), (3) superconductor critical current density, which can be improved by improving pinning and filament uniformity [18] and (4) piece length.

The main issues for cable design and fabrication are: (1) compaction, which should be large enough to ensure good mechanical stability and high overall current density while leaving enough void for helium cooling, (2) control of outer dimensions to achieve suitable coil geometry and mechanical properties, (3) limitation of critical current degradation [29, 30] and (4) control of interstrand resistance, which should not be too small to limit field distortions induced by coupling currents while ramping (see field quality section) and should not be too large to allow current redistribution among cable strands.
The interstrand resistance can be modified by oxidizing or by coating strand surface [31, 32]. Also, a thin, insulating foil (such as stainless steel) can be inserted between the two strand layers of the cable [33]. The strands used in HERA and LHC cables are coated with a silver-tin solder, called Stabrite. Half of the strands of the Tevatron cable are coated with Stabrite, while the other half is insulated with a black copper oxide, called Ebanol. UNK, SSC and RHIC cables rely on natural oxidation. Up to now, no foiled cable has been used in a magnet.

Note that at the end of cabling, the high purity copper of the strand matrix is heavily cold-worked and that it may require an annealing procedure.

3.3 Cable insulation

3.3.1 Insulation requirements

The main requirements for cable insulation are: (1) good dielectric strength in helium environment and under high transverse pressure (up to 100 MPa), (2) small thickness (to maximize overall current density in magnet coil) and good physical uniformity (to ensure proper conductor positioning for field quality), (3) retention of mechanical properties in a wide temperature range, and (4) ability to withstand radiations in an accelerator environment. In addition, the insulation system is required to provide a means of bonding the coil turns together to give the coil a rigid shape and facilitate its manipulation during the subsequent steps of magnet assembly. It is also desirable that the insulation be somewhat porous to helium for conductor cooling. Note that the dielectric strength of helium gas at 4.2 K is far worse than that of liquid helium and that it degrades significantly with increasing temperature [34].

3.3.2 Insulation of NbTi Cables

The insulation of Tevatron, HERA and UNK magnets, of most SSC magnets and of the early LHC models is constituted of one or two inner layers of polyimide film, wrapped helically with a 50-to-60% overlap, completed by an outer layer of resin-imregnated glass-fiber tape, wrapped helically with a small gap. The inner layer is wrapped with an overlap for at least two reasons: (1) the polyimide film may present pin holes which have to be covered (the probability of having two superimposed pin holes in the overlapping layer is very low) and (2) the Tevatron experience has shown that it was preferable to prevent the resin impregnating the glass wrap from entering in contact with the NbTi cable (the energy released by cracks in the resin is believed to be sufficient to initiate a quench) [p. 784 of Reference 4]. The outer layer is wrapped with a gap to set up helium cooling channels between coil turns. The resin is of thermosetting-type and requires heat to increase cross-link density and cure into a rigid bonding agent. Curing is done after completion of the winding and in a mold of very accurate dimensions to control coil geometry and Young’s modulus [35].

RHIC magnets and the most recent LHC models used a so-called all-polyimide insulation where the outer glass-fiber wrap is replaced by another layer of polyimide film with a polyimide adhesive on its surface [36]. The all-polyimide insulation has a better resistance to puncture but the softening temperature of the adhesive can be higher than the temperature needed to cure a conventional resin (225 °C for RHIC-type all-polyimide insulation compared to 135 °C for SSC-type polyimide/glass insulation).

3.3.3 Insulation of Nb₃Sn cables

The insulation of Nb₃Sn cables is usually based on a glass-fiber tape or sleeve put on the conductor prior to winding. At the end of the heat treatment needed for Nb₃Sn formation, the reacted coil is transferred to a precision molding fixture and is vacuum impregnated with
resin. The glass fibers used for the tape or the sleeve must be able to sustain the heat treatment without degradation. Also, all organic materials, such as sizing or finish, must be removed from the fibers to prevent the formation of carbon compounds that may lower the dielectric strength. The sizing removal is performed by carbonisation in air prior to conductor insulation.

Using such an insulation system adds to the difficulty of manufacturing Nb₃Sn coils for at least two reasons: 1) de-sized glass-fiber tapes or sleeves are fragile and easy to tear off by friction and 2) vacuum impregnation is a delicate operation. Furthermore, a full impregnation prevents any helium penetration in the coil greatly reducing cooling capabilities.

4. MAGNETIC DESIGN

4.1 Field produced by simple current distributions

4.1.1 Single current line in free space

Let \((O,\bar{x},\bar{y},\bar{z})\) designate a rectangular coordinate system and let \((-I,R,\theta)\) designate a current-line of intensity \((-I)\), parallel to the \(z\)-axis, and located at a position \(s = R \exp (i \theta)\) in the complex \((O,\bar{x},\bar{y})\) plane, as represented in Fig. 3(a). The magnetic flux density, \(\vec{B}\), produced by this current-line in free space can be computed using Biot and Savart's law. It is uniform in \(z\) and parallel to the \((\bar{x},\bar{y})\) plane and its \(x\)- and \(y\)-components, \(B_x\) and \(B_y\) are given by

\[
B_y + i B_x = \frac{\mu_0 I}{2\pi} \frac{1}{(z - s)}
\]

where \(\mu_0\) is the magnetic permeability of vacuum \((4\pi \times 10^{-7} \text{ H/m})\) and \(z = x + iy\).

The above expression can be expanded into a power series of the form [37]

\[
B_y + i B_x = \sum_{n=1}^{+} (B_n + i A_n) z^{n-1} \quad \text{for } z = x + iy, |z| < R
\]

where \(A_n\) and \(B_n\) are constant coefficients, referred to as normal and skew \(2n\)-pole field coefficients, given by

\[
B_n + i A_n = \frac{\mu_0 I}{2\pi R^n} [\cos(n \theta) - i \sin(n \theta)]
\]

Fig. 3 Representations of a single current-line
(a) in a vacuum and (b) inside a circular iron yoke.

4.1.2 Single current line within a circular iron yoke

Let us now assume that the current line of Fig. 3(a) is located inside a circular iron yoke of inner radius, \(R_y\), as represented in Fig. 3(b). The contribution of the iron yoke to the
The magnetic flux density can be shown to be the same as that of a mirror current line, of intensity, \((-I_m)\), and position, \(s_m\), in the complex plane, where

\[
I_m = \frac{\mu - 1}{\mu + 1} I \quad \text{and} \quad s_m = \frac{R_y^2}{s^*} \tag{15}
\]

Here \(\mu\) is the relative magnetic permeability of the iron yoke and \(s^*\) the complex conjugate of \(s\). Note that the mirror image method is only applicable if the iron yoke is not saturated and its permeability is uniform.

4.1.3 Quadruplet of current lines with dipole symmetry

Using the above expressions, the magnetic flux density produced by the quadruplet of current lines \((-I,R,\theta), (+I,R,\pi-\theta), (+I,R,\pi+\theta)\) and \((-I,R,-\theta)\), represented Fig. 4(a), can be estimated from the power series expansion

\[
B_y + iB_x = \sum_{k=0}^{+} B_{2k+1} z^{2k} \quad \text{for} \; z = x + iy, |z| < R \tag{16}
\]

where

\[
B_{2k+1} = \frac{2\mu_0 I}{\pi R^{2k+1}} \cos[(2k+1)\theta] \tag{17}
\]

The first term \((k = 0)\) of the series corresponds to a pure normal dipole field parallel to the \(y\)-axis. The \(B_{2k-1}\) coefficients are called the allowed multipole field coefficients of this current distribution.

Fig. 4 Examples of current-line distributions with selected symmetries (a) quadruplet of current-lines with an even symmetry about the \(x\)-axis and an odd symmetry about the \(y\)-axis and (b) octuplet of current-lines with even symmetries with respect to the \(x\)- and \(y\)-axes and odd symmetries with respect to the first and second bisectors.

4.1.4 Octuplet of current lines with quadrupole symmetry

Similarly, the magnetic flux density produced by the octuplet of current lines represented in Fig. 4(b) is given by

\[
B_y + iB_x = \sum_{k=0}^{+} B_{4k+2} z^{4k+1} \quad \text{for} \; z = x + iy, |z| < R \tag{18}
\]

where

\[
B_{4k+2} = \frac{4\mu_0 I}{\pi R^{4k+2}} \cos[(4k+2)\theta] \tag{19}
\]
The first term ($k = 0$) of the series corresponds to a pure normal quadrupole field whose axes are parallel to the first and second bisectors. For this current distribution, the allowed multipole field coefficients are the normal ($4k + 2$)-pole field coefficients.

4.1.5 Cos$\theta$ and sin$\theta$ current sheets

Let us now consider a cylindrical current sheet of radius, $R$, carrying a linear current density of the form $[-j\cos(\theta)]$ where $j$ is a constant (in $A/m$). The magnetic flux density produced within the sheet can be computed by dividing the sheet into elementary current lines of intensity $[-jR\cos(\theta)d\theta]$ and by integrating their contributions over $(2\pi)$. We get

$$B_y + iB_x = B_p = \frac{\mu_0 j}{2R^{p-1}} \quad \text{for } z = x + iy, |z| < R \quad (20)$$

Hence, a cos($\theta$)-type current sheet produces a pure normal $2p$-pole field.

Similarly, it can be shown that a cylindrical current sheet of radius, $R$, carrying a linear current density $[+j\sin(\theta)]$ produces a pure skew $2p$-pole field

$$B_y + iB_x = A_p = \frac{\mu_0 j}{2R^{p-1}} \quad \text{for } z = x + iy, |z| < R \quad (21)$$

4.1.6 Cylindrical current shells

Let us finally consider a cylindrical current shell of inner radius, $R_i$, outer radius, $R_o$, pole angle $\theta_0$, carrying a uniform current density $(-J)$ for $x, x > 0$ and $(+J)$ for $x, x < 0$, as represented Fig. 5(a). The magnetic flux density produced within the cylinder can be computed by dividing the shell into quadruplets of current lines having the symmetry of Fig. 4(a) and by integrating their contributions over a shell quadrant. It follows that the magnetic flux density is given by Eq. (10), but the expressions of the multipole field coefficients become

$$B_1 = \frac{2\mu_0 J}{\pi} (R_o - R_i) \sin \theta_0 \quad (22a)$$

and

$$B_{2k+1} = \frac{2\mu_0 J}{\pi(2k+1)(2k-1)} \left( \frac{1}{R_i^{2k-1}} - \frac{1}{R_o^{2k-1}} \right) \sin[(2k+1)\theta_0] \quad \text{for } k, k \geq 1 \quad (22b)$$

Note that $B_3$ (first allowed multipole field coefficient after $B_1$ in a current distribution with a dipole symmetry) is nil for $\theta_0 = \pi/3$.

![Fig. 5](image-url) Current shell approximations for the generation of multipole fields: (a) dipole field and (b) quadrupole field.

Similarly, it can be shown that magnetic flux density produced by the current shell of Fig. 5(b) is given by Eq. (12), where
\[ B_2 = \frac{2\mu_0 J}{\pi} \ln \left( \frac{R_o}{R_1} \right) \sin 2\theta_0 \]  

(23a)

and

\[ B_{4k+2} = \frac{\mu_0 J}{\pi k(4k+2)} \left( \frac{1}{R_1^{4k}} - \frac{1}{R_o^{4k}} \right) \sin [(4k+2)\theta_0] \quad \text{for } k, k \geq 1 \]  

(23b)

Note that \( B_2 \) corresponds to the quadrupole field gradient, \( g \), and that \( B_6 \) (first allowed multipole field coefficient after \( B_2 \) in a current distribution with a quadrupole symmetry) is nil for \( \theta_0 = \pi/6 \).

4.2 Two-dimensional geometry

4.2.1 Symmetry considerations

The field computations carried out in the previous section have shown that current distributions with the symmetries of Fig. 4(a) (i.e., even with respect to the \( x \)-axis and odd with respect to the \( y \)-axis) were fitted to the generation of dipole fields, while current distributions with the symmetries of Fig. 4(b) (i.e., even with respect to the \( x \)- and \( y \)-axes and odd with respect to the first and second bisectors) were fitted to the generation of quadrupole fields. Starting from these premises, the coil geometry can be optimized to obtain the required dipole or quadrupole field strength within the magnet aperture, with the smallest possible contributions from non-dipole or non-quadrupole terms.

4.2.2 Current shell approximations

The coil geometries the most commonly used for dipole and quadrupole magnets are approximations of the cylindrical current shells shown in Figs. 5(a) and 5(b). The approximation is obtained by stacking into an arch the slightly keystoned cables described in the conductor section. The low field and field gradient magnets for RHIC rely on a single coil layer while Tevatron, HERA, UNK, SSC and LHC magnets rely on two coil layers whose contributions add up. The high field LBNL model magnet D20 count four layers. In addition, in most accelerator magnet coil designs, copper wedges are introduced between some of coil turns to separate the conductors into blocks. The blocks’ angles are then optimized to eliminate high order multipole field coefficients and approach the ideal \( \cos \theta \) and \( \cos 2\theta \) conductor distributions [37]. By extension, such coil geometries are referred to as \( \cos \theta \) and \( \cos 2\theta \) designs. They are very compact and make the most effective use of conductors by bringing them close to the useful aperture.

In the case of Tevatron, HERA and UNK magnets, the cable keystone angle is large enough to allow the formation of an arch with the desired aperture. Furthermore, each coil turn is positioned radially. In the case of SSC and LHC magnets, the coil aperture is reduced to minimize the volume of superconductor. This results in a keystone angle requirement deemed unacceptable from the point of view of cabling degradation. Hence, in these magnets, the cables are not sufficiently keystoned to assume an arch shape and the wedges between conductor blocks must be made asymmetrical to compensate for this lack [38]. Also, the coil turns end up being non-radial, as illustrated in Fig. (6), which shows the conductor distribution in a quadrant of a 50-mm-aperture SSC dipole magnet coil (the vectors represent the components of the Lorentz force discussed in the mechanical design section).

Note that the magnetic flux density produced by the coil of Fig. (6) can be accurately computed by dividing each turn into two rows of elementary current lines parallel to the \( z \)-axis and approximately equal in number to the number of cable strands (p. 226 of Reference [39]).
4.2.4 Iron yoke contribution

The coils of particle accelerator magnets are usually surrounded by an iron yoke, which provides a return path for the magnetic flux while enhancing the central field or field gradient.

As an illustration, let us place the cylindrical current shells of Fig. 5(a) within a circular iron yoke of inner radius, \( R_y \). The contribution of the iron yoke to the normal \( (2k+1) \)-pole field coefficient, \( B_{yoke}^{2k+1} \), can be estimated as (p. 53 of Reference [2])

\[
B_{yoke}^{2k+1} = \frac{\mu - 1}{\mu + 1} \left( \frac{R_i R_0}{R_y^2} \right)^{2k+1} B_{shell}^{2k+1}
\]  

where \( \mu \) is the relative magnetic permeability of the iron yoke, \( R_i \) and \( R_o \) are the current shell inner and outer radii and \( B_{shell}^{2k+1} \) is the \((2k+1)\)-pole field coefficient produced by the current shell alone.

The above formula shows that the smaller \( R_y \), the larger the field enhancement. However, there are two limitations on how close the iron can be brought to the coils: (1) room must be left for the support structure, and (2) iron saturates for fields above 2 T, resulting in undesirable distortions (see field quality section).

As already mentioned, the Tevatron magnets use a warm iron yoke (i.e., placed outside the helium containment and vacuum vessel), but starting with HERA magnets, the iron yoke is included within the magnet cold mass. For SSC dipole magnets, the field enhancement due to the iron yoke is of the order of 20%. In LHC magnets, two coil assemblies (of opposite polarity) are placed within a common iron yoke. This twin-aperture design results in left/right asymmetries in the yoke surrounding each coil assembly taken individually which must be taken into account.

4.2.5 Operating margin

Equations 14(a) and 15(a) show that to achieve high fields and high field gradients, it is desirable to maximize the overall current density in the magnet coil. This can be done by three means: 1) maximizing the superconductor performance, 2) minimizing the copper-to-superconductor ratio in the cable strands and 3) minimizing the turn-to-turn insulation thickness. As explained in other sections, there are lower bounds on the values of copper-to-superconductor ratio and insulation thickness in order to limit conductor heating in case of quench and to ensure proper electrical insulation. As for the superconductor, the upper limit is the critical current density at the given temperature and magnetic flux density.

The magnetic flux density to which the conductor is exposed is non-uniform over the magnet coil, but the maximum current-carrying capability of the conductor is determined by the section where the magnetic flux density is the highest. In most cases, this corresponds to the pole turn of the innermost coil layer. Let \( B_p = f(I) \) designate the peak magnetic flux density on the coil as a function of supplied current, \( I \), and let \( I_c = f(B) \) designate the supposedly known cable critical current as a function of applied magnetic flux density, \( B \), at the operating temperature, \( T_0 \). The intersection between these two curves determines the maximum quench current of the magnet at \( T_0, I_{qm}(T_0) \).

In practice, magnets must be operated below \( I_{qm} \) so as to ensure that the superconductor is in the superconducting state and as to limit the risks of quenching. Let \( I_{op} \) designate the operating current, the current margin of the magnet, \( m_I \), at the operating temperature, \( T_0 \), is defined as
\[ m_1 = 1 - \frac{I_{op}}{I_{qm}(T_0)} \] (25)

The excellent quench performance of the HERA magnets \cite{6} suggests the current margin can be set to as little as 10\%, but it is safer to aim for 20\%.

In comparison to other superconducting magnets, a current margin of 10 to 20\% is quite small. This implies that storage ring magnets are operated very close to the superconductor critical surface and that they are very sensitive to any kind of disturbances that may cause a surface crossing and lead to a quench.

A particularity of a two-layer, \(\cos \theta\) dipole magnet coil design is that the peakfield in the outermost layer is quite a bit lower than in the innermost layer. Hence, when using the same cable for both layers, the outer layer is operated with a much higher current margin than the inner layer, which can be considered as a waste of costly superconductor. SSC and LHC dipole magnet coils use a smaller conductor for the outer layer than for the inner layer. This results in a higher overall current density in the outer layer and reduces the difference in current margins. Such action is referred to as conductor grading. The main disadvantage of grading is that it requires splices between inner and outer layer cables.

4.2.6 Limits of \(\cos \theta\) coil design

The \(\cos \theta\) coil design has been very successful until now, with a record dipole field of 13.5 T reached by the LBNL short model magnet D20 (using Nb\_3Sn cables at 1.8 K). However, it has two main drawbacks: (1) the coil ends are difficult to make (see section on coil ends), and (2) due to the Lorentz force distribution, there is a stress accumulation in the azimuthal direction which results in high transverse pressures on the midplane conductors (see Fig. (6)). For very high field magnets, requiring the use of A15 (or even possibly HTS) superconductors, which are strain sensitive, these high transverse pressures can result in significant critical current degradation \cite{40}.

![Conductor distribution in a quadrant of a 50-mm-aperture SSC dipole magnet coil](Fig. 6)

Alternative coil designs are being investigated which may allow a better stress management within the magnet coil. As an illustration, Fig. (7) presents a conceptual block design developed at BNL for a twin-aperture dipole magnet relying only on simple, racetrack coils \cite{41}. Note, however, that such designs make a less effective use of superconductor.

4.3 Coil-end design

One of the main difficulties of the \(\cos n\theta\) design is the realization of coil ends. In the coil straight section, the conductors run parallel to the magnet axis, but, in the coil ends, the conductors must be bent sharply with small radii of curvature to make U-turns over the beam tube that is inserted within the magnet aperture. This confers to the coil a saddle shape as illustrated in Fig. 8.
Sophisticated algorithms have been developed to determine the conductor trajectories which minimize strain energy [42]. These algorithms are coupled with electromagnetic computations to minimize field distortions. SSC and LHC magnets use precisely machined end spacers, designed by the optimization programs, which are positioned between conductor blocks [43]. In addition, the iron yoke does not extend over the coil ends to reduce the field on the conductors.

4.4 Sagitta

To limit the number of coil ends and of magnet interconnects around the accelerator ring, the arc dipole and quadrupole magnets are made as long as possible. The circulation of a charged beam in a dipole magnet, of magnetic length, \( l_d \), results in an angular deflection, \( \phi \), of the particle trajectory which can be estimated as

\[
\phi = \frac{0.3 \ q \ B_1 \ l_d}{p} \tag{26}
\]

Here, \( \phi \) is in radians and \( l_d \) is in meters, \( B_1 \) is the dipole magnetic flux density in teslas, \( q \) is the particle charge in units of electron charge, and \( p \) is the particle momentum in GeV/c.

As a result, long dipole magnets must be slightly bent to accompany the particle trajectory. This bending, which is implemented in the \((\hat{x},\hat{y})\) plane, is referred to as sagitta.
5. FIELD QUALITY

5.1 Multipole expansion

Except near the short coil ends, the magnetic flux density produced in the bore of a particle accelerator magnet can be considered as two-dimensional. It is conveniently represented by the power series expansion

\[ B_y + iB_x = B_{\text{ref}} 10^{-4} \sum_{n=1}^{\infty} (b_n + ia_n) \left( \frac{z}{R_{\text{ref}}} \right)^{n-1} \text{ for } z = x + iy, |z| < R_i \]

(27)

where \( B_x \) and \( B_y \) are the \( x \)- and \( y \)-components of the magnetic flux density, \( R_{\text{ref}} \) is a reference radius representative of the maximum beam size (\( R_{\text{ref}} = 10 \) mm for SSC and LHC), \( B_{\text{ref}} \) is the absolute value of the dipole or quadrupole component at \( R_{\text{ref}} \), \( a_n \) and \( b_n \) are the dimensionless normal and skew \( 2n \)-pole coefficients expressed in so-called units and \( R_i \) is the coil inner radius. Note the presence of the \( 10^{-4} \) scale factor.

Given the symmetries of magnet assemblies, only selected normal multipole coefficients are expected to be non-zero. These allowed multipole coefficients can be tuned up by iterating on the electromagnetic design. In practice, however, non-uniformities in material properties and manufacturing errors result in symmetry violations which produce un-allowed multipole coefficients. For instance, a top/bottom asymmetry in a dipole magnet produces a non-zero skew quadrupole coefficient \( (a_2) \), while a left/right asymmetry produces a non-zero normal quadrupole coefficient \( (b_2) \). These unwanted coefficients can only be eliminated by improving material selection, tooling and assembly procedures.

5.2 Field quality requirements

On the accelerator point of view, the beam optics is primarily governed by integrated field effects over the magnet ring. The main field-quality requirements are: (1) suitable dipole-field integral and small dipole-field angle variations (the former to ensure that the integrated bending angle over the magnet ring is \( 2\pi \) and the latter to ensure that the particle trajectory is plane), (2) accurate quadrupole alignment and suitable quadrupole field integral (the former to avoid coupling of particle motions along the \( x \)- and \( y \)-axes and the latter to ensure proper focusing), and (3) small high-order multipole coefficients (to ensure large beam dynamic aperture). In the case of high-order multipole coefficients, it is customary to specify tables of mean values and standard deviations over the entire magnet population [44]. The tables of mean values are referred to as systematic multipole specifications while that of standard deviations are referred to as random multipole specifications. The specified values are all expressed at the reference radius, \( R_{\text{ref}} \).

In large machines such as SSC or LHC, the dipole and quadrupole field integrals must be controlled with a relative precision of the order of \( 10^{-3} \). The variations in dipole field angles must be kept within a few milli-radians and the tolerance on quadrupole alignment is of the order of 0.1 mm. Systematic and random multipole specifications are given up to the 18th or 20th pole and get tighter with increasing pole order, typically from a few tenths of a unit for low order coefficients to a few thousandths of a unit for higher order coefficients.

5.3 Geometric errors

5.3.1 Types of geometric errors

The specifications on multipole coefficients require that the individual conductors and the yoke surrounding the coil assembly be positioned with a very good accuracy (typically: a few hundredths of a millimeter in the two-dimensional cross-section). Improper positioning results in geometric errors that distort the central field.
The geometric errors can be classified in at least five categories: (1) errors in coil inner and outer radii and in yoke inner radius, (2) errors in coil pole angle, wedge angle and conductor angular distribution, (3) symmetry violations in coil assembly, (4) centering errors with respect to the iron yoke and (5) residual twist of magnet assembly.

5.3.2 Effects of azimuthal coil size mismatch

A common cause of geometric error is a mismatch between the azimuthal sizes of the various coils constituting a coil assembly. Such mismatch results in displacements of the coil assembly symmetry planes which produce non-zero, low-order un-allowed multipole coefficients [45]. For instance, a mismatch between the azimuthal sizes of the top and bottom coils used in a dipole-magnet coil assembly causes an upward or downward displacement of the coil parting planes which produces a non-zero skew quadrupole coefficient \(a_2\). Similarly, a systematic mismatch between the left and right sides of the coils used in a dipole magnet coil assembly causes a rotation of the coil parting planes which produces a non-zero skew sextupole coefficient \(a_3\). The effects on \(a_1\) can be limited by randomly mixing coil production, while the occurrence of a systematic \(a_2\) can only be avoided by correcting tooling.

5.4 Iron saturation

When the field in the iron yoke is less than 2 T, the relative magnetic permeability of the yoke can be considered as uniform, and the iron contribution to the central field increases linearly as a function of transport current in the magnet coil. For fields above 2 T, parts of the iron start to saturate and their relative magnetic permeability drops. As a result, the iron contribution becomes a less-than-linear function of transport current. This relative decrease in iron contribution appears as a sag in the magnet transfer function [38]. (The transfer function is defined as the ratio of \(B_{\text{ref}}\) to the transport current.) The transfer function sag can exceed a few percent in dipole magnets but is usually negligible in quadrupole magnets.

In the case of a single-aperture magnet with a symmetrical iron yoke, the saturation first occurs in the pole areas producing a positive shift in normal sextupole coefficient \(b_3\) which compensates partially the effects of pole saturation. The midplane saturation can be forced to occur sooner by punching notches (i.e., removing matter) at appropriate locations in the yoke. As an illustration, Fig. (9) presents measurements of \(b_3\) as a function of current in the central part of a SSC dipole magnet prototype. The measurements clearly show the effect of pole saturation at high currents (the origin of the hysteresis is explained in the next section). In the case of a twin-aperture dipole, the central part of the yoke saturates before the outer parts, resulting in left/right asymmetries in the yoke contributions which affect the normal quadrupole coefficient \(b_2\). The saturation effects in \(b_2\) are of opposite sign in the two apertures.

In any case, the iron contribution depends on the packing factor of the yoke laminations which must be tightly controlled over the magnet length. Also, the iron yoke must be carefully aligned to limit magnet assembly twist.
5.5 Superconductor magnetization

5.5.1 Critical-state model

According to the so-called critical-state model, bipolar magnetization currents are induced at the periphery of the superconducting filaments in the cable strands each time the field to which the filaments are exposed is varied [46]. The magnetization currents distribute themselves with a density equal to the superconductor critical current density at the given temperature and field, $J_C$, in order to screen the filament cores from the applied field change. Unlike regular eddy currents, the magnetization currents do not depend on the rate of field variations. Also, because they can flow with zero resistance, they do not decay as soon as the field ramp is stopped. They are called persistent magnetization currents.

5.5.2 Effects of superconductor magnetization

In an accelerator magnet cycled in current, the bipolar shells of magnetization currents induced in the filaments behave as small magnetic moments which contribute to — and distort — the central field. The magnetic moments depend on $J_C$ and are proportional to filament diameter. Their distribution follows the symmetries of the transport-current field (i.e., the field produced by the transport current in the magnet coil) and, if the superconductor properties are uniform, only the allowed multipole coefficients are affected. Computer models have been developed which can accurately predict the field distortions resulting from superconductor magnetization [47].

The field distortions are the most significant at low transport current, where the transport-current field is low and $J_C$ is large. They are progressively overcome as the transport-current field increases and $J_C$ diminishes and become negligible at high transport current. They change sign and regain influence as the transport current is ramped down. As a result, the allowed multipole coefficients exhibit sizable hysteresis as a function of transport current, which depend on magnet excitation history. This is illustrated in Fig. (9) which shows measurements of $b_3$ as a function of current in the central part of a SSC dipole magnet (as explained in the previous section, the distortions at high field result from iron yoke saturation).

The field distortions resulting from superconductor magnetization are one of the major drawbacks of using superconducting magnets in a particle accelerator. They can be reduced by reducing filament size (typically, to 5 μm for SSC and LHC strands), but they cannot be
eliminated. The powering cycle of the magnets must be adapted to avoid brutal jumps between the two branches of the multipole coefficient hystereses while the beam circulates. Also, elaborate beam optics correction schemes must be developed, which can include superconducting, high-order multipole magnets (Chapter 9 of Reference [2]).

5.5.3 Time decay

In addition, the effects of superconductor magnetization are not indefinitely persistent, but exhibit a slow time decay, which, at low transport current, can result in significant drifts of the allowed multipole coefficients [48, 49]. These drifts are particularly disturbing during the injection phase of machine operation, where the magnet current is maintained at a constant and low level for some period of time [50]. Also, they complicate the early stages of acceleration, for, as the current is increased at the end of injection, the drifting multipoles snap-back rapidly to values on the hysteresis curves [51]. Part of the observed time decay can be attributed to flux creep in the superconductor [52], but flux creep cannot account for the large drifts observed after a high current cycle [49]. The nature of the other mechanisms that may be involved is not well understood.

5.6 Coupling Currents

As described in the conductor section, accelerator magnet coils are wound from Rutherford-type cables, which consist of a few tens of strands twisted together and shaped into a flat, two-layer slightly keystone cable. The cable mid-thickness is smaller than twice the strand diameter, which results in strand deformation and large contact surfaces at the crossovers between the strands of the two layers. Furthermore, and as explained in the mechanical design section, the coils are pre-compressed azimuthally during magnet assembly. Large pressures are thus applied perpendicularly to the cable that keep the strands firmly in contact. The large contact surfaces and the high pressures can result in low contact resistances at the strand crossovers.

In the steady state, the transport current flows in the superconducting filaments which offer no resistance. When the cable is subjected to a transverse varying field, the network of low interstrand resistances allow the formation of current loops which are superimposed on the transport current. The loop currents, referred to as *interstrand coupling currents*, circulate along the superconducting filaments and cross-over from strand to strand through the interstrand resistances. Unlike persistent magnetization currents, the interstrand coupling currents are directly proportional to the rate of field variations and they start to decay as soon as the field ramp is stopped.

Interstrand coupling currents have three main effects on magnet performance [39]: 1) quench current degradation (for they are superimposed on the transport current), 2) heat dissipation (when crossing the interstrand resistances), and 3) field distortions. This last issue is the most critical for accelerator magnet applications.

The coupling current contribution to the central field does not depend on transport current and increases linearly as a function of current ramp rate. If the interstrand resistance is uniform throughout the coil assembly, the coupling current distribution follows the symmetries of the transport-current field and only the allowed multipoles are affected. In practice, however, there can be large coil-to-coil differences as well as large non-uniformities within the coils themselves which result in sizable effects in the un-allowed multipole coefficients. This is illustrated in Figs. 10 (a) and 10(b) which present plots of skew and normal sextupole field coefficient ($A_3$ and $B_3$) as a function of current, measured at various ramp rates in the central part of a SSC dipole magnet prototype (note that the transport-current
contribution has been subtracted from the data). No particular treatment was applied to the strands of the cable used in this prototype.

The effects of interstrand coupling currents can be limited by ensuring that the interstrand resistances are not too low. However, and as mentioned in the conductor section, the interstrand resistances should not be too large either to allow some possibility of current redistribution among cable strands.

Fig. 10 Effects of interstrand coupling currents on multipole field coefficients as measured as a function of ramp rate in the central part of a SSC dipole magnet [39]: (a) skew sextupole field coefficient \( A_3 \) and (b) normal sextupole field coefficient \( B_3 \). The transport-current contribution is subtracted from the data.

5.7 Longitudinal periodicity

When measuring the field along the axis of an accelerator magnet with a fine spatial resolution, all multipole coefficients appear to exhibit periodic oscillations [53, 54]. The amplitude of the oscillations vary as a function of space, transport current, excitation history and time, but the wavelength is always approximately equal to the twist pitch length of the cable used in the innermost coil layer.

The longitudinal periodic oscillations are believed to result from imbalances in the current distribution among cable strands. The current imbalances may have at least three origins: (1) non-uniformities in the properties of cable strands, (2) non-uniformities in the solder joints connecting the coils in series to the current leads and (3) large and long-lasting interstrand coupling current loops superimposed on the transport current [55]. Such current loops could be induced by spatial variations in the time-derivative of the field to which the cable is exposed as it turns around the coil ends or exits towards the current leads [56–58].

The oscillation wavelength is too short to affect beam optics but may be an issue for magnetic measurements. It is recommended that the measurements be averaged over an integer number of cable pitch lengths. Also, the slow decay of the large interstrand coupling current loops associated with these periodic oscillations may contribute to the drifts of the allowed multipole coefficients observed at low and constant transport current (see section on superconductor magnetization) [59].

6. MECHANICAL DESIGN
6.1 Support against the Lorentz force

6.1.1 Components of the Lorentz force

The high currents and fields in an accelerator magnet coil produce a large Lorentz force on the conductors. In a dipole coil, the Lorentz force has three main components which are represented Fig. 6 (38, 60): (1) an azimuthal component which tends to squeeze the coil towards the coil assembly midplane (which, in the coordinate system defined previously, corresponds to the horizontal \((\bar{x}, \bar{z})\) plane), (2) a radial component which tends to bend the coil outwardly, with a maximum displacement at the coil assembly midplane (along the horizontal \(x\)-axis), and (3) an axial component, arising from the solenoidal field produced by the conductor turnaround at the coil ends and which tends to stretch the coil outwardly (along the \(z\)-axis).

6.1.2 Stability against mechanical disturbances

Since accelerator magnets are operated close to the critical current limit of their cables, the electromagnetic works produced by minute wire motions in the coil are of the same order of magnitude as the energy depositions needed to trigger a quench [61]. If the motions are purely elastic, no heat is dissipated and the coil remains superconducting, but if the motions are frictional, the associated heat dissipation may be sufficient to initiate a quench. This leaves two possibilities: either to prevent wire or coil motions by providing a rigid support against the various components of the Lorentz force or to reduce to a minimum the friction coefficients between potentially moving parts of magnet assembly.

6.1.3 Conceptual design

The mechanical design concepts used in present accelerator magnets are more or less the same and were developed at the time of the Tevatron [4, 62]. In the radial direction: the coils are confined within a rigid cavity defined by laminated collars which are locked around the coils by means of keys or tie rods. In the azimuthal direction: the collars are assembled so as to pre-compress the coils. In the axial direction: the coils either are free to expand or are restrained by means of stiff end-plates.

The use of laminated collars pioneered at the Tevatron was a real breakthrough in achieving a rigid mechanical support while keeping tight tolerances over magnet assemblies which are a few meters in length and which must be mass-produced. The laminations are usually stamped by a fine blanking process allowing a dimensional accuracy of the order of one hundredth of a millimeter to be achieved.

6.2 Azimuthal pre-compression

6.2.1 Preventing collar pole unloading

As described above, the azimuthal component of the Lorentz force tends to squeeze the coil towards the midplane. At high fields, it may happen that the coil pole turns part from the collar poles, resulting in variations of coil pole angle which distort the central field and creating a risk of mechanical disturbances. To prevent conductor displacements, the collars are assembled and locked around the coils so as to apply an azimuthal pre-compression. The pre-compression is applied at room temperature and must be sufficient to ensure that, after cooldown and energization, there is still contact between coil pole turns and collar poles.

6.2.2 Pre-compression requirement

To determine the proper level of room temperature azimuthal pre-compression, at least three effects must be taken into account: (1) stress relaxation and insulation creep following
the collaring operation, (2) thermal shrinkage differentials between coil and collars during cooldown (if any) and (3) stress redistribution due to the azimuthal component of the Lorentz force. In addition, the collaring procedure must be optimized to ensure that the peak pressure seen by the coils during the operation (which may be significantly higher than the residual pre-compression) does not overstress insulation (p. 1326 of Reference [60]).

The pre-compression loss during cooldown, $\Delta \sigma$, can be estimated from

$$\Delta \sigma \hat{A} E_{\text{coil}} (\alpha_{\text{coil}} - \alpha_{\text{collar}})$$

where $E_{\text{coil}}$ is the coil Young's modulus in the azimuthal direction, and $\alpha_{\text{coil}}$ and $\alpha_{\text{collar}}$ are the thermal expansion coefficients of the coil (in the azimuthal direction) and of the collars, integrated between room and operating temperatures.

6.2.3 Choice of collar material

To limit cooldown loss, it is preferable to use for the collars a material whose integrated thermal expansion coefficient matches more or less that of the coil. For NbTi coils with polyimide/glass or all-polyimide insulation, this suggests the use of aluminum alloy (see Table 2). However, and as will be described in the next section, it is also desirable that the collars be as rigid as possible or have an integrated thermal expansion coefficient approaching that of low carbon steel. This favors the use of austenitic stainless steel, which has a lower integrated thermal expansion coefficient and whose Young's modulus is 195 GPa at room temperature and 203 GPa at 4.2 K, compared to 72 GPa at room temperature and 80 GPa at 4.2 K for aluminum alloy.

When assessing the respective merits of austenitic stainless steel and aluminum alloy, it should be noted that austenitic stainless steel presents a better resistance to stress cycling at low temperature [63], but that it has a higher density (7800 kg/m$^3$ compared to 2800 kg/m$^3$ for aluminum alloy) and that it is more expensive.

There is no ideal solution between stainless steel and aluminum alloy and magnets with both types of collar materials have been built: HERA dipole magnets and most LHC dipole magnet prototypes use aluminum alloy collars while Tevatron dipole magnets and most SSC dipole magnet prototypes rely on stainless steel collars. In any case, and whichever collar material is chosen, a thorough mechanical analysis of the structure under the various loading conditions is required.

6.3 Radial support

6.3.1 Limiting radial deflections

As described above, the radial component of the Lorentz force tends to bend the coil outwardly, with a maximum displacement at the coil assembly midplane. At high fields, this bending results in shear stresses between coil turns and in an ovalization of the coil assembly along the horizontal $x$-axis which generates field distortions. To prevent displacements or deformations, the radial deflections of the coil assembly must be limited to, typically, less than 0.05 mm.

6.3.2 Seeking yoke support

The main support against the radial component of the Lorentz force is provided by the collars, whose stiffness and radial width must be optimized to limit collared-coil assembly deflections. However, in the magnetic design of high field magnets, the field enhancement provided by the iron yoke is maximized by bringing it as close as possible to the coil. This reduces the space left for the collars, whose rigidity then becomes insufficient to hold the
Lorentz force. In such magnets, the yoke and helium containment shell must also be used as part of the coil support system.

The mechanical design of magnets where the yoke is needed to support the collared-coil assembly is complicated by the fact that the collar material (stainless steel or aluminum) shrinks more during cooldown than the low-carbon steel used for the yoke (see Table 2). This thermal shrinkage differential must be compensated to ensure that, when the magnet is cold and energized, there is a proper contact between the collared-coil assembly and the yoke along the horizontal \(x\)-axis.

The aforementioned thermal shrinkage differential, \(\Delta r\), can be estimated as

\[
\Delta r = R_{\text{collar}} (\alpha_{\text{collar}} - \alpha_{\text{yoke}})
\]

where \(R_{\text{collar}}\) is the collar outer radius and \(\alpha_{\text{yoke}}\) is the thermal expansion coefficient of the yoke, integrated between room and operating temperatures.

To limit contact loss along the horizontal diameter it is preferable to use for the collars a material whose integrated thermal expansion coefficient approaches that of low-carbon steel. This suggests the use of austenitic stainless steel (see Table 2). However, and as was described in the previous section, it is also desirable to limit the cooldown loss of coil pre-compression, which favors the use of aluminum alloy.

**Table 2**

Integrated thermal expansion coefficients between 4.2 K and room temperature (10\(^{-3}\) m/m)

<table>
<thead>
<tr>
<th>Material</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-carbon steel</td>
<td>2.0</td>
</tr>
<tr>
<td>Stainless steel (304/316)</td>
<td>2.9</td>
</tr>
<tr>
<td>Copper (OFHC)</td>
<td>3.1</td>
</tr>
<tr>
<td>Aluminium</td>
<td>4.2</td>
</tr>
<tr>
<td>Insulated cable (polyimide/glass)</td>
<td>5.1 a)</td>
</tr>
<tr>
<td>Insulated cable (all-polyimide)</td>
<td>5.6 a)</td>
</tr>
</tbody>
</table>

\(a)\) transverse direction; design specific

### 6.3.3 Mechanical design with positive collar/yoke interference

If the thermal shrinkage differential between collar and yoke is not too large (as in the case of stainless-steel collars), it can be compensated by introducing a positive collar-yoke interference at room temperature. The axis along which this interference is introduced drives the choice of yoke split orientation: the SSC dipole magnet prototypes built at BNL use a horizontally-split yoke with a positive collar-yoke interference along the vertical \(y\)-axis as shown in Fig. 11(a), while the SSC dipole magnet prototypes built at FNAL use a vertically-split yoke with a positive collar-yoke interference along the horizontal \(x\)-axis as shown in Fig. 11(b) [64]. Both types of magnets performed very well.
6.3.4 Mechanical design with yoke midplane gap

For large thermal shrinkage differentials (as in the case of aluminum collars), the required positive collar-yoke interference at room temperature would over stress the collared-coil assembly and a different mechanical design must be used. The twin-aperture LHC dipole magnet prototypes with aluminum collars rely on a vertically-split yoke with an open gap at room temperature and a welded outer shell made of a material (stainless steel or aluminum) that shrinks more during cooldown than the low carbon steel yoke [65].

In these magnets, the yoke is designed so that, when placed around the collared-coil assembly at room temperature with no pressure applied to it, there remains an opening between the two yoke halves of the order of the expected thermal shrinkage differential. Furthermore, the outer shell is designed so as to apply on the yoke halves a compressive load which forces a progressive closing of the yoke midplane gap during cooldown. This compressive load arises from weld shrinkage at room temperature and from thermal shrinkage differential between yoke and shell during cooldown. As a result, the two yoke halves follow the shrinkage of the collared-coil assembly and maintain a contact along its horizontal diameter.

A crucial issue in such a design is the ability of keeping a tight tolerance (of the order of 0.1 mm) on the yoke midplane gap during magnet production (for a gap too close may result in coil overstressing while a gap too open may result in contact loss during cooldown). In some LHC prototypes, the yoke midplane gap is controlled by means of aluminum spacers located between the two yoke halves [66]. The spacers are dimensioned to have a spring rate similar to that of the collared-coil assembly and they prevent the gap from closing at room temperature. During cooldown, however, they shrink more than the yoke and cease to be effective.

6.3.5 RHIC magnets

In RHIC magnets, collar and yoke designs are altogether simplified by replacing the collars by reinforced plastic spacers and by using directly the yoke to pre-compress the one-layer coils [67]. It remains to be seen if this structure could be scaled-up to higher field magnets.
6.4 End support

As described above, the axial component of the Lorentz force tends to stretch the coil outwardly along the z-axis. In magnets where the yoke is not needed to support the collared-coil assembly, a clearance can be left between the two. If the axial stresses resulting from the Lorentz force do not exceed the yield stress, it is possible to let the collared-coil assembly free to expand within the iron yoke. This is the case of the quadrupole magnets designed at CEA/Saclay for HERA, SSC and LHC [68]. However, in magnets where there is contact between collar and yoke, it is essential to prevent stick/slip motions of the laminated collars against the laminated yoke and to provide a stiff support against the axial component of the Lorentz force [60, 69]. The ends of SSC and LHC dipole magnet coils are contained by thick stainless steel end plates welded to the shell.

7. MAGNET COOLING

7.1 Superconductor critical temperature

The superconducting state only exists at temperatures below the so-called critical temperature, $T_C$. For NbTi, $T_C$ can be estimated as a function of applied magnetic flux density, $B$, using

$$T_C = T_{C0} \left(1 - \frac{B}{B_{C20}}\right)^{1.7}$$

where $T_{C0}$ is the critical temperature at zero field (about 9.2 K) and $B_{C20}$ is the upper critical magnetic flux density at zero temperature (about 14.5 T).

7.2 Boiling and supercritical helium cooling

To achieve low temperatures and ensure stable operation against thermal disturbances, the accelerator magnet coils are immersed in liquid helium. Helium is a cryogenic fluid whose pressure-temperature phase diagram is presented in Fig. 12 [70]. Its boiling temperature is 4.22 K at 1 atmosphere (1 atmosphere $\approx$ 0.1 MPa).

Small superconducting magnet systems usually rely on boiling helium at 1 atmosphere [71]. Using boiling helium offers the advantage that, as long as the two phases are present, the temperature is well determined. However, in large scale applications, such as superconducting particle accelerators, the fluid is forced to flow through numerous magnet cryostats and long cryogenic lines, where heat leaks are unavoidable. The heat leaks result in increases in vapor contents and create a risk of gas pocket formation that may block circulation.

The aforementioned difficulty can be circumvented by taking advantage of the fact that helium exhibits a critical point at a temperature of 5.2 K and a pressure of 0.226 MPa (see Fig. 12). For temperatures and pressures beyond the critical point, the liquid and vapor phases become indistinguishable. The single-phase fluid, which is called supercritical, can be handled in a large system without risk of forming gas pockets. However, its temperature, unlike that of boiling helium, is not constant and may fluctuate as the fluid circulates and is subjected to heat losses.
The cryogenic systems of Tevatron, HERA, and RHIC combine single-phase and two-phase helium [71]. In the case of Tevatron and HERA, the inside of the magnet cold masses are cooled by a forced flow of supercritical helium while two-phase helium is circulated in a pipe running at the cold mass periphery (around the collared-coil assembly for Tevatron magnets, in a bypass hole in the iron yoke for HERA magnets). In the case of RHIC, only supercritical helium is circulated through the magnet cold masses, while so-called re-coolers, consisting of heat exchangers using two-phase helium as primary fluid, are implemented at regular intervals along the cryogenic lines. In these systems, the boiling liquid is used to limit temperature rises in the single-phase fluid.

7.3 Superfluid-helium cooling

A particularity of helium is the occurrence of superfluidity [70]. When cooling down boiling helium at 1 atmosphere, it stays liquid until a temperature of the order of 2.17 K, where there appears a phase transition. For temperatures below 2.17 K (at 1 atmosphere) helium looses its viscosity and becomes a superconductor of heat. This property, unique to helium, is called superfluidity. Superfluidity is very similar to superconductivity, except that, instead of electrical conductivity, it is the thermal conductivity that becomes infinite. The transition temperature between the liquid and superfluid phases depends on pressure. It is called the lambda-temperature, $T_\lambda$.

The LHC magnets are cooled by superfluid helium and their operating temperature is set at 1.9 K [72]. Decreasing the temperature improves dramatically the current carrying capability of NbTi and allows higher fields to be reached. (For NbTi, the curve "critical current density as a function of field" is shifted by a about +3 T when lowering the temperature from 4.2 K to 1.9 K.) The feasibility of large scale cryogenic installation relying on superfluid helium has been demonstrated by Tore Supra, a superconducting tokamak built at CEA/Cadarache (Commissariat à l’Energie Atomique at Cadarache near Aix en Provence in the South of France) and operating since 1988 [73].

7.4 Magnet cryostat

To maintain the magnet cold masses at low temperature it is necessary to limit heat losses. There are three mechanisms of heat transfer [74]: (1) convection, (2) radiation and (3) conduction. The convection losses are eliminated by mounting the cold masses into cryostats which are evacuated [71, 75]. The radiation losses, which vary like the fourth power of wall temperature, are reduced by surrounding the cold masses with thermal shields at intermediate temperatures. The main sources of conduction losses are the support posts, the power leads and the cryogenic feedthroughs which must be designed to present high thermal resistances.
8. QUENCH PERFORMANCE

As explained in the operating marging section, the maximum quench current, \( I_{qm} \), of a magnet at a given operating temperature can be estimated from the critical current of the cable and the peak field on the coil. It corresponds to the ultimate current carrying capability of the cable and can only be raised by lowering the operating temperature.

When energizing a superconducting magnet, the first quenches usually occur at currents below \( I_{qm} \) (Chapter 5 of Reference [76]). In most cases, however, it appears that, upon successive energizations, the quench currents gradually increase. This gradual improvement is called the magnet's training. The training often leads to a stable plateau corresponding to the maximum quench current.

Quenches below the expected maximum quench current have at least four origins: (1) energy deposition in the magnet coil resulting from frictional motions under the Lorentz force, (2) energy deposition from beam losses, (3) heat dissipation from coupling currents in the cable and (4) current imbalances among cable strands. Quenches of the first origin are revealing of flaws in the mechanical design or in the assembly procedures which must be analyzed and corrected. The effects of beam losses can be reduced by implementing an intercepting screen within the beam tube. Coupling losses and current imbalances are only of concern for fast current cycles.

When operating an accelerator made of several hundreds or even several thousands of superconducting magnets, it cannot be tolerated that magnets quench at random. Hence, the magnets must be designed with a safe margin above the maximum operating current of the machine. In addition, systematic tests must be carried out before installing the magnets in the tunnel to ensure that their quench performance is adequate and does not degrade upon extended current and thermal cycling [77].

9. QUENCH PROTECTION

9.1 The effects of a quench

Although most R&D programs have been successful in developing magnet designs that can be mass-produced and meet accelerator requirements, quenches do occur in accelerator operations. These quenches must be handled in order to avoid any damage of the quenching magnet, to ensure the safety of the installation and to minimize down time.

The most damaging effect of a quench is that, once a volume of conductor has switched to the normal resistive state, it dissipates power by the Joule effect (Chapter 9 of Reference [76]). Most of this power is consumed locally in heating the conductor. In a very short time (typically: a few tenths of a second), the conductor temperature can reach room temperature, and, if the magnet is not discharged, keeps on increasing.

To discharge a quenching magnet, all its stored magnetic energy must be dissipated into resistive power. If the quench propagates very slowly, there is a risk that a large fraction of the stored energy be dissipated in a small volume of conductor. In the case of a string of magnets connected electrically in series, it may even happen that the energy of the whole string be dissipated in the quenching magnet. Hence, to prevent burnout, it is necessary to ensure that the normal resistive zone spreads rapidly throughout the quenching coil. This can be done by means of heaters, implemented near the magnet coils and fired as soon as a quench is detected. These heaters are referred to as quench protection heaters.
In comparison to other superconducting magnets, storage ring magnets do require an active quench protection system because of the rapidity of the temperature rise resulting from the high current density and the low fraction of stabilizing copper in the cable strands.

9.2 Conductor heating

9.2.1 Maximum temperature requirement

The temperature rise consecutive to a quench must be limited for at least three reasons: (1) to restrict the thermal stresses induced in the quenching coil, (2) to prevent degradation of superconductor properties, and (3) to avoid insulation damage.

For most materials, thermal expansion starts to be significant for temperatures above 100 K. The critical current density of NbTi is affected by exposure to temperatures above 250 °C. The degradation amplitude depends on the temperature level and on the duration of the exposition: at 250 °C, it takes of the order of 1 hour to get a significant degradation, while it may take less than a minute at 400–450 °C [78]. Finally, the polyimide materials used to insulate NbTi cables loose most of their mechanical properties for temperatures above 500 °C. It follows that an upper limit for conductor heating consecutively to a quench is 400 °C. Most magnets are designed not to exceed 300 to 400 K, and whenever possible, the limit should be set to 100 K.

9.2.2 Estimating hot spot temperature

The volume of conductor that heats up the most significantly during a quench is the spot where the quench first originated. It is called the hot spot. An upper limit of the hot spot temperature, $T_{\text{max}}$, can be determined by assuming that, near the hot spot, all the power dissipated by the Joule effect is used to heat up the conductor. Integrating the heat balance equation yields

$$S^2 \int_{T_0}^{T_p} \frac{dT}{C(T)} \frac{1}{\rho(T)} = \int_{t_0}^{+} dt I(t)^2$$

(31)

where $C$ is the overall specific heat per unit volume of conductor, $\rho$ is the overall conductor resistivity in the normal state, $S$ is the conductor cross-sectional area, $I$ is the current, $T_0$ is the coil temperature before quench and $t_0$ is the time of quench start.

The left member of Eq. (31) depends only on conductor properties while the right member depends only on the characteristics of current decay. The right-hand side integral, divided by $10^6$, is called the MIIT integral (Mega I times I versus Time integral) and its value is referred to as number of MIITs. The maximum temperatures computed from the numbers of MIITs have been shown to be in fairly good agreement with actual measurements of hot spot temperatures on quenching magnets [79].

9.2.3 Limiting hot spot temperature

The hot spot temperature can be limited by acting on either member of Eq. (31). Regarding the left member, the only conceivable action is to reduce the overall conductor resistivity by increasing the copper-to-superconductor ratio. However, and as explained in the conductor section, the copper-to-superconductor ratio must also be optimized to ensure a high overall critical current. Regarding the right member, the MIIT integral can be minimized by: (1) detecting the quench as soon as possible, (2) turning off the power supply (case of a single
magnet) or forcing the current to bypass the quenching magnet (case of a magnet string),
(3) firing the quench protection heaters and (4) discharging the quenching magnet or the
magnet string.

9.3 Quench detection

The magnets are connected to quench detection systems which monitor the occurrence
of a resistive voltage in the coil windings or the coils leads. The resistive voltage has to be
discriminated from inductive voltages arising from magnet ramping. The inductive
components are canceled out by considering voltage differences across two identical coil
assemblies or two identical parts of a given coil assembly (e.g., the upper and lower half coils
in a dipole magnet). When the resistive voltage exceeds a preset threshold over a time
exceeding a preset duration, the detection system generates a trigger which signals the
occurrence of a quench.

9.4 Protection of a single magnet

9.4.1 Time constant of current discharge

Let us first consider the case of a single magnet. Once a quench is detected, the power
supply is turned off and the magnet is switched to an external dump resistor, $R_{\text{ext}}$. The time
constant of the current decay, $\tau$, is given by

$$\tau = \frac{L_m}{R_q(t) + R_{\text{ext}}}$$

where $L_m$ is the magnet inductance and $R_q(t)$ is the developing resistance in the quenching
coils. Furthermore, the total voltage across the magnet, $V_m$, is given by

$$V_m = R_{\text{ext}} I(t)$$

where $I$ is the current intensity.

To limit the number of MIITs, it is desirable to have a short time constant. Equation
(32) shows that small $\tau$ values are obtained either by means of a large $R_{\text{ext}}$ or by ensuring that
$R_q(t)$ increases rapidly. For some magnets, an external resistor can be used to extract a
significant fraction of the stored magnetic energy. However, it is also required to keep $V_m$ to
a reasonable level (typically: less than 1 kV) to avoid insulation breakdown. Given the order
of magnitude of $I$ (up to 15 kA), this imposes a small $R_{\text{ext}}$ (typically: a few hundredth of
ohms) which, in case of a quench, is soon overcome by $R_q(t)$. Hence, in storage ring magnets,
the current decay is largely dominated by the resistance development in the quenching coils
and the time constant can only be reduced by speeding up $R_q(t)$.

9.4.2 Maximum voltage to ground

The developing resistance in the quenching coil separates the coil impedance into
several parts (p. 137 of Reference [2]): un-quenched parts across which the voltage is mainly
inductive and quenched parts across which the voltage is mainly resistive. The resistive and
inductive voltages compensate each other partially so that their sum equals $V_m$. The voltage
distribution with respect to ground depends on the respective sizes and locations of these
various parts. The more uniform the quench development, the lower the maximum voltage to
ground. As an illustration, Fig. 13 shows the voltage distribution in a quenching magnet.
Here, $V_m$ is assumed to be nil and $R_q$ is assumed to be concentrated at about two-thirds of the
magnet length.
9.4.3 Quench protection heaters

As described earlier, to speed up and uniformize quench development, accelerator magnets rely on quench protection heaters which are fired as soon as a quench is detected. The heaters are usually made of stainless-steel strips, which are copper clad at regular intervals along their lengths and which are placed on the outer surface of the coil assemblies. Note, however, that the heater firing unit relies on a capacitor bank and that it takes some time for the energy to be released. Note also that the heaters have to be electrically insulated from the coil and that this electrical insulation introduces a thermal barrier. As a result, there is a non-negligible delay between the firing of the heaters and its effect on the coils, during which, one has to rely on natural quench propagation [80]. The heaters and their implementations in the magnet assembly are optimized to reduce this delay.

9.5 Protection of a magnet string

In an accelerator, the magnet ring is divided into several sectors constituted of series-connected magnets. The sectors are powered independently and are electrically independent. Once a quench is detected in a magnet, the power supply of the sector to which the magnet belongs is turned off and the sector is discharged over a dump resistor.

Unlike in the case of a single magnet, the current decay rate in the sector must be limited for at least two reasons: (1) prevent the induction of large coupling currents in the magnet coils (which may quench the remaining magnets in the sector, resulting in global warming and significant helium venting), and (2) avoid the occurrence of unacceptable voltages to ground (because of the large overall inductance of the sector). A too slow decay rate, however, creates the risk that a significant fraction of the total energy stored in the sector be dissipated in the quenching magnet, resulting in destructive overheating.

These contradictory considerations can be conciliated by forcing the current to bypass the quenching magnet and by ramping the current down at the desired rate in the remaining un-quenched magnets. The bypass elements consist of diodes (or thyristors) connected in parallel to individual or small groups of magnets, as shown in Fig. 14. As long as the magnets are superconducting, the current flows through the magnets. Once a magnet has quenched and starts to develop a resistive voltage, the main current is bypassed through the diode connected in parallel and the quenching magnet is discharged over the diode circuit. The time constant of the discharge is similar to that given by Eq. (32), except that $R_{\text{ext}}$ has to be replaced by the resistance associated with the bypass element, $R_b$. 

Fig. 13 Voltage distribution in a quenching magnet. The total voltage across the magnet is assumed to be nil and the developing resistance is assumed to be concentrated at about two third of the magnet length [2].
HERA, RHIC and LHC rely on silicon diodes which are mounted inside the helium cryostats and operate at cryogenic temperatures. The main requirements for these cold diodes are [81]: (1) small forward voltage and low dynamic resistance (to limit power dissipation in the diodes), (2) good radiation hardness and (3) large backward voltage. In the case of Tevatron, which has a short current ramp time resulting in large inductive voltages across the bypass elements, the diodes are replaced by thyristors operating as fast switches [82]. The thyristors are located outside the magnet cryostats and require additional power leads and cryogenic feedthroughs.

The protection system of the magnet ring must be carefully designed and thoroughly tested before starting up the machine. The system tests are usually carried out on a cell or a half-cell representative of the magnet lattice and all failure modes are investigated [83–85].

10. BRIEF SUMMARY

As of today, two large superconducting storage rings, Tevatron and HERA, have been built and are reliably operating, and work is under way on two other superconducting colliders: RHIC and LHC. The construction of RHIC is near completion and the industrial contracts for the mass production of LHC magnets will soon be awarded.

Since the time of Tevatron, a factor of about two has been gained on the critical current density of NbTi at 4.2 K and 5 T and a dipole field of 10.5 T has been reached on a short model magnet relying on NbTi cables at 1.8 K. In the last years, encouraging results have been obtained on a couple of short dipole magnet models relying on Nb$_3$Sn cables, which may open the 10 to 15 T range.

ACKNOWLEDGMENTS

The author wishes to thank L. Bottura, H. Desportes, J.M. Rifflet and F. Simon for their comments on the manuscript.
REFERENCES


[72] Ph. Lebrun, Superfluid helium cryogenics for the large hadron collider project at CERN, Cryogenics, 34 ICEC Supplement: 1–8, 1994.


FIELD DYNAMICS IN SUPERCONDUCTING MAGNETS FOR PARTICLE ACCELERATORS

L. Bottura
CERN, Switzerland

Abstract
The field quality in superconducting magnets for particle accelerators shows significant dependence on ramp-rate and powering history. The main effects are outlined, based on measurement results, and the basic understanding of the effects are discussed. The dynamic characteristics bear implications for the magnetic measurement techniques and data treatment.

1. INTRODUCTION

Magnetic fields in particle accelerator magnets are generated over a very large dynamic range. At the extreme of long time scales we have superconducting dipole magnets that operate with large charge-up times, in the range of tens of seconds and up to steady-state. The other extreme, very short time scales, is well represented by the fast kicker magnets that generate fields over time scales comparable to the spacing between bunches in a particle beam, in the range of some tens of ns. Several effects come into play over such a wide span of time scales, covering different physical and engineering aspects such as eddy currents in conducting parts, magnet inductance and capacitance, power dissipation, power supply design and tuning, current distribution and skin effects in cables, superconductor AC loss and stability, and, not least, field quality. This incomplete list just hints at the complexity of the general problem of magnet design and operation in conjunction with field dynamics.

This chapter is limited to a portion only of the domain sketched above, namely the main aspects of field dynamics and its measurement in connection with the use of superconducting magnets. The motivation is that superconducting magnets are becoming more and more common in large size and high energy particle accelerators. Examples of working superconducting accelerators are the proton-antiproton collider Tevatron [1] and the proton-lepton collider HERA [2] at DESY. At present (1997) the Relativistic Heavy Ion Collider (RHIC) [3] is completing installation and will soon come into operation, while the Large Hadron Collider (LHC) [4] is in the prototyping phase and will soon start construction. As the accelerator performance becomes more demanding, and the design capability evolves, the importance of dynamic effects and their control in superconducting magnets grows.

Bending (dipole) and focusing (quadrupole) magnets in accelerators are generally operated between a low field level, at which particle injection takes place, and the coast flat-top, after the acceleration of the beam to its nominal energy. It is common practice to condition the magnets by means of a precycle procedure, aimed at bringing the magnet into reproducible conditions. The magnets are ramped between the injection level and the coast flat-top. As we will see, significant dynamic behaviour in superconducting magnets appears both during the ramps and during constant current plateaux. We will start in Section 2 subdividing the components of the magnetic field according to the different steady-state or dynamic origin. Sections 3, 4 and 5 will be dedicated to the phenomenology of dynamic fields in superconducting magnets going for each effect from the discovery to their present understanding. Section 6 will briefly list dynamic effects not related to superconductors.
Finally, in Section 7 we discuss the peculiarities of measurement of time variable fields in accelerator magnets. Appendix 1 is dedicated to the magnetic field formalism, while Appendix 2 gives some details of the behaviour of superconducting cables in a variable magnetic field. The treatment there is cursory because of obvious limitations, and we must refer to the literature quoted for a deeper insight.

2. A CATALOGUE OF FIELD COMPONENTS

A systematic approach to the magnetic field analysis in an accelerator magnet is to break the total field generated in the bore into its components of different origin. Following commonly accepted practice [5-8], we can identify for steady-state operation the following components:

- **geometric**, related to the cable positions in the winding pack, the accuracy of their placement and movements during energization;
- **iron magnetisation**, accounting for the magnetisation and saturation of the iron yoke as a function of the excitation field;
- **persistent currents magnetisation** of the superconducting filaments.

Of the components above, the first two (geometric and iron magnetisation) do not exhibit *dynamic* behaviour. They are proportional to the excitation current. The third, the persistent current magnetisation, is in principle also of steady-state nature, but can show a long-term variation as will be discussed later on. Most important, the persistent currents magnetisation has a large hysteresis that appears as a difference in the ramp-up and ramp-down branches of the magnet loadline. These three contributions, in the case of a *perfect* magnet, will appear only on *allowed* harmonics, i.e. those permitted by the symmetry conditions of the coil. As discussed in Appendix 1, this is evident for symmetric geometry. If the iron geometry and superconductor properties also respect the magnet symmetry conditions, the resulting magnetisations, both for iron and superconductor, will have the same degree of symmetry and thus only contribute to allowed harmonics.

The steady-state properties of the field components are not the object of this chapter. Still they are discussed here because, as we will see later on, all dynamic effects will cause a deviation from this *ideal* situation, either through additional allowed harmonics, or because of the appearance of non-allowed harmonics. An example of the three different contributions in steady-state can be clearly seen in the measurement of normal sextupole in a superconducting dipole as shown in Fig. 1. We have plotted there the normal sextupole component as measured in steady-state conditions at different levels of current during the ramp-up and ramp-down in an LHC dipole. The average value of the normalised sextupole for ramp-up and -down is constant for fields below approximately 6 T, according to the linear contribution associated to the winding geometry. Above 6 T we see that the average of the curves deviates from the initial constant owing to the iron saturation. Finally, the superconductor magnetisation is responsible for the hysteresis in the two curves, also clearly showing the field dependence of the magnetisation. In the next sections we will add three additional effects of dynamic nature:

- **coupling currents** (ramp-rate dependent) magnetisation in the superconducting strands and cables;
- **field periodicity**, related to the current distribution in the superconducting cable;
- **field drift** during constant current plateaux.
3. COUPLING CURRENT EFFECTS DURING RAMPS

As discussed in Appendix 2, eddy currents are induced in superconducting strands and cables in the presence of a changing magnetic field. The eddy currents tend to flow in the superconductor, where they find virtually zero resistance, and cross over along the minimum resistance path in the copper matrix of a strand, or in the contact points between strands in a cable. These eddy currents are the result of the electromagnetic coupling between filaments or strands, and are often referred to as coupling currents in the superconducting strand or cable. As to the magnetic field in the bore, and for common accelerator operation with ramp times much larger than the coupling currents time constant, we expect in first approximation a cable magnetic moment proportional to the field ramp-rate and inversely proportional to the transverse strand and cable resistances (see Appendix 2 for details).

The problem of the field distortion produced by coupling currents magnetisation was tackled already during the ISABELLE project, when Courant [9] derived an analytical approximation for the harmonics generated by cable coupling currents under uniform ramp-rate in the single-layer, superconducting dipoles and quadrupoles. Coupling currents magnetisation is not an issue for the normal operation of accelerators like the Tevatron or HERA, either because of the large interstrand resistance in the cable or because of moderate ramp-rates, both resulting in small coupling currents magnetisation effects. The problem was not addressed extensively during the design and manufacturing of either accelerators.

For the Superconducting Super-Collider (SSC) ramped-field distortions associated to coupling currents became again an issue, mainly because of the operating requirements on the High Energy Booster (HEB) ring. The SSC - HEB needed fast cycling times, and for this reason the SSC prototype magnets were measured systematically sweeping the field with different ramp-rates. A typical result obtained from this type of measurements is shown in Fig. 2. There we show the normal sextupole in an SSC dipole prototype DCA312, as obtained with a rotating coil measurement on-the-fly (as the field is changing). As compared to the steady-state conditions we see that the normal sextupole hysteresis amplitude depends now on the ramp-rate. We see here the first feature of ramped operation, namely that both ramp-up and ramp-down branches are displaced by an amount approximately proportional to the ramp-rate. As shown in Fig. 2, if the harmonic coefficients are plotted in absolute terms (instead of normalised units, as in Fig. 1) the amount of displacement is indeed a constant for a given ramp-rate over the whole field range.
This can also be shown by plotting the amplitude of the hysteresis cycle as a function of the ramp-rate, as in Fig. 3 for the normal sextupole in the same magnet, measured at three different positions along the magnet length. The hysteresis amplitude varies linearly with the ramp-rate. In all the positions the intercept (the steady-state hysteresis amplitude) is the same, as we would expect in a magnet with uniform superconductor properties in longitudinal direction. However we see here a second feature of ramped operation, namely that the ramp-rate dependence of the amplitude of the hysteresis cycle can be a function of the longitudinal position along the magnet.

The picture is however not complete. In fact, an additional discovery from swept measurements of SSC dipoles was that several magnets could show a ramp-rate dependent hysteresis in the non-allowed harmonics. This hysteresis, again with an amplitude proportional to the magnet ramp-rate if given in absolute terms, disappeared in steady-state conditions. An example of this anomalous behaviour is shown in Fig. 4, where we report the skew sextupole measurement in the same SSC dipole DCA312 mentioned previously.

As shown by several authors [7, 10–13], the ramp-rate dependent hysteresis is mainly generated by the cable magnetisation due to coupling currents associated to a field change normal to the broad face of the cable. As mentioned above, and discussed more extensively in Appendix 2, this magnetisation component is proportional to the field change rate, so that the resulting contribution to the harmonics is proportional to the magnet ramp-rate as we have
indeed observed. In addition the coupling currents magnetisation is inversely proportional to the transverse resistance of the cable. This last however is known to depend critically on several factors, among them the surface conditions of the strands, their ageing, heat treatment conditions during coil fabrication, and possibly the electromagnetic pressure on the strand contacts at operation. Hence in a coil we can have an arbitrary distribution of interstrand resistances along the magnet length and within the winding cross section. This distribution does not necessarily respect the geometrical symmetries. The consequence is that the magnetisation can vary on a cable-by-cable basis within the winding and in principle all harmonics can be present during a ramp.

As described in [11, 14], it is possible to reconstruct the distribution of the interstrand resistance in a winding pack, based on the measurement of the ramp-rate dependent harmonics solving an inverse problem based on models analogous to those discussed in Appendix 2. Such a procedure was followed for the SSC dipole DCA312, for which we have reported the measurements in Figs. 2 through 4. The results are reported in Fig. 5, where we show the computed interstrand conductance (the inverse of the resistance) distribution that explains the allowed and non-allowed harmonics measured. The interstrand conductance obtained from the reconstruction is indeed non-uniform in the winding, showing peaks close to the midplane in the upper pole, a region which seemed particularly delicate being coincident with the location of ramp-rate related quenches. This magnet was examined in detail, measuring the interstrand resistance by direct methods, and showed a satisfactory correlation between the direct measurement and interstrand resistance reconstruction [15, 16].
4. FIELD PERIODICITY AND CURRENT DISTRIBUTION

In a longitudinal scan of the field performed in a HERA dipole using a three-axis hall probe, Brueck et al. noticed that the local value of the sextupole was periodic along the length of the magnet [17]. Figure 6 shows results of these measurements, obtained after a pre-cycle 0 A – 5.5 kA – 50 A - 250 A (curve a). The average sextupole was approximately on the hysteresis curve as would be obtained measuring with a long probe, but the local value oscillated periodically with a considerable amplitude. After a subsequent cycle 250 A – 2 kA – 250 A the average was shifted from negative to positive values according to the expected variation of the magnetisation sextupole (see also Fig. 1 for comparison), but the amplitude of the periodic pattern was practically unchanged (curve b). Similar results were obtained lowering the current to 100 A (curve c). The periodic pattern disappeared as soon as the magnet was quenched (increasing its temperature), proving that the periodicity was associated with the superconducting state.

![Figure 6 Periodicity of the normal sextupole as measured by Brueck et al. [17] in a HERA dipole. Reproduced by permission of IEEE. © 1991 IEEE.](image)

The most remarkable fact was however that the periodicity length was coincident with the cable twist pitch, $95 \pm 2$ mm in the case of the measurements shown. Indeed, this is a help in understanding the origin of this phenomenon, as we will discuss later on. Motivated by the discovery of this fine structure in the sextupole, other measurements were performed on HERA, RHIC, ISABELLE, SSC and LHC dipoles [10, 17-24] using Hall probes and rotating coils. The periodicity was found in all dipoles tested, with an amplitude and phase strongly dependent on the previous powering history of the magnet. No systematic trend could be observed comparing the results of different magnets, apart from the fact that the periodicity did appear on all harmonics, on both skew and normal components, with oscillation wavelength identical to the cable twist pitch. We show this feature in Fig. 7 for an LHC dipole prototype.

At constant current conditions the periodicity appeared to change, both in amplitude and phase, exhibiting different time scales. The fastest variations took place just after ramps on typical time scales of the order of 100 to 1000 s, while the slowest changes could last several hours and longer [17, 18, 23].

The explanation for the field periodicity is a non-uniform current distribution in the superconducting cables. If we take a single superconducting cable and we assume that the current distribution is not uniform, we see easily that scanning the cable the magnetic field is stronger in positions closer to the strands with higher current and weaker when closer to strands with lower currents. A very clear experiment showing this behaviour was performed...
by Verweij on a short length of Rutherford cable scanned by means of an array of Hall probes [25]. The current distribution was driven in the experiment by a localised pulsed field. The situation in a magnet is similar, with all cables in the winding of the straight section having a different current distribution, not necessarily related among each other. Each cable contributes an oscillating field in the magnet bore, containing in principle all harmonic components. The periodicity of this oscillating field is of course the cable twist pitch. We expect in this case a periodic pattern on all harmonics, allowed and non-allowed, as indeed we have shown in Fig. 7.

The reasoning above explains the presence of a periodic pattern in the field, but still leaves us puzzled about the origins of the current distribution in a superconducting cable. A non-uniform current distribution in steady-state or transient operation can have several possible causes, as discussed briefly in Appendix 2. To demonstrate the existence of different sources of current imbalance, measurements were performed on LHC dipole prototypes in periodic powering conditions established ramping the magnet continuously with a trapezoidal waveform [24]. The field periodicity was measured with an array of adjacent rotating coils.
The measurements showed that the longitudinal field periodicity could be separated into two clear time scales.

On the slower time scale (larger than 1000 s) the field periodicity had large scattering among different magnets but had little dependence on the length. The faster time scale, in the range of 100 to 300 s could be attributed to long-range coupling currents driven by the variation of the field-change rate in the coil heads, a phenomenon similar to that evidenced in the experiment of Verweij [25]. In the magnet heads the field has a strong variation along the cable, so that the strands are not fully transposed with respect to the field changes. This field discontinuity produces long-range current loops diffusing into the straight part of the magnet (see Appendix 2 and Refs. [26-28]). A spectacular measurement of this effect is shown in Fig. 8, where we plot the amplitude of the field periodicity as scanned in periodic conditions along the LHC dipole prototype MTP1EH, just after ramp-down to the lower current plateau.

Fig. 8 Peak-to-peak amplitude of the normal sextupole periodic pattern as measured just after current ramp-down in periodic conditions in the LHC dipole prototype MTP1EH. Periodic conditions were established by cycling the magnet with a trapezoidal waveform with ramp-rate of 20 A/s. Note the symmetry with respect to the magnet centre and the strong contribution in the magnet ends.

The presence of a periodic pattern with such a short wavelength as the cable twist pitch is of course of no major concern for the beam optics. The particle will integrate the short-scale field variation, sensing only the average value. However, as we will discuss in the next section, we believe at present that a non-uniform current distribution — visible through the associated periodic pattern — bears consequences for the long-term stability of the field. In addition the current distribution has implications for performance limitations during ramping (see for instance [10]) that are not the object of this chapter. Finally, a periodic variation of the harmonics along the magnet length poses problems for the correct measurement of the average value of the harmonic itself. We will deal with this issue in Section 7.

5. FIELD DRIFT

An unexpected and surprising phenomenon during the first operation of the Tevatron collider was the evidence of large chromaticity drifts during periods when the excitation current of the magnets was constant [1]. Data of chromaticity taken from several stores of different durations were converted into an equivalent sextupole change in the main dipoles. The results are shown in Fig. 9, and clearly show that the deduced sextupole was drifting in time during the injection porch. This indirect evidence was supported by later direct measurements on single Tevatron magnets [29-31] that indeed showed a decay of the normal sextupole.
Fig. 9 Sextupole change during a Tevatron injection, deduced from chromaticity measurements in different stores [1]. Reproduced by permission of IEEE. © 1987 IEEE.

At the restart of ramping, after the plateau, the sextupole returned to its original value in a few seconds. This sextupole snap-back could be observed in the Tevatron through collateral effects, such as emittance blow-up and beam losses. Tables were generated based on independent magnetic measurements of dipole magnets and used in Tevatron to cope with the variations of chromaticity [31, 32].

This puzzling behaviour motivated a parametric investigation on the effect of several factors affecting the sextupole decay in the Tevatron magnets [29]. This study showed, among other things, that the sextupole drift at injection was increased:

- pre-cycling the magnet at high operating current. Higher precycle currents corresponded to stronger drifts;
- increasing the duration of the precycle flat-top;
- repeating the pre-cycling procedure several times before the measurement.

In summary, the magnet seemed to show a memory of the previous powering history. This memory could be erased only by quenching the magnet. Because of the relevance for accelerator operation, both in terms of correction of the drift and of machine reproducibility, similar measurements were soon performed on the HERA production magnets [33], and on SSC prototype magnets [34, 35]. Both confirmed the memory effects and the dependence on powering history and in particular on precycle parameters such as maximum current reached and flat-top duration. Stops at intermediate field levels during the precycle and waiting times before reaching the injection level, where measurements were performed, also affected the sextupole drift [35]. Finally, Gilbert showed that a lower strand magnetisation was associated with smaller sextupole drift [34]. As an example of the dependence of the sextupole drift on the powering history, we show in Fig. 10 the measured decays of the module of the normal sextupole in a HERA magnet, as a function maximum field reached during the precycle. In all cases an initial quench was used to erase all previous memory. We see from there another feature that was recognised soon, in the search of an explanation for the drift, namely that the time dependence could be reasonably approximated by means of a single logarithmic decay, with a slope depending on the powering history.

This logarithmic dependence seemed to suggest a thermally activated flux creep in the superconductor (see Appendix 2) as the responsible mechanism for field decay. However, the early tentatives of explanation of the decay of the allowed harmonics based on flux creep could not be confirmed owing to several reasons:
the field drift measured in magnets was much larger than the expected variation based on the flux creep rate measured in cables (see for instance [36] and [37] for data on flux creep);

- the memory and pre-cycle dependence effects mentioned above were not consistent with a flux creep theory (flux creep in a superconductor does not depend on the powering history);

- a temperature drop during the field drift at constant current could not completely stop the harmonics variation [35] although it is known that a temperature decrease stops the flux creep.

Because of the recognition of a systematic direction in the sextupole drift, independent of the direction of the field ramp-rate, the early measurements on the Tevatron magnets had already shown that the drift had to be associated with a slow change of DC magnetisation, rather than with long time constants of cable eddy currents [30]. Brueck et al., measuring the dipole component with a NMR probe and the other harmonics with a rotating coil, showed that in addition there was a good correlation between the dipole and the sextupole decays in the HERA dipole magnets [32, 38]. In all cases the field drift was in the direction of decreasing cable magnetisation. Present measurements on LHC dipoles [23, 39] have confirmed that although the spread in the drift among magnets is large, all allowed harmonics are affected by field drift in a systematic way, namely that corresponding to a decreasing cable magnetisation. On the other hand un-allowed harmonics do not have systematic behaviour and the differences in drift among magnets translates in a spread with (ideally) zero average.

The idea of a magnetisation loss during constant current plateaux and of its recovery at the restart of a ramp also explains the appearance of the so called snap-back at the end of the injection phase in Tevatron, mentioned earlier. As shown in Fig. 11 for an LHC dipole prototype, after the drift at injection the normal sextupole returns approximately to the initial value within few mT of upwards ramps of the dipole field. After the snap-back the sextupole evolution follows the normal ramp-up branch, that is the curve that it would have followed without a stop at the injection field. The snap-back phase takes a limited field change, in the case of Fig. 11 approximately 20 mT. The speculation is that this field change re-establishes
the magnetisation pattern that was somehow perturbed during the injection plateau. An interesting feature is that this field change does not have to be concentrated at the end of the injection plateau for the magnetisation to be re-established. In the second measurements reported in Fig. 10 the injection plateau was substituted with a very slow 15 mT field ramp (approximated by single 0.7 mT steps). Clearly the snap-back has nearly disappeared at the end of the measurement time corresponding to the constant current injection plateau in the previous measurement. In fact, as we see from the details of Fig. 10, a series of mini-snap-backs was generated coincident with each single current step in the approximation of the continuous ramp. The magnetisation loss was compensated in a distributed way rather than in a single step.

![Graph](image)

Fig. 11 Decay and snap-back of the sextupole in the LHC dipole prototype MTP1N2 at approximately 0.6 T (close to the nominal injection field) after a precycle to 8.5 T – 0.04 T – 0.6 T. The operating current is maintained in one case constant (marked as current plateau) and changed in 1 A steps over 20 minutes, for a total of 20 A, in the other case (marked continuous ramp). The dipole measurement for the second case is shown on the second axis. Plotted as a function of time on the left, and as a function of the dipole measured on the right.

What is the reason of the decay of the magnetisation? As we stated before, the decay cannot be completely explained by a flux creep model. In fact, the most plausible explanation at the moment is based on the effect of current distribution and redistribution in the cable. This idea, originated by R. Stiening [40], has been further expanded by later workers [7, 39]. In summary, any change of current distribution in a cable is associated with a periodic variation of the local magnetic field (mostly the self-field) along the cable. In turn any field variation causes a change in the magnetisation state of the superconducting filaments. It can be shown [39] that the net change of the magnetisation of the filaments is always in the direction of a decreasing absolute value of the average cable magnetisation. This indeed explains the systematic drift of the allowed multipoles in the direction of decreasing magnetisation contribution. The diffusion of the current profile in the cable has very long time constants (an evaluation is given in Appendix 2) which are coherent with the characteristic times observed on the field drift (hundreds of seconds and above). Finally, the internal field changes necessary to explain the drift of the harmonics observed is small, in the...
range of 10 mT. Such a field change can be generated in a typical Rutherford cable for accelerators by a current redistribution among strands of some 10 A, a value which is also coherent with the expected current imbalances, e.g. generated by the localised field variations discussed in the previous section (see also Ref. [24] and Appendix 2 for the expected orders of magnitude).

6. OTHER DYNAMIC EFFECTS

Superconducting magnets require, as conventional magnets, strong structural components. The high field generated in the bore is efficiently shielded by a large iron yoke. And in addition to conventional magnets a cryostat with several thermal shields and complex cryogenic connections for cooling and venting is required to insulate the cold mass from the ambient temperature conditions. All these components are metallic, conducting and potentially they can house eddy-current loops. This is indeed the largest source of possible dynamic effects of origin other than superconducting cable magnetisation. As in conventional magnet technology care is taken to segment massive structural or magnetic components, such as the laminated iron yoke. Owing to the limited dynamic range of operation of superconducting magnets (low dB/dt), and also because the superconducting cables are placed in close proximity to the bore, all eddy current effects in conventional structures are orders of magnitude below the effects discussed in the previous sections and, in first approximation, can be neglected.

7. MEASUREMENT OF DYNAMIC FIELDS

The measurement of field and field errors in dynamic conditions in accelerator magnets is not, by itself, a self-standing topic. However, care must be used to extend the standard measurement techniques, based on the use of rotating coils [41], search coils [42], Hall generators [43] or NMR probes [44] to the case of time-varying fields. A first boundary can be traced based on the typical time scale of the field variation in connection with the allowable bandwidth of the measurement system. As we have seen the field dynamics in superconducting magnets exhibits time scales ranging typically from some ms (the coupling current time constant in a strand) to several hundred thousands of s (for current redistribution along the cable length). A time scale below 1 s usually prevents the use of standard rotating coils and NMR equipment, because of the lower limit on the measurement cycle time. The fast field changes are therefore the typical domain of fixed pick-up coils. On the other hand for slow field changes, with typical time scales of the order of 100 s and above, the sensitivity of the fixed coils is limited by the low voltage pick-up induced by the flux change rate as compared to unavoidable noise sources. Hence the domain of slow field variations is usually covered by rotating coils and NMR devices. Hall generators are DC devices with a fast response time [43], as compared to all time scales expected in a superconducting magnet, and can be used therefore in all relevant conditions.

We can trace a second line on the main interest of the measurement, namely whether we are measuring the main field component, that is the magnet strength, or the higher-order harmonics, i.e. the field quality of the magnet. For measurements of harmonics the obvious choice is the rotating coil method, that has superior accuracy on higher-order terms. Both NMR probes and Hall generators have a finite spatial resolution and are therefore quickly bounded in the accuracy of the measurement of high-order harmonics. Their application is therefore limited to the main field component and the lower-order errors. Fixed coils deliver generally a combination of allowed harmonic components, and are also mainly suitable for the main-field component, although measurements of changing field repeated at different angles
using a flat fixed coil can be used to reconstruct the field harmonics. Note that this technique needs reproducible conditions for the measurements.

A third boundary can be finally traced on the local or integrated nature of the measurement, when referred to the longitudinal dimension of the magnet. As an example we are confronted with this problem when we try to map the field periodicity along the magnet bore, as compared on the other hand to the measurement of the field and field harmonics integrated along the whole magnet length. The cable twist pitch, generally in the range of 10 cm, is the shortest scale appearing in a superconducting magnet. Compared to this scale NMR and Hall generators are small devices, and therefore suitable for local measurements. Coils can be manufactured in lengths smaller than the cable twist pitch. Practical problems of coil winding impose a lower limit on the coil length of the order of some cm.

We give in the next sections a few practical examples of measurement devices and techniques specifically tailored to dynamic integrated or local fields. The principles of the measurement devices themselves are not the object of this chapter and will be treated elsewhere in the course, or can be found in the references quoted.

7.1 Measurements of ramped fields using a rotating coil

Rotating coils can be used for ramped fields with the limitation mentioned above, namely that the time scale of the variation is longer than the measurement cycle time. This is the case, as an example, in the measurement of a superconducting accelerator magnet during the nominal ramps from injection to maximum energy. In this case an additional problem is posed when the magnetic field has a significant change in one turn of the rotating coil. To demonstrate this fact, we can take as an example a radial coil of sensitivity $K_{n}^{rad}$ rotating in a time varying field [41]. The magnetic fluxed linked with the coil during a turn is given by:

$$\psi(\theta, t) = \sum_{n=1}^{\infty} K_{n}^{rad} \left( B_{n}(t) \cos(n\theta(t)) - A_{n}(t) \sin(n\theta(t)) \right)$$  \hspace{1cm} (1)

where we see the dependence of the terms on the rotating angle and time. The voltage pickup is caused by a change in the rotation angle $\theta(t)$ and by the variation of the harmonic coefficients in time $B_{n}(t), A_{n}(t)$. The result of the measurement is therefore a harmonic function with changing amplitude. This fact prevents the direct use of the Fourier analysis on the measured fluxes, as now the harmonic coefficients sought at a given time $t$ would be polluted by the effect of their variation during the rotation time. Ideally, the harmonic analysis would need a set of angular points measured simultaneously, that is a snapshot of the flux dependence on $\theta$ at a given time. On the other hand in the measurement we are scanning the angles in sequence, forcibly at increasing time. In spite of this difficulty it is possible to recover the instantaneous value of the harmonic coefficients using a series of subsequent measurements. For what has been said above, the purpose of the analysis is to reconstruct the value of the flux for all angles at a given time. A table containing the value of the flux measured at a given angle and time can be built taking subsequent, continuous measurements (note that both time and angle must be read by the acquisition system). This tables defines a surface $\psi(\theta, t)$ that can be interpolated or fitted. From this analytical approximation we can calculate the flux at any given time [13]. This flux slice is then suitable for harmonic analysis. Such a technique has been used at SSC for the analysis of the ramped measurements presented in Section 3.

An alternative and extremely simple approach is to operate the rotating coil system so that subsequent measurements are taken changing the direction of rotation, in what has been called a washing machine mode. For a constant rotation speed we see at once that the average of two subsequent measurements is identical to a linear interpolation of the $\psi(\theta, t)$ surface referred to the average time between the two rotations. As the average is an extremely easy operation, this procedure can be implemented directly on the system that controls the measurement. This technique is used at CERN for the routine measurements of the LHC magnets [23, 41].
7.2 Measurements of ramped fields using a fixed coil

Fixed coils are the basic measurement technique for time varying fields when the field variation is too fast to allow the harmonic method. The principle is extremely simple, the voltage caused by the variation of the magnetic flux linked with the coil is integrated in time. The calibrated coil surface is then used to derive the average field in the coil area. Fixed coils are clearly dedicated to the measurement of the main field component, and are used routinely at CERN to measure the influence of cable coupling currents on the bending dipoles [26]. The fixed coil measurement can be obtained, as an example, holding a radial rotating coil in a given position and sweeping the magnetic field. This set-up is sensitive to the dipole and all higher-order allowed harmonics. Correction for the influence of higher-order harmonics, mainly sextupole in this set-up, is generally not necessary owing to the dominance of the dipole component. If necessary, it can be applied either based on independent rotating coil measurements of the harmonics, or repeating the fixed coil measurement at several equally spaced angular positions and separating the harmonics via a Fourier decomposition.

As for higher-order harmonics, the influence of the cable coupling currents on the main field component is proportional to the ramp-rate of the magnet. This known characteristic allows one to separate easily the coupling current effect from the steady-state current-field characteristic of the magnet. Several sweeps are taken at different ramp-rates and the magnet current is used to synchronise the different sweeps. At a given magnet current the cable coupling current effect is then determined as the slope of the B(dI/dt) curve obtained plotting the field reading vs. the current ramp-rate. It is worth noting that, because coupling current effects are generally at least three orders of magnitude smaller than the main field itself, the actual synchronisation of the readings requires much care in this type of measurement. For instance a systematic lag in the reading of the current with respect to the reading of the pick-up voltage would also introduce an apparent deviation proportional to the ramp-rate that would add to the sought effect.

7.3 Measurements of field periodicity with Hall generators

Hall generators have been used to measure the longitudinal periodicity of the sextupole component and its time variation at HERA [17] and BNL [18]. The principle of the device, as described by Brueck, [17], is to mount three hall generators on a probe, spacing them at 120°, and adjust the gains of their control amplifiers so that the sum of the signals is sensitive to the sextupole and higher allowed harmonics only. This is obtained by adjusting the gains so that the dipole component of the field is cancelled. The compensation ratio needed is of the order of $10^{-7}$ to guarantee that the sextupole signal is not polluted by spurious dipole reading. To achieve this the Hall generators are temperature controlled, and must be carefully selected [18]. The signal-to-noise ratio is finally increased using a lock-in amplifier. The HERA sextupole detector was equipped with a gravity sensor, thus allowing the adjustment of the direction of the probe to measure the normal sextupole component only. This arrangement has the advantage of providing direct and fast measurements of local sextupole values, and has been used to produce the measurements reported in Fig. 6. The typical spatial resolution is determined by the dimension of the active region of the Hall detector (approximately 5 mm were achieved in the case of the HERA detector), while a time resolution of the order of 0.3 s was achieved. A final refinement consists in using two such sensors, spaced by half a twist pitch [18]. The half-sum of the read-outs of the two sensors gives directly the average value of the periodicity, independently of the periodicity amplitude and the position of the probe. On the other hand the amplitude of the periodicity is obtained from the maximum of the difference of read-outs obtained by scanning the magnet bore over a length of half a cable twist pitch.
7.4 Measurements of field periodicity with rotating coils

Generally rotating coils are built with a long and slender geometry, and are used for measurements of integrated fields over lengths much longer than the cable twist pitch. Still, pushing the winding technique, it is possible to manufacture coils with short length and sufficient sensitivity to allow local measurements of the field and its harmonics. The advantage of such a technique, as compared to the Hall generators probe described in the previous section, is that all harmonics can be obtained in a single measurement cycle. Rotating coils of short length have been used, for instance, at HERA [17], SSC [10], and CERN [23, 24]. A rotating coil of short length is of course identical in principle to its long-length counterpart, apart from geometrical effects due to the coil ends that are now no longer negligible as compared to the coil length.

The short coil provides the integral measurement of the field along its length, a good approximation to the field value in its longitudinal centre of gravity when the length of the coil is much smaller than the characteristic length of the field variation. Coils for local measurements have been built with typical length of the order of 20 to 40 mm, i.e. several times shorter than the typical cable twist pitch. One coil, or a compensated set, measures in any case a single position. The measurements of periodicity must be done then scanning the magnet length, with the obvious drawback that it is not possible to distinguish the spatial and temporal variations. Therefore scans with a short coil are suited for known steady conditions only. This technique has been used to produce the periodicity results of Fig. 7.

In order to overcome this limitation an array of adjacent short coils has been developed at CERN [45]. Each single coil is purely radial, 25 mm long, and is provided with a compensation coil for the dipole suppression. Seven such coil groups are mounted on a support, covering a total length of 175 mm, and are read out simultaneously by a group of digital integrators. This probe, combined with the techniques discussed in Section 7.1, allows ramped measurement of local harmonics, from which average and amplitude of the periodicity can be reconstructed. The results reported in Fig. 8 have been obtained with this array of short rotating coils.

CONCLUSIONS

Among the dynamic effects treated here, coupling current effects have known origin, are reproducible and, within certain limits, calculable. They can introduce field distortions on any harmonic component, and therefore must be controlled. Their control in accelerator magnets wound with Rutherford cables is based on the interstrand contact resistance and on its uniformity, which in turn has deep implications on the manufacturing process.

Regarding the two other effects, field periodicity and field drift, the basic understanding is there, but the modelling is much more difficult. In particular the reproducibility of the sextupole variations at injection is still a hot issue in the two superconducting accelerators operating at present, Tevatron and HERA. In both machines care is taken to limit the history dependence and bring the magnet into a known, reproducible state, from where dedicated procedures are used to correct the drift and the snap-back [31, 46, 47]. Because of the limits on the prediction capability, the main approach followed nowadays in the characterisation of superconducting magnets with respect to their injection behaviour is based on cold measurements and their parametrisation.
REFERENCES


[13] T. Ogitsu, Influence of Cable Eddy Currents on the Magnetic Field of Superconducting Particle Accelerator Magnets, PhD, Institute of Applied Physics, Tsukuba, Japan, 1994


[38] H. Brueck, et al., Time Dependent Field Distorsions from Magnetisation Currents in the Superconducting Hera Magnets, Cryogenics, 30, 605-609, 1990


[55] M. Wilson, Rate Dependent Magnetisation in Flat Twisted Superconducting Cables, RHEL Internal Report, RHEL/M/A26, 1972


[57] L. Krempansky, AC Losses in Flat Twisted Superconducting Cables, CERN Internal Report, SPS/EA/78-2, 1978


[61] B. Turck, Influence of a Transverse Conductance on Current Sharing in a Two-layer Superconducting Cable, Cryogenics, 14, 448, 1974


APPENDIX 1. MAGNETIC FIELD IN ACCELERATOR MAGNETS

For accelerator magnets it is generally accepted, and indeed common practice, to express the magnetic flux density \( B \) in the \( x-y \) plane normal to the beam using the following harmonic expansion in terms of the complex variable \( \zeta = x + iy \):

\[
B(\zeta) = B_y + iB_x = \sum_{n=1}^{\infty} C_n \left( \frac{\zeta}{R_{ref}} \right)^{n-1}
\]  

(A1.1)

where the coefficients \( C_n \) appearing above are the complex harmonic coefficients, and \( R_{ref} \) is an arbitrary reference radius. Equation (A1.1) implies that the field is 2-dimensional, with no component along \( z \). This is usually the case for accelerator magnets that have long longitudinal dimension compared to the winding cross section. The harmonic coefficients \( C_n \) can also be explicitly written as a sum of their real (normal) and imaginary (skew) parts:

\[
C_n = B_n + iA_n
\]  

(A1.2)

Uppercase notation defines the coefficients in non-normalised terms, i.e. given in units of T at the reference radius. More commonly we refer to normalised coefficients, which we will indicate with lowercase letters:

\[
c_n = b_n + i a_n
\]  

(A1.3)

In general the normalised coefficients are obtained for a normal magnet of order \( m \) (where a dipole has \( m = 1 \)) using:

\[
c_n = 10^4 \frac{C_n}{B_m} = 10^4 \left( \frac{B_n}{B_m} + i \frac{A_n}{B_m} \right) = b_n + i a_n
\]  

(A1.4)

where the factor \( 10^4 \) is inserted for convenience, as field errors are generally small, typically of the order of \( 10^{-4} \) of the main field component. Although the normalised harmonic coefficient are dimensionless, they are usually quoted in so-called units, a unit being the result of the normalisation and scaling of Eq. (A1.4).

It is useful to recall here two expressions for the multipoles generated by a single current line normal to the complex plane and placed at a co-ordinate \( R=R_x + i R_y \):

\[
C_n = -\frac{\mu_0 I}{2\pi R_{ref}} \left( \frac{R_{ref}}{R} \right)^n
\]  

(A1.5)

where \( I \) is the current flowing in \( z \) direction. Similarly we can give the multipoles generated by a magnetic moment \( m=m_x + i m_y \), which has no \( z \) component and is therefore in the complex plane \( \zeta \):

\[
C_n = -n \frac{\mu_0 m^*}{2\pi R_{ref}^2} \left( \frac{R_{ref}}{R} \right)^{n+1}
\]  

(A1.6)

where \( m^* \) denotes the complex conjugate of the quantity \( m \). Note that both a single current line and a magnetic moment generate all harmonics.

Most superconducting accelerator magnets are enclosed in an iron yoke which increases the field and shields the exterior from the intense magnetic field. We can give the multipoles generated by an ideal, infinitely large iron shell, with inner radius \( R_{iron} \) and permeability \( \mu \), concentric with the superconducting coil. In the case of the current line, the contribution of the iron is:
\[ C_n = -\frac{\mu_0 I'}{2\pi R_{ref}} \left( \frac{R_{ref}}{R'} \right)^n \]  
(A1.7)

where the current \( I' \) and the radius \( R' \) are obtained mirroring the original current line:

\[ I' = \frac{\mu - 1}{\mu + 1} I \]  
(A1.8)

\[ R' = \frac{R^2_{iron}}{R'^*} \]  
(A1.9).

For a magnetic moment we have that the iron contributes with:

\[ C_n = -n\frac{\mu_0 m^*}{2\pi R_{ref}^2} \left( \frac{R_{ref}}{R} \right)^{n+1} \]  
(A1.10)

where the mirror magnetic moment is given by:

\[ m' = -\frac{R^2_{iron}}{R^2} \frac{\mu - 1}{\mu + 1} m^* \]  
(A1.11)

Given the expressions above, we can see easily that a symmetry condition on the magnet geometry and magnetisation results directly in restrictions on the orders that are allowed in the harmonic expansion. To demonstrate this we recall that a perfectly symmetric multipole magnet of order \( m \) is such that the geometry of the winding and iron is rotationally symmetric by the angle \( \pi/m \). In fact after such a rotation we obtain an identical magnet if in addition we invert the current direction. We can express these symmetry conditions on a general harmonic coefficient \( C_n \) by writing that:

\[ C_n' = C_n e^{\frac{-\pi}{m}} = -C_n \]  
(A1.12)

where the primed coefficient indicates the value of the harmonic after rotation and we have used a known property of the rotation of reference frame on the harmonic coefficients. The condition expressed by Eq. (A1.12) can be satisfied only when:

\[ e^{\frac{-\pi}{m}} = -1 \]

that is, when the harmonic order \( n \) is such that:

\[ n = m(2k + 1) \]  
(A1.13)

where \( k \) is an arbitrary non-negative integer number. The harmonics satisfying Eq. (A1.13) are said to be allowed by the symmetry, while all other harmonics are un-allowed, that is not permitted by the symmetry.

Finally, accelerator magnets are usually produced and positioned so that they generate a pure normal or skew multipole of order \( m \). A normal multipole magnet has top-bottom symmetry in the geometry and current. As a consequence the magnetic field on the midplane has strictly \( y \) direction. This implies immediately that the imaginary (skew) part of any allowed harmonic coefficient must be zero. Similarly a skew magnet has top-bottom symmetric geometry and antisymmetric current. In this case the field on the midplane has \( x \) direction, so that all allowed coefficients have zero real (normal) part. The result is that in a perfectly symmetric normal multipole magnet only the normal allowed harmonics are present, and similarly for a perfectly symmetric skew multipole magnet, where only skew allowed harmonics are present.
APPENDIX 2. SUPERCONDUCTING CABLE IN A MAGNETIC FIELD

For reasons of electrodynamic and thermal stability, technical superconducting materials presently in use are manufactured in the form of twisted multifilamentary strands. The common choice for accelerator magnets is to cable the strands in flat, keystone cables, of the so-called Rutherford type. These cables offer the advantage of high compaction fraction with minimal distortion and degradation of the superconducting strand, in conjunction with good mechanical stability, achieving high operating current density in the range of 300 to 500 A/mm².

Within the scope of this chapter we are interested in the contribution of the superconducting cable to the magnetic field. In general terms any superconductor behaves as a diamagnetic material. A field variation causes long-lasting eddy currents that tend to shield the interior of the superconducting strand or cable. Depending on the path followed, the shielding currents can have a persistent nature (when they flow along a completely superconducting path, e.g. within a superconducting filament) or time dependent nature (if they flow along a partially resistive path, coupling superconducting filaments or strands). These current loops are associated with a magnetisation per unit volume \( \mathbf{M} \), defined here as:

\[
\mathbf{M} = \mathbf{B} - \mu_0 \mathbf{H}
\]  

(A2.1)

where \( \mathbf{H} \) is the magnetic field and \( \mathbf{B} \) is the magnetic flux density. In the next sections we will discuss the major sources of magnetisation in a Rutherford cable, taking the cable volume as reference for the definition of \( \mathbf{M} \).

As we have recalled in Appendix 1, in accelerator magnets the field changes are usually normal to the strands and cables and uniform along the magnet length. We will therefore limit the treatment to the case of uniform normal field variations, neglecting the small errors due to the real strand orientation in a cable. We will assume in addition uniform cable properties. Hence the magnetisation will also be normal to the strand and cable, and will have opposite direction to the field change. For purely normal magnetisation we can obtain the magnetic moment \( \mathbf{m} \) associated with the unit volume magnetisation \( \mathbf{M} \) simply multiplying this last by the cross section of the cable.

2.1 Strand magnetisation due to persistent currents

Let us take an isolated superconducting filament exposed to a field change. It will tend to screen its bulk by means of currents flowing on its surface [48-52]. In an ideal superconductor the screening currents would be confined to an infinitesimal layer, and thus attain infinite current density. In reality the current carrying capacity of a superconductor is limited to the critical current density, function of the magnetic field and temperature. The screening of the filament bulk is obtained then by a finite thickness layer that grows from the filament surface towards the interior as the field increases. In this phase the magnetic field penetrates from the exterior into the filament. This penetration phase proceeds until the screening layers have occupied all the volume available in the filament. We say then that the superconductor is fully penetrated, and the field at which this state is reached is called the penetration field \( B_p \). A reversal of the direction of the field change will initiate a new screening layer making its way towards the filament interior, removing the previous screening current layer. For a cylindrical filament, assuming that the critical current has a negligible variation in the filament, we can compute analytically the magnetisation due to the screening currents. If the external field changes in a cyclic regime, each time reversing completely the screening current patterns, we obtain that the change \( \Delta M \) of the module of the magnetisation \( \mathbf{M} \) after a field change \( \Delta B \) is [51, 52]:

\[
\Delta M = \frac{4\mu_0}{3\pi} \lambda J \int \left[ 1 - \left( 1 - \frac{\Delta B}{2B_p} \right)^3 \right] \quad \text{for} \quad \Delta B \leq 2B_p
\]  

(A2.2)
in the penetration phase, until the maximum trapped magnetisation is reached:

\[ M = \frac{2\mu_0}{3\pi} \lambda J_c D \quad \text{for} \quad \Delta B > 2B_p \quad \text{(A2.3)} \]

after full penetration. The first penetration field (for virgin initial state) is given by:

\[ B_p = \frac{\mu_0 J_c D}{\pi} \quad \text{(A2.4)} \]

and as implied by Eqs. (A2.2), (A2.3), we remark that full penetration is obtained in non-virgin state after a field change twice as large. Above we have used \( J_c \) for the field dependent critical current density and \( D \) for the filament diameter. The factor \( \lambda \) is the ratio of superconductor in the strand cross section, appearing because we have referred the magnetisation to the unit volume of the cable (we neglect for simplicity voids in the cable). As it appears from Eqs. (A2.2), (A2.3) the magnetisation associated with filament shielding currents does not depend on the field change rate, consistent with the fact that the superconducting currents are of persistent nature. At constant field a magnetisation is trapped in the filaments. At the field reversal the trapped magnetisation reverses causing a large hysteresis for field changes larger than the penetration field.

As we have stated above, the magnetisation trapped in the superconducting filament is ideally of persistent nature. In reality some small decrease can be observed monitoring the magnetisation evolution as a function of time, a phenomenon called flux creep. The first measurement of flux creep by Kim et al. [53], was attributed by Anderson to a thermally activated process [54]. Anderson showed in particular that the decay of the magnetisation is proportional to the logarithm of time, a relation that has been confirmed experimentally on accelerator cables [36, 37].

2.2 Strand magnetisation due to coupling currents

The filaments in a single strand are electromagnetically coupled [51], meaning that magnetic flux changes transverse to the strand induce eddy currents that circulate in the superconducting filaments and close resistively across the strand matrix. For this reason these currents are often called coupling currents. Twisting of the filaments reduces the linkage of field changes, and thus limits the magnitude of the coupling currents. Similarly to an R-L circuit, coupling currents are established and decay with a characteristic time constant \( \tau \) that depends on the twist pitch of the filaments in the strand \( l_p \) and on the matrix (effective) transverse resistivity \( \rho_{\text{eff}} \). For NbTi strands used in accelerator magnets this time constant is of the order of 10 ms.

As for the persistent currents it is possible to calculate analytically the ramp-rate dependent magnetisation of a circular superconducting strand subjected to a transverse field change \( \dot{B} \). Assuming that the coupling currents are fully established (i.e. for ramp times much larger than the time constant \( \tau \) of the currents), the module of the magnetisation \( M \) is given by[64]:

\[ M = \frac{\mu_0}{\rho_{\text{eff}} \left( \frac{l_p}{2\pi} \right)^2} \dot{B} \quad \text{(A2.5)} \]

proportional to the ramp-rate.

2.3 Cable magnetisation due to coupling currents

A superconducting, flat cable for particle accelerator magnets responds to field changes in a manner similar to the filaments in a strand. In this case the superconducting strands themselves are coupled. Coupling currents flow along the strands and cross-over at the points where the strands touch each other. We can identify at least two such type of contacts, namely that of crossing strands, touching in a point, and that of adjacent strands, touching
ideally along a line. It is customary to characterise the contacts through two resistances, referred to a single contact of two strands, the transverse $R_t$ and the adjacent $R_a$ resistances.

Distributing the contact resistances along the length of the strand, Wilson [55] came to a continuum approximation for the strand currents and found a solution for the magnetisation of the cable associated to coupling current loops closing across the transverse and adjacent contacts. A different approach was followed by Morgan [56], who modelled the cable by means of an equivalent electrical network after lumping the contacts at discrete points, and found results similar to Wilson. Both approaches were augmented in later works (see for instance [57-59]). In both cases the field variation was considered to be uniform in space, and an infinite cable length with uniform properties was considered. In addition the field variation was assumed to last much longer than the time constants of the coupling currents. In reality these assumptions do not always hold in a magnet. The field variation has significant gradients both along the developed cable length and across the cable width. Cable properties, and in particular contact resistances, are not necessarily uniform along the cable length. Finally joints and splices provide specific boundary conditions generating current diffusion waves along the cable length. All these phenomena do not allow a simple analytic treatment, but can be tackled numerically (see for example [13, 26, 28, 60]).

It is still useful to recall the analytic results of Wilson to provide an estimate of the magnetisation of the cable. In the case of a homogeneous field variation $\hat{B}_\perp$ normal to the wide face of an infinitely long cable with constant contact resistances a convenient expression for the magnetisation associated to fully established coupling currents is [55]:

$$M \approx \mu_0 L_p \left[ \frac{(N^2 - N)}{120} \frac{\alpha}{R_t} + \frac{N^2}{96} \frac{1}{\alpha R_a} \right] \hat{B}_\perp \tag{A2.6}$$

where $N$ is the number of strands, $\alpha$ is the aspect ratio of the cable (ratio of width to thickness) and $L_p$ is the cable twist pitch. The first term in brackets originates from coupling currents closing at the crossing of strands, while the second term is due to currents closing on adjacent strands. We see at once that as in general $\alpha \gg 1$, the second term can be neglected when the transverse and adjacent contact resistances are of the same order of magnitude. Under the same assumptions above, for a uniform field variation $B_\parallel$ normal to the thin face of the cable the magnetisation is given by:

$$M \approx \mu_0 L_p \left[ \frac{N^2}{128} \frac{1}{\alpha^2 R_a} \right] \hat{B}_\parallel \tag{A2.7}$$

### 2.4 Non-uniformities and current distribution in cables

As mentioned in the previous section, a Rutherford cable does not necessarily have uniform properties along its length, and is generally subjected to position dependent field variations during operation in an accelerator. In addition, joints and splices within a coil cause electrical discontinuities and can introduce differences among the series resistances of the strands. All these deviations from the ideal conditions discussed in the previous sections translate into a non-uniform distribution of current between the strands. A current imbalance in the strands of the cable can have several effects. With regard to the main scope of this chapter, a current imbalance is responsible for a periodic pattern in the field harmonics along the length of an accelerator magnet. In addition we believe that the changes in the current distribution related to the generation and decay of the current imbalance are the main cause of the long term drift of the magnetic field through the interaction of the local field in the cable with the magnetisation of the superconducting filaments.
The general problem of current distribution and redistribution in a superconducting cable has no simple analytical solution. A simplified approach that has been repeatedly followed is to consider the case of an ideal two-strand twisted cable, powered with a time-varying current and subjected to a time-varying field [61-64]. Under the further simplifying assumption of an infinite and uniform cable length it is possible to find a closed form solution to this problem that can be used as a guideline in the interpretation of the more general case of a full-size cable. In the ideal two-strand cable with uniform and distributed transverse conductance per unit length \( G' \) it can be shown that the current in each strand satisfies a diffusion equation with a diffusivity coefficient \( \delta \) [7]:

\[
\delta = \frac{1}{L'G'}
\]  

(A2.8)

where \( L' \) is the inductance per unit length of strand. In this approximation any current change propagates along the two-strand cable from the voltage source points as a diffusion wave over the length \( \xi \) with a characteristic time \( \tau_D \):

\[
\tau_D = \frac{\xi^2}{\delta}
\]  

(A2.9)

As we will evaluate in the next section, the characteristic time \( \tau_D \) can be extremely long, in the range of several hundreds s and above. In fact, this is a general characteristic of current redistribution in a superconductor, namely the long time necessary to establish the steady-state conditions.

The source points themselves, as listed previously, are associated with voltage imbalances between the strands, either at the cable ends (joints and splice resistance differences) or distributed along the cable (variation in the flux linkage with the field changes, changes in the cable properties). A case of major interest for an accelerator magnet is that of a localised field change rate \( \dot{B} \) transverse to the cable. This is for instance the situation in the curved ends of a long magnet, where the field and thus the field change rate are different from the values in the straight part. As shown by several authors (see for instance [63, 64]) each discontinuity in \( \dot{B} \) can be a source of long range coupling current loops. Compared to the previous section, where we have considered an infinitely long cable subjected to a uniform transverse field variation, in the case of a localised \( \dot{B} \) the currents coupling the strands do not necessarily compensate after half a twist-pitch, owing to the lack of periodicity in the impressed voltage. Non-homogeneous cable properties can have a comparable effect on the periodicity condition for the strand coupling [60] and also cause long-range currents.

Numerical models, based on the network approximation for the Rutherford cable discussed in the previous section, are generally used to treat accurately the full-size cable geometry in addition to an arbitrary distribution and magnitude of the 26-28, 60, 65]. Based on results of numerical studies, Verweij has derived expressions for the magnitude \( \Delta I \) of the long range currents caused by a localised field change. They apply to Rutherford cables of width \( w \) with uniform interstrand resistance \( R_s \), obtained by cabling \( N \) strands of diameter \( d \) [25, 27] subjected locally to a variation of the field change rate \( \dot{B} \). In the case that the diffusion (characteristic) length \( \xi_D \) is significantly smaller than the cable length, the current imbalance \( \Delta I \) in a strand is given by:

\[
\Delta I \approx 0.88 \frac{w \xi_D}{R_s} \left( 1 - e^{-\frac{N}{9.6}} \right) \Delta \dot{B}
\]  

(A2.10)

and the diffusion length of the current imbalance along the cable is:

\[
\xi_D \approx 0.5 \sqrt{\frac{R_s L_p \pi d^2}{2 \rho N}}
\]  

(A2.11)

where \( \rho \) is an effective strand resistivity that is used in the model to represent the longitudinal electric field associated to current flow in or out of the superconducting filaments, and usually taken in the range of \( 10^{14} \) \( \Omega \)m.
2.5 Orders of magnitude

To give an order of magnitude of the magnetisation associated with the different components we can take a typical cable for use in the inner layer of the LHC main bending dipoles [4]. The main characteristics of this cable are reported in Tab. 1. Magnetisation in the superconducting filaments will be largest at the lowest field levels (when the critical current density is large). On the other hand the magnetisation from the coupling currents will be largest when the ramp-rate is highest. For LHC particle injection is foreseen at 0.54 T, where NbTi has a critical current density in the range of 20,000 A/mm². The highest ramp-rate on the cables is of the order of 7 mT/s, of which we can take approximately 5 mT/s for both components normal and parallel to the broad face. If we compute the magnetisation components as discussed in the previous sections we obtain:

- filaments magnetisation (Eq. (A2.3)) 14 (mT)
- filaments coupling currents (Eq. (A2.5)) 0.3 (mT)
- cable magnetisation normal to the broad face (Eq. (A2.6)) 3.1 (mT)
- cable magnetisation normal to the thin face (Eq. (A2.7)) 0.001 (mT)

We see clearly that for this typical conductor design and operating conditions the dominant magnetisation is due to the filament persistent currents. Generally, in the range of cable parameters given above, the magnetisation due to coupling currents within the strand, and the cable magnetisation due to field changes normal to the thin face of the cable are negligible.

To estimate the characteristic time, length and magnitude of a diffusing current imbalance, we need a value for the interstrand conductance and inductance per unit length. A reasonable approximation is obtained taking:

$$G' = \frac{2}{R_{c}L_{p}} \approx 1.8 \text{ (MSiemens)}$$

and

$$L' = \frac{\mu_{0}}{\pi} \left[ \ln \left( \frac{w}{d} \right) + 0.25 \right] \approx 1.2 \text{ (} \mu\text{H)}.$$ 

Using now Eq. (A2.8) we obtain a diffusivity coefficient $\delta$ of the order of 0.46 m²/s. If we take a typical total cable length of 300 m, representative of a complete inner layer of an LHC main bending dipole, the characteristic time for the diffusion of current along the cable $\tau_{\delta}$ given by Eq. (A2.9), is then of the order of $2 \times 10^5$ s, more than 50 hours. The characteristic diffusion length $\xi_{\delta}$, for a pointwise source of current imbalance, e.g. the coil ends of the LHC dipoles where the transverse field change rate drops from 7 mT/s to virtually zero, can be computed using Eq. (A2.11), from which we obtain $\xi_{\delta} \approx 1.3$ m. Finally, we can estimate the current imbalance in a strand associated with the same discontinuity in $\theta$ using Eq. (A2.10), from which we obtain that $\Delta I \approx 12$ A. This imbalance is significant if compared, for example, to the average current carried by each strand during the LHC injection phase, of the order of 30 A.
Table A2.1

Typical dimensions and major characteristics of a flat cable for the LHC dipoles. Quantities are either nominal values (geometry) or expected range (electrical characteristics).

<table>
<thead>
<tr>
<th>Strand</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Diameter</td>
<td>$d$ (mm)</td>
<td>1.065</td>
</tr>
<tr>
<td>Copper:NbTi ratio</td>
<td>(-)</td>
<td>1.6</td>
</tr>
<tr>
<td>Filling factor</td>
<td>$l$ (-)</td>
<td>0.38</td>
</tr>
<tr>
<td>Filament size</td>
<td>$D$ (mm)</td>
<td>7</td>
</tr>
<tr>
<td>Twist pitch</td>
<td>$l_p$ (mm)</td>
<td>25</td>
</tr>
<tr>
<td>Critical current density</td>
<td>$J_c$</td>
<td></td>
</tr>
<tr>
<td>at 0.5 T, 1.8 K</td>
<td>(A/mm$^2$)</td>
<td>$\approx 20000$</td>
</tr>
<tr>
<td>at 8 T, 1.8 K</td>
<td>(A/mm$^2$)</td>
<td>$\approx 2000$</td>
</tr>
<tr>
<td>Transverse resistivity</td>
<td>$\rho_{ot}$ (Ω m)</td>
<td>$\approx 4.5 \times 10^{-10}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cable</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of strands</td>
<td>$N$ (-)</td>
<td>28</td>
</tr>
<tr>
<td>Cable dimensions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>thin edge</td>
<td>$h_1$ (mm)</td>
<td>1.72</td>
</tr>
<tr>
<td>thick edge</td>
<td>$h_2$ (mm)</td>
<td>2.06</td>
</tr>
<tr>
<td>width</td>
<td>$w$ (mm)</td>
<td>15.0</td>
</tr>
<tr>
<td>Aspect ratio</td>
<td>$a$ (-)</td>
<td>7.9</td>
</tr>
<tr>
<td>Twist pitch</td>
<td>$L_p$ (mm)</td>
<td>110</td>
</tr>
<tr>
<td>Cross contact resistance</td>
<td>$R_c$ (mΩ)</td>
<td>$\approx 10$</td>
</tr>
<tr>
<td>Adjacent contact resistance</td>
<td>$R_a$ (mΩ)</td>
<td>$\approx 10$</td>
</tr>
</tbody>
</table>
MATERIALS

J. Billan
CERN, Geneva, Switzerland

Abstract
In particle accelerators, machine accuracy and magnetic field quality are increasing. Magnetic materials are important elements in the fabrication of magnets and take a non-negligible part in the overall level of accuracy. The selection of these materials has been considered carefully since the beginning of CERN and this chapter is mainly devoted to the different systems in use for such a selection.

1. INTRODUCTION
The natural magnetism of lodestone (Fe₃O₄) was already known 4000 years ago and used by Chinese sailors as compasses to aid navigation, but the industrial use of magnetic materials started really only in this century. Nowadays the most important applications in which magnetic materials are involved concern 50 or 60 Hz devices such as transformers, generators, motors, etc, for electric power generation, and therefore, for economic reasons, the most important parameter which is measured is the core loss in Watt/kg. On the contrary, in particle accelerator machines, magnetic fields are generally fixed or slowly varying and core loss is no longer the important parameter. Magnetic materials are mainly used in electromagnets for energy saving in the magnet yoke, field shaping in the yoke pole-faces and shielding to protect the environment from stray field. Magnetic fields which are used to bend, focus and stabilise the particle beams now require an accuracy in the 10⁻⁶ range, an excellent stability with time and a minimum dispersion from magnet to magnet in order to get the highest possible luminosity. Magnetic materials participate in this goal and should be selected and controlled very carefully. To this end, dedicated apparatus has been designed and is explained in this chapter.

2. THEORETICAL ASPECTS
2.1 Origin of magnetism in matter
In vacuum, a magnetic field is created by the displacement of electric charges, more commonly by electric currents circulating inside a conductor. This field \( B \) at one point expresses the fact that a unit electric charge \( q = 1 \) placed at this point and moving at a unit speed \( v = 1 \) “feels” other electric charges moving around it with a force \( F \):

\[
F = q(v \times B)
\]

(1)

In matter, magnetism is generally not created by electric charges moving in straight lines, but mostly by current loops such as electrons orbiting around their nucleus, or by the spin of particles around themselves (electrons, protons and even neutrons). Magnetic field coming from protons and neutrons in an atom nucleus is approximately 1000 times smaller than the field created by electrons, essentially due to their mass ratio. Therefore magnetism in matter originates essentially from electrons, although the very tiny magnetism coming from the nucleus is used in certain applications such as nuclear magnetic resonance techniques (NMR).

A small current loop \( I \) of surface area \( S \) creates a magnetic moment \( \mu \):

\[
\mu = I S \quad \text{(A.m²)}
\]

(2)
In matter we define as magnetisation $M$, the density of magnetic moments at one point:

$$M = \frac{d\mu}{dv} \quad \text{(A.m}^3) \tag{3}$$

Under the action of external excitation fields ($H$), $M$ is no longer nil, and $M$ and $H$ are proportional:

$$M = \chi H \quad \tag{4}$$

where $\chi$ is the susceptibility of the material. As defined in Eq. (3), susceptibility is related to unit volume and in this case is dimensionless, but it can also be related to unit mass or mole.

Then the total field inside matter is the sum of the external field and the internal magnetisation. The total field, $B$, called magnetic induction, is expressed in Tesla, the external field, $H$, called magnetic field and the magnetisation, $M$, developed inside matter are both expressed in A.m$^3$. $B, H$ and $M$ are exactly of the same nature although two different units are used. The ratio between these two units is:

$$\mu_0 = 4\pi.10^{-4} \quad \text{(m.kg.s}^2.\text{A}^{-3}) \tag{5}$$

and the relation between these three quantities is:

$$B = \mu_0(H + M) \tag{6}$$

In the case of “non-magnetic” materials, essentially diamagnetic and paramagnetic materials as we will see further, $c$ is much smaller than 1 and Eq. (6) is generally rewritten as:

$$B = \mu_0(1 + \chi)H \tag{7}$$

On the contrary, in the case of magnetic materials, $\chi$ is generally much larger than 1 and is replaced by the relative permeability, $\mu_r$,

$$\mu_r = 1 + \chi \quad \tag{8}$$

then,

$$B = \mu_0.\mu_r.H \tag{9}$$

as $\chi, \mu_r$ are dimensionless. By analogy, $\mu_r$ is called the permeability of free space (vacuum permeability) but has the dimension mentioned in Eq. (5).
At the atomic scale, electron magnetism is partly self compensating because the electrons have spin vectors alternatively up and down. Uncompensated spins create a net magnetic moment to the atom. But the magnetic moment vector directions between atoms are generally not ordered and the global magnetisation inside matter is null. A very important exception arises for ferromagnetic materials for which magnetic moments remain naturally aligned, creating a spontaneous magnetisation over macroscopic regions called "domains".

2.2 Diamagnetism

In fact all materials are diamagnetic because diamagnetism has its origin in the Lenz law applied to the current loops representing the orbit of electrons. The magnetic moment of such a current loop varies in order to cancel the flux variation through this loop created by an applied external field. Diamagnetic materials are repelled from a magnetic field source. This effect is independent of temperature and shows no hysteresis, but is always very weak. As an exception, materials which exhibit a superconducting state can be considered as perfectly diamagnetic because they prevent any penetration of an external field. In other materials, paramagnetic and ferromagnetic, diamagnetism is hidden by stronger magnetic moments and on the contrary, are attracted by a magnetic field source. In diamagnetic materials, the global magnetisation is in the opposite direction to the external field and the susceptibility is negative \((-10^4 < \chi_v < -10^5)\).

2.3 Paramagnetism

For several reasons atoms can present a permanent magnetic moment e.g. odd number of electrons, incomplete inner electronic shells (transition and rare-earth elements), but in the absence of an external field they compensate each other and their net effect is zero. Contrary to diamagnetism which is purely an orbit effect, paramagnetism is mostly an electron spin effect. When applying an external field, a torque tends to align all elementary magnetic moments with the field giving then a non-zero net effect. The global magnetisation is in the same direction as the external field and then the susceptibility of a paramagnetic material is positive \((10^3 < \chi_v < 10^5)\).

3. FERROMAGNETISM

3.1 Origin

Like paramagnetic materials, ferromagnetic materials have permanent magnetic moments mostly due to the electron spin effect (> 98%). Normally the exclusion principle says that neighbouring electrons should be of opposite spins and then compensate each other. But for some elements a smallest energy level exists when some electrons of the atom have parallel spins instead of opposite. In this case spins line up from atom to atom and give rise to a spontaneous magnetisation over large regions of material. Such a magnetisation is not induced by an external field as in the case of dia- or paramagnets and reaches from the origin its maximum value, the saturation magnetisation \(M\).

3.2 Atomic structure of ferromagnetic materials [2]

Only a few elements in the periodic classification, Fe, Co and Ni, with respectively 26, 27 and 28 electrons per atom are ferromagnetic (gadolinium with 64 electrons is also slightly ferromagnetic). One of the characteristics of these atoms is that electrons occupy the 42 shell before the 3d shell is completed (Fig. 2). The 4s electrons, the "conduction" electrons, are free to wander between atoms. Some electrons of the 3d shell, the "magnetic" electrons, by interaction with the 4s

![Fig. 2 Electron configuration in an iron atom](image-url)
electrons, tend to align their spins. But other metals, such as Mn, have also an unfilled 3d shell and nevertheless are not ferromagnetic. The difference in these two behaviours lies in the strength of the magnetic coupling between spins of electrons of adjacent atoms (Fig. 3). The energy which aligns spin moments is called exchange energy. This energy can be positive or negative depending on whether spins line up in parallel (ferromagnetism) or antiparallel fashion. The correlation in Fig. 4 shows that for ferromagnetic elements the ratio of the atomic separation distance to the radius of the 3d shell is larger than 1.5. If only a few number of elements fulfil such a condition, ferromagnetism is a rather common phenomenon and has been observed in metal alloys, ceramics (ferrites, garnets), ionic solids (CrBr₃), some semiconductors (MnSb), in amorphous or crystalline structure of these materials.

Fig. 3 Two neighbouring atoms

Fig. 4 The interaction force related to the distance between atoms

3.3 Temperature dependence

In ferromagnetic materials, all points are spontaneously magnetised at the maximum magnetisation $M_{sat}$. But this is only true at $T = 0$ K, where all magnetic moments are perfectly aligned. Above $T = 0$ K, due to disordering created by the lattice thermal vibrations, as in paramagnetic materials (section 2.3), the magnetisation value, $M$, decreases down to zero at the Curie temperature $T_C$. Above this temperature, the material behaves as a paramagnetic material.

Fig. 5 Spontaneous magnetisation versus temperature (nickel) [2]

<table>
<thead>
<tr>
<th>$\mu_0 M_{sat}$ (Tesla, 0 K)</th>
<th>Curie temp. $T_C$ (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iron</td>
<td>2.18</td>
</tr>
<tr>
<td>Nickel</td>
<td>0.64</td>
</tr>
<tr>
<td>Cobalt</td>
<td>1.81</td>
</tr>
</tbody>
</table>
3.4 Magnetic domains

In ferromagnetism there is an apparent contradiction between theory, which explains that magnetic materials are fully magnetised even in the absence of an external field, and practice which generally shows that such materials exhibit no field or fields much smaller than the saturation field. In fact, theory is correct at a microscopic scale but at larger scales the material appears divided into small domains which, if they are all individually fully magnetised, have different directions of magnetisation between them and cancel each other. To cope with the natural tendency to reduce the energy level, the domain structure changes until there is no external flux ("demagnetised" state), as shown in Fig. 6. The energy is divided by approximately the number of domains (Fig. 6b and c), or even reduced to zero with triangular prism domains at the surface of the crystal (Fig. 6d and e).

![Fig. 6 The formation of domains in a single crystal [3]](image)

A change in the spin direction between two domains is made over a very short distance of about 300 atoms. This very special region (Fig. 7), corresponds to a minimum of the sum of the "exchange" energy, which increases with the (square of the) angle between adjacent spins and tends to thicken the wall, and the magnetocrystalline "anisotropy" energy which arises because the magnetisation differs depending on crystallographic directions. This is a minimum when spins are parallel to the easy magnetisation direction and tends to limit the wall thickness. Domains are also limited by grain boundaries, dislocations, strains or impurities.

Three different types of domain boundaries have been identified in ferromagnets as shown in Fig. 8. The first is the Bloch wall which corresponds to a 180° twist of the spin vector directions through the boundary. The second one is the Néel wall which corresponds to a 180° rotation in a same plane. The last type is the crosstie wall which consists of tapered Bloch wall lines jutting out in both directions from a Néel wall spine. Bloch and Néel walls are common in bulk ferromagnets, whereas crosstie walls are mostly present in ferromagnetic films.

![Fig. 7 Change of the spin direction across a domain boundary [1]](image)
4 INFLUENCE OF AN EXTERNAL FIELD

4.1 The magnetisation curve

In the absence of an external field, a ferromagnetic material is generally in an "unmagnetized" state. Inside a domain the magnetisation is at a maximum (saturation), but as all domains compensate each other, at large distances the mean magnetisation, $M$ is nil. The equilibrium established to minimise the energy with the formation of domains is modified under the action of the energy supplied by an external field.

A typical "magnetisation curve" is shown in Fig. 9. To be clear it should be pointed out that if the mean magnetisation varies strongly with $H$, the local magnetisation, always saturated, changes only in direction. Two mechanisms are responsible for this: wall domain displacement, which requires a rather low energy, and rotation of the magnetisation in the direction of $H$ which needs much higher applied fields. At low fields, domains with an easy magnetisation direction parallel to the applied field grow first, but imperfections in the crystal prevent walls from moving below a certain field. In this region the action of the field is still weak but is (nearly) reversible. At medium field, the energy is strong enough to push the walls beyond the crystal imperfections, the domain then occupies all the crystal. Such imperfections are responsible for irreversibilities in the wall displacements (Barkhausen effect), and are the cause of hysteresis. At higher fields, magnetisation still increases but only by rotation of domain magnetisation directions. The saturation $M_S$, is reached when they are all parallel to the applied field.

For practical reasons the magnetisation curve is generally not $M = M(H)$, but $B = B(H)$, as shown in Fig. 10. As $B = \mu_0(M + H)$, and $M$ is much greater than $H$ in most parts of the first magnetisation curve, both curves are comparable. The only difference occurs in the saturation region where the first curve is asymptotic to a constant value: $M$, while the second is asymptotic to the straight line $B = \mu_0(M_s + H)$.

4.2 The permeability

As explained in Section 2, the permeability is the ratio between $B$ and $H$
\[ \mu = \frac{B}{H} \]

\( B \) and \( H \) having different units in the SI unit system, \( B \) in Tesla, \( H \) in A.m\(^{-1}\), then \( \mu \) is not dimensionless. But generally speaking, the permeability is the relative permeability \( \mu \) such that:

\[ \mu = \mu_0 \cdot \mu_r \]

where \( \mu_r \) is dimensionless (N.B. Do not confuse the magnetic moment mentioned in the beginning of this paper and the permeability which will be largely used in the following.) The permeability is defined from the first magnetisation curve as shown in Fig. 10 and is often expressed as a function of the induction in the material as shown in Fig. 11, an example of measurement results on LEP dipole yoke steel. Initial and maximal permeability are the two most relevant figures of permeability.

![Fig. 10 The first magnetisation curve and permeability](image)

![Fig. 11 Permeability vs. induction. Decarburized steel of the LEP dipole yoke.](image)

4.3 The hysteresis loop

After a first magnetisation curve, irreversibilities coming from imperfections in the material prevent the magnetisation from coming back to zero when the external field is removed (Fig. 12). The residual magnetisation at zero external field is named "remanent field": \( B_r \). It is in fact a magnetisation but it is expressed as an induction, in Tesla.

Reversing the excitation field \( H \), the residual magnetisation decreases to zero for a particular field \( H_C \), called "coercive force". \( H_C \) is not the same for the two types of curve: \( M = M(H) \) or \( B = B(H) \). For low-\( H_C \)-value (soft materials), the difference is generally negligible, but for high-\( H_C \) (hard materials or permanent magnets), the difference becomes really significant. In all cases \( H_{CM} > H_{CB} \).

Increasing the field further on the negative side of the curve, more and more domains align with it until a new saturation state is reached. Both saturation states are exactly symmetrical. Coming back to zero and reversing the field again, the complete hysteresis loop is drawn. But the loop is not exactly closed the first time and several loops are necessary to completely close it. The number of "stabilisation" cycles depends on the required accuracy level and on the maximum field level. In the 10-3 range and near saturation, two to three cycles are sufficient. But at lower field and better accuracy, more cycles are necessary.
In AC applications, an interesting feature of this loop is its enclosed area which is proportional to the energy spent to overcome all the irreversibilities in domain movements during the cycle:

\[ W \text{(joules/m}^3\text{)} = \int H \cdot dB \approx \text{loop area} \]

More commonly, \( W \) is expressed in joules/kg. This energy is called the hysteresis loss. It varies with the maximum field and is an intrinsic characteristic of the material. As an example, \( W \) can be divided by a factor 10 for a cold-rolled steel sheet taken before and after a decarburization process. Such a process, largely used in CERN magnet steel production, consisting of 900°C annealing and hydrogen or ammonia decarburation, enlarges the iron grain size and reduces the impurity level, suppressing most of the irreversibility sources. Hysteresis losses are not the only source of losses. Eddy current losses are generally even more important in soft materials used in AC machines such as generators, motors, transformers, etc, which represent far larger production quantities than electro-magnets for particle physics accelerators. Eddy current losses are proportional to the square of lamination thickness, frequency, and field level and inversely proportional to the electrical resistivity of the material. Therefore, reducing the thickness and increasing the electrical resistivity of the magnetic material are important in such devices.

Another feature is that the slope \( dB/dH \) is maximum near \( H_C \). Then to measure \( H_C \) with accuracy, the stabilisation of the magnetisation around \( H_C \) can take a certain time (≈1 min.). After a complete demagnetisation, cycles between symmetric excitation fields, ±\( H_{\text{max}} \), look like those shown in Fig. 13 if the material is not saturated:

\[ H_C < M_{\text{max}} < M_{\text{saturation}} \]

A first deduction is that points \((H_{\text{max}}, B_{\text{max}})\), follow the first magnetisation curve. This property can be useful when measuring the permeability, with search coils, of incompletely demagnetised materials. Another deduction is that \( H_{C_{\text{sat}}} \) is reached before the complete saturation. This is of interest in the design of apparatus measuring \( H_C \) (coercimeter) since the excitation circuit can be much smaller than for a permeameter. For iron, \( H_C \) is maximum at an excitation of 1200 A.m\(^{-1}\) while saturation in a permeameter requires 24000 A.m\(^{-1}\).
5. CLASSIFICATION OF MAGNETIC MATERIALS

Before reviewing the measurement systems of magnetic materials mostly used in accelerator techniques, it is necessary to briefly survey those materials. Figure 14 shows their classification by their two most important characteristics: permeability and coercivity.

As correlation between $\mu$ and $H_C$ exists, up to a certain extent, a measurement of $H_C$ (which is easier) gives a good indication about $\mu$ and can be used to control steel production at better cost. As already mentioned about the hysteresis loop, magnetic materials are divided into two categories: soft and hard. Soft magnetic materials are easy to magnetise and easy to demagnetise while hard magnetic materials are hard to magnetise and hard to demagnetise.

5.1 Soft magnetic materials [4]

Two groups can be distinguished:

Very high $\mu$ materials: mainly the nickel-iron based alloys with $\mu$ between $10^4$ and $10^6$ increasing with the content of nickel (with a maximum around 80%) and very sensitive to heat treatments and degree of cold working. The most common example is Mumetal or 78 Permalloy (78% Ni), with $\mu > 10^5$ and $H_C \approx 1\text{A.m}^{-1}$, but limited to fields smaller than 0.5T. Alloys with 50% Ni (Supranhyster 50, 50 Permalloy) have smaller $\mu$ ($10^4$ to $10^5$), and higher $H_C$ ($\geq 6\text{A.m}^{-1}$) but are still very good up to 1 T. As a drawback, all must be heat treated at very high temperatures (1100°C), and only after shaping. Lately amorphous metals having the same magnetic properties as the above have appeared. They are easier to use and less sensitive to plastic deformations but are only available as very thin (<25 microns) and limited width (<100 mm) sheets. Their resistivity being 3 to 4 times higher than Ni-based materials, they can be used in very special AC applications. But all these materials are expensive.

High $\mu$ materials: can be used in most magnetic applications where a permeability between $10^3$ and $10^4$ is sufficient but they have a higher flux density and are much less expensive than the previous types. The most common is the 3% silicon-iron material. Its high resistivity makes it widely used in standard AC applications. In accelerator applications it is used for small magnet yokes such as steering and corrector magnet yokes. Its advantages are availability, punchability, precisely defined characteristics, and low $H_C$ value, a very important feature for the linearity of magnets working with increasing or decreasing field of either polarity. In large accelerator DC magnet yokes, decarburized steel is preferred because it is less expensive, laminations can be thicker and also because it can be used in 0.2 T higher fields. Its permeability is still around 1000 at 1.6T while 3%Ni-Fe laminations already reach this value at 1.4T.
Ferrites behaves as ferromagnetic materials but the origin of their magnetism is different. In ferrimagnets, there are two kinds of magnetic ions of opposite spin directions but of different magnetic moments, giving therefore a net spontaneous magnetisation as in ferromagnets. Ferrites can be hard or soft, but all are insulators and have no eddy current losses up to microwave application. Soft ferrites are often used in toroid shape as HF transformer cores.

Table 2

Properties of soft magnetic materials

<table>
<thead>
<tr>
<th>Material</th>
<th>$\mu$</th>
<th>$B_S$ (T)</th>
<th>$B_{OPT}$ (T)</th>
<th>$H_C$ (A.m$^{-1}$)</th>
<th>$T_C$ (°C)</th>
<th>$\rho$ (μΩ/cm)</th>
<th>Price (CHF/kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>80%Ni-Fe</td>
<td>$10^4$-$10^6$</td>
<td>0.8</td>
<td>0.7</td>
<td>1</td>
<td>460</td>
<td>15</td>
<td>70-150</td>
</tr>
<tr>
<td>50%Ni-Fe</td>
<td>$10^3$-$10^5$</td>
<td>1.6</td>
<td>1.0</td>
<td>5-10</td>
<td>500</td>
<td>35</td>
<td>30-60</td>
</tr>
<tr>
<td>3%Si-Fe</td>
<td>$5.10^4$-$10^6$</td>
<td>2.0</td>
<td>1.3</td>
<td>30-50</td>
<td>750</td>
<td>47</td>
<td>2-5</td>
</tr>
<tr>
<td>Very low-C steel</td>
<td>$10^3$-$8.10^3$</td>
<td>2.18</td>
<td>1.5</td>
<td>40-80</td>
<td>770</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>Amorphous</td>
<td>$10^4$-$5.10^6$</td>
<td>0.8</td>
<td>0.5</td>
<td>&lt;1</td>
<td>350</td>
<td>130</td>
<td></td>
</tr>
<tr>
<td>Soft ferrite</td>
<td>1400</td>
<td>0.25</td>
<td>0.8</td>
<td>110</td>
<td></td>
<td>10$^4$</td>
<td></td>
</tr>
</tbody>
</table>

5.2 Hard magnetic materials [5]

Hard magnetic materials are more commonly called permanent magnets. A high coercivity is the first criteria of a good hard material but as for soft material, a high-saturation induction is required. Both of these quantities appear in the second quadrant of the hysteresis loop which is also called the demagnetisation curve. This represents the reverse field which must be applied to a material to suppress its remanent magnetisation. Such a curve is essential for permanent magnets when they are used as magnetic field generators. The remanent field $B_r$ is the maximum induction attainable by the magnet when it is “unloaded” or “short-circuited” by an infinite permeability material. On the contrary a zero induction in the magnet ($H_C$ point), can only be obtained with the help of an external demagnetising field. This is impossible in a "passive" circuit, i.e. without any current or other permanent magnets. Permanent magnets are very attractive in accelerator applications because they do not need any excitation coil and generator which are expensive parts of a machine. But their high price limit their use to very specific devices such as ondulators, radiation sources in electron storage rings [6]. Table 3 summarises the properties of the most relevant permanent magnets.

Table 3

Characteristics of the main types of permanent magnets

<table>
<thead>
<tr>
<th>Material</th>
<th>$B_r$ (T)</th>
<th>$\mu_r$</th>
<th>$H_{CS}$ (kA.m$^{-1}$)</th>
<th>$H_{CM}$ (kA.m$^{-1}$)</th>
<th>$B_h_{max}$ (kJ.m$^{-3}$)</th>
<th>Tempco(Br) (%°C)</th>
<th>$\rho$ (μΩ/cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alnico5</td>
<td>1.1-1.3</td>
<td>1.5-5</td>
<td>40-55</td>
<td>40-55</td>
<td>30-50</td>
<td>-0.02</td>
<td>0.5</td>
</tr>
<tr>
<td>Ferrite</td>
<td>0.25-0.4</td>
<td>1.1-1.2</td>
<td>150-250</td>
<td>160-300</td>
<td>10-30</td>
<td>-0.2</td>
<td>&gt;10$^8$</td>
</tr>
<tr>
<td>SmCo5</td>
<td>0.8-0.9</td>
<td>1.03-1.1</td>
<td>600-700</td>
<td>1200-1600</td>
<td>140-200</td>
<td>-0.04</td>
<td>0.6</td>
</tr>
<tr>
<td>Sm2Co17</td>
<td>0.9-1.0</td>
<td>1.03-1.1</td>
<td>650-750</td>
<td>800-1200</td>
<td>150-250</td>
<td>-0.03</td>
<td>0.8</td>
</tr>
<tr>
<td>NdFeB</td>
<td>1.1-1.2</td>
<td>1.04-1.1</td>
<td>800-900</td>
<td>1200-1500</td>
<td>200-300</td>
<td>-0.11</td>
<td>1.5</td>
</tr>
</tbody>
</table>
5.2.1 Alnico permanent magnet

Alnico permanent magnets, based on an Al-Ni-Co(Fe) system, have been the most popular type for the last 50 years. They are cheap and have a high $B_r$ value (1.2 Tesla) but they can easily be demagnetised by a rather small field ($< 0.1$ Tesla). Figure 15 [1], shows the demagnetisation curve of Alnico 5, one of This line has a slope $\mu_r$ called reversible permeability, which has a small value ($\approx 3$), and is nearly constant. Another important parameter is the product $B_rH_r$ ($\approx 35$ kJ.m$^{-3}$ or Am$^{-1}$T), which is twice the magnetic energy density of the magnet and must be taken into account in the design of a magnetic circuit excited by a permanent magnet. The $B_r$ temperature sensitivity of this material is smaller than most of the others: - 0.02%/°C, the best materials of this type (8%Al, 12%Ni, 23%Co, 3%Cu, balance Fe). The two load lines correspond, for example, to the demagnetising effect of a variable air gap in a magnetic circuit using an Alnico permanent magnet. Between the extreme load lines, the magnetisation is reversible and follows nearly a straight line: the "recoil curve" (in fact a thin minor loop).

This line has a slope $\mu_r$ called reversible permeability, which has a small value ($\approx 3$), and is nearly constant. Another important parameter is the product $B_rH_r$ ($\approx 35$ kJ.m$^{-3}$ or Am$^{-1}$T), which is twice the magnetic energy density of the magnet and must be taken into account in the design of a magnetic circuit excited by a permanent magnet. The $B_r$ temperature sensitivity of this material is smaller than most of the others: - 0.02%/°C.

5.2.2 The rare-earth permanent magnets

A revolutionary type of material based on rare-earth elements now makes possible totally new magnetic circuits. Its main feature is that its magnetisation is (nearly) insensitive to external fields, because first $\mu_r$ is very near to 1 ($H_{eb} \approx M_s$), and second $H_{CM}$ is much larger than $M_s$. Practically, this means that thin permanent magnets can even be magnetised in a direction perpendicular to their length, and that this magnetisation is not changed by fields smaller than 1.5 Tesla. Conversely the magnetisation requires very high fields: from 3
to 10 Tesla. Two different rare-earth magnets are commercially available: the samarium-cobalt compounds: SmCo5, Sm2Co17, and the neodymium-iron-boron compound: Nd2Fe14B (see Fig. 16). Their energy product, $B_H$, is nearly 10 times larger than that of other types of magnet, and their volume in a given magnetic circuit can be reduced accordingly. But this does not compensate for their price even if NdFeB is less expensive than SmCo. From the magnetic characteristic point of view, NdFeB is also better than SmCo, but it is not rust-proof and it is more sensitive to temperature.

5.2.3 Ferrite magnets

Another important type of permanent magnet is the "ferrite" (or "ceramic") type. Hard ferrites have much weaker magnetic characteristics than those of the above-mentioned types, and are mainly used in AC applications thanks to their high resistivity.

5.3 Stainless steel

Non-magnetic materials are a really big issue in accelerator equipment design. Among metals, aluminium is certainly one of the more widely used, but its strength and elastic modulus are often too low. Titanium is much better but it is expensive and of reduced availability, and its use is limited, for example, in special vacuum chamber. Ceramic are excellent in many aspects but have two major limitations: price and brittleness.

Non-magnetic stainless steels are very important as structural elements thanks to their high mechanical strength, but they need a very careful selection to be really non-magnetic. This is the reason why low-$\mu$ permeameters have been developed to measure permeabilities between 1 and 2 at room and cryogenic temperatures. The austenitic steels ($\gamma$Fe) are paramagnetic but precipitation of martensite (ferromagnetic) can occur at low temperatures or under plastic deformation seriously increasing their susceptibility. The two principal austenitic steels are the Fe-Cr-Ni and Fe-Mn-Cr ternary alloys. The minimum susceptibility of such materials is in the $10^{-2}$–$10^{3}$ range while paramagnetic materials are in the $10^{3}$–$10^{6}$ range.

6. MEASUREMENT TECHNIQUES
6.1 Strategy in control of material quality

Magnetic properties are not exactly intrinsic properties of a specific material but depend also on many other "external" parameters. The accuracy level required in accelerator techniques demands the knowledge of magnetic properties, such as permeability and coercivity, within a few percent. At such an accuracy level, machining of samples is already a very critical issue. Cold work and heating generated by machining can drastically modify these two characteristics. Another issue is the easiness of using the device, easiness of assuring a good reproducibility and to save time in the operation. For that reason, CERN has always produced its own measuring devices because those existing on the market are not well adapted to its purpose.

The procedure for sample fabrication must be very well defined in order to limit at the lowest possible level heating and cold work. The single-sheet coercimeter has from this point of view the big advantage not to need any sample machining, thus permitting fast and reliable measurements. With the coercimeter the number of sample has been largely increased for a better control of the quality and reproducibility of the production.

Decarburized steel represents by far the largest quantities of magnetic materials used at CERN: 11'000 tons for LEP and 50'000 tons ordered for LHC. This steel is produced as 1.5-mm and 5.8-mm-thick lamination sheets respectively. For such a production two apparatus are used: the ring permeameter and the coercimeter. The permeameter allows the precision measurement of permeability and coercivity on a limited number of samples, while the coercimeter essentially measures the coercivity but on a much larger number of samples.
Thanks to a good correlation between coercivity and permeability, the whole production is sufficiently well controlled by the coercimeter, at less expense. The permeameter is used as a reference.

A third apparatus, the Epstein frame, widely used in industry, has been adapted to our needs to measure thinner steel laminations such as the 0.5-mm-thick silicon-steel laminations. The sampling is quite easy because it only requires a simple cutting of the laminate. The Epstein frame can be used as a permeameter at the exception of the saturation where it is less and less precise due to the increasing flux leaks at the corners.

Low-μ permeameters have been developed for room- and low-temperature use to control the residual magnetism of some stainless steels. At room temperature, a hand-held version can be used to select the better stainless steel already at the supplier store.

6.2 Soft materials

6.2.1 Principle

All the three hereafter described apparatus follow the same principles, for that reason they have been designed in a similar way. Samples are different from each other, but they all need a bipolar power supply and one or two integrators. The different excitation circuits have been designed in such a way as to be excited by the same power supply. Obviously, electronic equipment, procedure and software are identical. The schematic principle is shown in Fig. 17.

The basic procedure is to make a “stable” hysteresis cycle, as schematized in Fig. 18. The material should have never “seen” any flux superior to $\Phi_{\text{max}}$ or it has to be demagnetized progressively down to $\Phi_{\text{max}}$. Generally, only one “stabilization” cycle is sufficient. Then the “measurement” cycle is done reading the four flux variations $\Delta \Phi_1$, $\Delta \Phi_2$, $\Delta \Phi_3$, and $\Delta \Phi_4$. From these four values we deduce:

$$\Phi_{\text{max}} = \frac{\Delta \Phi_1 - \Delta \Phi_2 - \Delta \Phi_3 + \Delta \Phi_4}{4}$$  \hspace{1cm} (10)

and

$$\Phi_{r} = \frac{\Delta \Phi_1 + \Delta \Phi_2 - \Delta \Phi_3 - \Delta \Phi_4}{4}$$  \hspace{1cm} (11)

From the four values $\Delta \Phi_1$, $\Delta \Phi_2$, $\Delta \Phi_3$, and $\Delta \Phi_4$, a simple analysis allows to estimate the correctness of the cycle. An insufficiently “stabilized” cycle will not be closed and the sum of
these four values will not be null, nevertheless this is generally negligible. On the contrary, a drift in the integrator, which gives the same effect, is much more frequent and has to be checked periodically and compensated for. If the drift can be considered as constant, it is eliminated in Eqs. (10) and (11). If the cycle is not symmetrical, modulus of \( \Delta \Phi_1 \) and \( \Delta \Phi_3 \), \( \Delta \Phi_2 \) and \( \Delta \Phi_4 \) respectively, will be different. Beyond a certain difference, a warning can be displayed and the measurement redone. Such diagnostics eases at the validation of the measurement.

Let's define:
\[
\begin{align*}
N & \quad \text{number of excitation turns} \\
N_s & \quad \text{number of turns of the detection coil} \\
S & \quad \text{the sample cross-section area} \\
I_{\text{max}} & \quad \text{the maximum excitation current value} \\
I_{\text{eff}} & \quad \text{the “effective” length of the “mean” flux line}
\end{align*}
\]

\( I_{\text{eff}} \) is very well defined in a ring sample, is still well defined, but to a lesser extent, in an Epstein frame due to the presence of the corners of the sample, but it is much less well defined in a coercimeter due to the presence of air-gaps and yokes. Nevertheless, by measurement and calculations a good estimation of \( I_{\text{eff}} \) can be obtained. From Eq. (10), we can calculate:

\[
B_{\text{max}} = \frac{\Phi_{\text{max}}}{N_s S}
\]

and

\[
H_{\text{max}} = \frac{N I_{\text{max}}}{I_{\text{eff}}}
\]

hence:

\[
\mu = \frac{B_{\text{max}}}{H_{\text{max}}} = \frac{\Phi_{\text{max}} \cdot I_{\text{eff}}}{\mu_0 N_s S \cdot N I_{\text{max}}}
\]

From Eq. (11), we calculate the half aperture of the cycle at zero excitation. \( \Delta \Phi \), and applying the following method, described in Fig. 18, we measure the two currents: \( +I_c \) and \( -I_c \) which are necessary to suppress the remanent field in the sample.

Hence:

\[
H_c = \frac{N I_c}{I_{\text{eff}}}
\]

In coercimeter where there are yokes, the coercive force of the yokes (\( H_{c,yoke} \)), has to be taken into account, even if it is very low (\( H_c \) mumetal \( \approx 1 \text{ A.m}^{-1} \)):

\[
H_c = \frac{N I_c - H_{c,yoke} l_{\text{yoke}}}{I_{\text{eff}}}
\]

\( l_{\text{yoke}} \) is the mean flux line length inside the yoke and is approximately 30\% larger than \( I_{\text{eff}} \). Hence such a correction represents 1\% to 3\% for a decarburized steel.

6.2.2 Ring permeameter [8]

It is well known that any discontinuity, air gap or ends in a magnetic circuit create a demagnetising field which perturbs the applied field and decreases the induction level in the
sample. Apart from infinitely long straight samples, only ring samples are free from demagnetising fields but the excitation field inside the sample is not constant and varies with the inverse of the radius.

\[ H(r) = \frac{N \cdot I}{2 \pi \cdot r} \]

and its mean value is:

\[ \overline{H} = \frac{N \cdot I}{2 \pi \cdot (r_2 - r_1)} \int_{r_1}^{r_2} \frac{dr}{r} \]

true at

\[ r_0 = \frac{r_2 - r_1}{\ln \left( \frac{r_2}{r_1} \right)} \]

where \( r_1 \) and \( r_2 \) are the inner and outer radii of the sample and \( r_0 \) the mean radius. A ring sample is not too difficult to machine even if certain precautions are necessary: no heating or surface cold working during machining. The most important problem is to wind the search and excitation coils around each sample. This problem can be alleviated by splitting the coils into two halves but at the expense of using very good quality interconnections since there are now two connections per turn, all in series, in each coil (see Fig. 19). This is the split-coil ring permeameter which is the CERN standard and reference permeameter since 1967 [8]. At cryogenic temperature, the same ring sample is used but windings have to be wind around it and that makes the measurement complicated and time consuming.

Figure 20 shows the permeability of LEP steels measured with this apparatus. The accuracy of the apparatus is:

\[ B: \quad \pm (10^{-3} + 50 \text{mT}) \]

\[ \mu: \quad \pm 3 \cdot 10^{-3} \quad \text{if } m > 1000 \]
\[ \pm 5 \cdot 10^{-3} \quad \text{if } m < 1000 \]

\[ H_c: \quad \pm (5 \cdot 10^{-3} + 0.5 \text{ A.m}^{-1}) \]

For \( H_c \) two precautions have to be taken in order to obtain such accuracy:

i) \( H_c \) decreases with temperature:

\[ \frac{1}{H_c} \frac{\partial H_c}{\partial T} \approx -10^{-3} / ^\circ C \]
ii) A ring permeameter is not able to take into account any anisotropy in the material. Unlike the single-sheet coercimeter, it only gives an averaged value. Moreover, with thin laminations of relatively large anisotropy, as several rings are stacked, a significant error can occur due to a "magnetic short-circuit" between high-$H_C$ and low-$H_C$ regions when they are superimposed. As an example: a sample of eight rings of 1.5-mm thick decarburized steel laminations with 15% $H_C$ anisotropy, gave 2% difference depending upon whether the rings were superimposed, crossed or parallel. The measured $H_C$ value is always smaller.

At CERN, the ring permeameter is taken as a reference instrument. The only limitation is the large number of contacts of each coil. For the search coil the limited number of turns prevents measurements at very low fields (limited accuracy of the initial permeability). For the excitation coil, the heat dissipation limits the maximum excitation field to 24000 Am$^{-1}$ (0.03 T) for some seconds, hence the study of the saturation region. However such limitations are not important for designing electromagnets.

6.2.3 The Epstein frame

Shown in Fig. 21, this apparatus is made of four stacks of strips placed to form a square. At the corners, the strips are overlapped and clamped in order to reduce the air gaps. As in a ring permeameter, detection and excitation coils are spread as uniformly as possible around the frame, with the exception of the corners. Contrary to their rather limited application at CERN, Epstein frames are widely used in industry for AC measurements (50Hz losses). The accuracy depends strongly on the geometry of the frame. The effect of the air gaps at the corners is negligible when the strips have high shape factors (width/thickness $> 10$ and length/width $> 15$) and permeabilities smaller than 10000. A correct estimation of the mean effective length of the flux lines, permits a 1% accuracy on $\mu$ and $H_C$.

6.2.4 The single-sheet coercimeter [9]

In accelerators the reproducibility between magnets of the same kind should be as close as possible. Magnets being powered in series, only the mechanical accuracy and the magnetic properties of the yokes can cause differences between them. A steel with good magnetic properties is important, but its reproducibility from yoke to yoke is still more important. $H_C$ and $\mu$ are both very important and only a ring permeameter can measure them simultaneously. However, as already mentioned in Section 6.1, a certain correlation exists between them and a coercimeter, much easier to use, is very useful in controlling the reproducibility of the steel production. If this correlation is excellent at low field, it still good up to 1 Tesla with correlation coefficients superior to 80% for a low-carbon
steel. \( H_e \) is much easier to measure with a coecimeter for several reasons:
- The excitation field can be limited to 1200 Am\(^{-1}\), \( H_e \) being maximum at 1.5 T
- The influence of an air gap is very small
- Using a zero method, the search coil needs no special accuracy
- The lamination thickness has a very small influence
- An anisotropy measurement is possible
- For the coecimeter described below there is no sample preparation
- Each measurement takes only a few minutes

The coecimeters built at CERN were designed to measure steel sheets without preparation and just before being punched. The length of the yokes should be equal to or greater than the width of the steel sheet. To guarantee a very good flux return with a minimum coercivity, a Mumetal (78\% Ni-Fe) has been chosen as yoke material.

Figure 23 shows a schematic outline of these coecimeters. To take advantage of the high permeability of Mumetal, its maximum flux density has been limited to 0.25T thus determining the maximum sheet thickness (LEP: 3mm max, 1.5mm nominal, LHC: 6mm max, 5mm nominal). A maximum excitation current of 8A is sufficient in the 12 turns (LEP) or 17 turns (LHC) of the main excitation coil. A four-turn auxiliary excitation coil creates a flux circulating only in the yokes and is used to measure the mean value of the four air-gap heights. Small corrections (<1\%) due to the influence on the field distribution in the air-gap of the air-gap height, and also of the lamination thickness and of the width of the contacts, have been calculated and can be applied.

A 50-turn search coil (Fig. 23) measures the flux variation in the steel sheet and two other similar coils surrounding the yokes are used to measure the air gap height but also the flux division between the two yokes signalling thus any dissymmetry in the system.

The accuracy of the apparatus comes mainly from \( l_{eff} \) which has first to be calculated, the measurement then comparing very well with the ring permeameter. With some precautions an accuracy of 1\% is possible for \( H_e = 50 \) Am\(^{-1}\). One precaution is to ensure that the earth field direction is perpendicular to the excitation field of the coecimeter. The earth field horizontal component is about 16 Am\(^{-1}\) and part of it is not shielded by the yokes. Some typical results of coecimeter measurements are shown in Figs. 24 and 25.
6.3 Low-\( \mu \) materials

6.3.1 Low-\( \mu \) permeameter at ambient temperature [10]

Stainless steels which have excellent mechanical properties should be non-magnetic and need to be carefully selected to be used in the field of accelerator magnets (see Section 5.3). The apparatus shown in Fig. 26 has been designed to measure the residual permeability of stainless steels. This apparatus is based on the measurement of the flux variation in a magnetic circuit when approaching a slightly ferromagnetic material to the air gap of the circuit. A search coil detects the flux created by a permanent magnet and a return yoke creates the air gap. The operation consists of placing the magnetic circuit in contact with the sample to be measured, resetting the integrator of the search coil, and measuring the flux variation caused by moving the circuit 12 mm apart from the sample. A calibration made with respect to three stainless steel samples measured with a magnetic balance.

Susceptibilities measured with this apparatus range between 0 to 1, with an accuracy of 1\% \pm 0.0001 for massive and flat pieces. Corrections should be applied for thin or rounded pieces. A special application was the measurement of the nickel layer thickness (\( \approx 7\mu \)) of the LEP vacuum chamber through the 3 mm-thick lead shield.

6.3.2 Low-\( \mu \) permeameter at low temperature [11]

This apparatus has been designed to measure the residual magnetisation of stainless steels. As shown in Fig. 27, a superconducting solenoid creates an excitation field between 0.1 and 6 tesla, a 12000 turns search coil is placed in the centre of the solenoid and the measurement is made by plunging the sample from a remote position into the search coil. The voltage created by the flux variation due to residual magnetisation in the sample is integrated...
and gives the permeability of the material. An anti-cryostat allows for measurements at higher temperatures: 78K and 300K. For materials with permeability smaller than 2, 5mm-diameter and 40mm-long samples are used. For such low permeabilities, the demagnetising effect of the ends is negligible. For higher permeabilities (up to 20), such as steels at the onset of the saturation region, a measurement is still possible on condition that samples are made with a smaller cross-section, in order to limit the demagnetising effect. Two types of 40mm-long steel samples have been made: 2mm-diameter cylinders and 1mm-square bars. The weight allows a good estimation of the sample cross-section. An accuracy of 1% is still easily obtained. For steel, both apparatus are complementary and measurements compare well in the range 0.1 to 1 tesla.

Fig. 27 The 4.2K low-μ permeameter

Fig. 28 Permeability at 4.2K of two

REFERENCES


OVERVIEW OF MAGNET MEASUREMENT METHODS

K.N. Henrichsen
CERN, Geneva, Switzerland

Abstract
Electromagnets used as beam guiding elements in particle accelerators and storage rings require very tight tolerances on their magnetic fields along the particle path. Also large volumes of magnetic field in modern spectrometer magnets must be mapped in order to allow for the precise tracking of secondary particles. The article describes the methods and equipment used for these types of magnetic measurements. Short descriptions are given of the magnetic resonance techniques, the fluxmeter method, the Hall generator and the fluxgate magnetometer. A few of the more exotic methods are also briefly mentioned. References of historical nature as well as citations of recent work are indicated. It is mentioned when sensors and associated equipment are commercially available.

1. INTRODUCTION

Before computers became common tools, electromagnets were designed using analytical calculations or by measuring representative voltage maps in electrolytical tanks and resistive sheets. Magnetic measurements on the final magnets and even on intermediate magnet models were imperative at that time.

Nowadays it has become possible to calculate strength and quality of magnetic fields with an impressive accuracy. However, the best and most direct way to verify that the expected field quality has been reached is to perform magnetic measurements on the finished magnet. It is also the most efficient way of verifying the quality of series produced electromagnets in order to monitor wear of tooling during production.

It is curious to note that while most measurement methods have remained virtually unchanged for a very long period, the equipment has been subject to continual development. In the following only the more commonly used methods will be discussed. It is noticeable that these methods are complementary and that a wide variety of the equipment is readily available from industry. For the many other existing measurement methods, a more complete discussion can be found in the two classical bibliographical reviews [1, 2]. An interesting description of early measurement methods can be found in [3].

2. MEASUREMENT METHODS

2.1 Choice of measurement method

The choice of measurement method depends on several factors. The field strength, homogeneity and variation in time, as well as the required accuracy all need to be considered. Also the number of magnets to be measured can determine the method and equipment to be deployed. As a guide, Fig. 1 shows the accuracy which can be obtained in an absolute measurement as a function of the field level, using commercially available equipment. An order of magnitude may be gained by improving the methods in the laboratory.

2.2 Magnetic resonance techniques

The nuclear magnetic resonance technique is considered as the primary standard for calibration. It is frequently used, not only for calibration purposes, but also for high precision measurements. However, it is generally not suitable for use in production because of the large amount of time required for the measurement.

Overview of Magnet Measurement Methods

K.N. Henrichsen
CERN, Geneva, Switzerland

Abstract
Electromagnets used as beam guiding elements in particle accelerators and storage rings require very tight tolerances on their magnetic fields along the particle path. Also large volumes of magnetic field in modern spectrometer magnets must be mapped in order to allow for the precise tracking of secondary particles. The article describes the methods and equipment used for these types of magnetic measurements. Short descriptions are given of the magnetic resonance techniques, the fluxmeter method, the Hall generator and the fluxgate magnetometer. A few of the more exotic methods are also briefly mentioned. References of historical nature as well as citations of recent work are indicated. It is mentioned when sensors and associated equipment are commercially available.

1. INTRODUCTION

Before computers became common tools, electromagnets were designed using analytical calculations or by measuring representative voltage maps in electrolytical tanks and resistive sheets. Magnetic measurements on the final magnets and even on intermediate magnet models were imperative at that time.

Nowadays it has become possible to calculate strength and quality of magnetic fields with an impressive accuracy. However, the best and most direct way to verify that the expected field quality has been reached is to perform magnetic measurements on the finished magnet. It is also the most efficient way of verifying the quality of series produced electromagnets in order to monitor wear of tooling during production.

It is curious to note that while most measurement methods have remained virtually unchanged for a very long period, the equipment has been subject to continual development. In the following only the more commonly used methods will be discussed. It is noticeable that these methods are complementary and that a wide variety of the equipment is readily available from industry. For the many other existing measurement methods, a more complete discussion can be found in the two classical bibliographical reviews [1, 2]. An interesting description of early measurement methods can be found in [3].

2. MEASUREMENT METHODS

2.1 Choice of measurement method

The choice of measurement method depends on several factors. The field strength, homogeneity and variation in time, as well as the required accuracy all need to be considered. Also the number of magnets to be measured can determine the method and equipment to be deployed. As a guide, Fig. 1 shows the accuracy which can be obtained in an absolute measurement as a function of the field level, using commercially available equipment. An order of magnitude may be gained by improving the methods in the laboratory.

2.2 Magnetic resonance techniques

The nuclear magnetic resonance technique is considered as the primary standard for calibration. It is frequently used, not only for calibration purposes, but also for high precision measurements. However, it is generally not suitable for use in production because of the large amount of time required for the measurement.
solids by two independent research teams [6–8]. Since then, the method has become the most important way of measuring magnetic fields with very high precision. Based on an easy and precise frequency measurement it is independent of temperature variations. Commercially available instruments measure fields in the range from .011 T up to 13 T with an accuracy better than 10 ppm.

In practice, a sample of water is placed inside an excitation coil, powered from a radio-frequency oscillator. The precession frequency of the nuclei in the sample is measured either as nuclear induction (coupling into a detecting coil) or as resonance absorption [9]. The measured frequency is directly proportional to the strength of the magnetic field with coefficients of 42.57640 MHz/T for protons and 6.53569 MHz/T for deuterons. The magnetic field is modulated with a low-frequency signal in order to determine the resonance frequency [10].

The advantages of the method are its very high accuracy, its linearity and the static operation of the system. The main disadvantage is the need for a rather homogeneous field in order to obtain a sufficiently coherent signal. A small compensation coil, formed on a flexible printed circuit board and providing a field gradient, may be placed around the probe when used in a slightly inhomogeneous field. A correction of the order of 0.2 T/m may be obtained [10]. The limited sensitivity and dynamic range also set limits to the suitability of this method. It is, however possible to use several probes with multiplexing equipment, if a measurement range of more than half a decade is needed.

Pulsed NMR measurements have been practiced for various purposes [11, 12], even at cryogenic temperatures [13]. But equipment for this type of measurement is not yet commercially available.

Finally, it should be mentioned that a rather exotic method of NMR measurements using water flowing in a small tube has given remarkably good results at low fields [14–16]. It fills the gap in the measurement range up to 11 mT, for which NMR equipment is not yet commercially available. In addition, it provides a method of measurement in environments of strong ionizing radiation such as in particle accelerators. It was tested for measurements in the bending magnets installed in LEP. A resolution of 0.0001 mT was reached in the range from the remanent field of 0.5 mT, up to the maximum field of 112.5 mT and a corresponding reproducibility was observed [17]. The remarkable sensitivity and resolution of this measurement method makes it very suitable for absolute measurements at low fields. In fact, it was even possible to detect the Earth magnetic field outside the magnet, corresponding to an excitation frequency of about 2 kHz. However, the operation of this type of equipment is rather complicated due to the relatively long time delays in the measurement process.

Electron spin resonance (ESR) [18-21] is a related and very precise method for measuring weak fields. It is now commercially available in the range from 0.55 mT to 3.2 mT with a reproducibility of 1 ppm and is a promising tool in geology applications.

Magnetic resonance imaging (MRI) has been proposed for accelerator magnet measurements [22]. It is a very promising technique which has proven its quality in other applications. However, the related signal processing requires powerful computing facilities,
2.3 The fluxmeter method

This method is based on the induction law. The change of flux in a measurement coil will induce a voltage across the coil terminals. It is the oldest of the currently used methods for magnetic measurements, but it can be very precise [23]. It was used by Wilhelm Weber in the middle of the last century [24] when he studied the variations in strength and direction of the earth’s magnetic field. Nowadays it has become the most important measurement method for particle accelerator magnets. It is also the most precise method for determining the direction of the magnetic flux lines; this being of particular importance in accelerator magnets. The coil geometry is often chosen to suit a particular measurement. One striking example is the Flux ball [25] whose complex construction made it possible to perform point measurements in inhomogeneous fields.

Measurements are performed either by using fixed coils in a dynamic magnetic field, or by moving the coils in a static field. The coil movement may be a rotation through a given angle, a continuous rotation or simply a movement from one position to another. Very high resolution may be reached in field mapping using this method [26].

Very high resolution may also be reached in differential fluxmeter measurements using a pair of search coils connected in opposition, with one coil moving and the other fixed, thus compensating fluctuations in the magnet excitation current and providing a much higher sensitivity when examining field quality. The same principle is applied in harmonic coil measurements, but with both coils moving. A wide variety of coil configurations are used, ranging from the simple flip-coil to the complex harmonic coil systems used in fields of cylindrical symmetry.

2.3.1 Induction coils

The coil method is particularly suited for measurements with long coils in particle accelerator magnets [27, 28], where the precise measurement of the field integral along the particle trajectory is the main concern. Long rectangular coils were usually employed and are still used in magnets with a wide horizontal aperture and limited gap height. In this case, the geometry of the coil is chosen so as to link with selected field components [29]. The search coil is usually wound on a core made from a mechanically stable material, in order to ensure a constant coil area, and the wire is carefully glued to the core. Special glass or ceramics with low thermal dilatation are often used as core materials. During coil winding the wire must be stretched so that its residual elasticity assures a well-defined geometry and mechanical stability of the coil.

Continuously rotating coils with commutating polarity were already employed in 1880 [3]. The harmonic coil method has now become very popular for use in magnets with circular cylindrical geometry, in particular superconducting beam transport magnets. The coil support is usually a rotating cylinder. This method has been developed since 1954 [30, 31]. The induced signal from the rotating coil was often transmitted through slip rings to a frequency selective amplifier (frequency analyzer), thus providing analog harmonic analysis.

The principle of a very simple harmonic coil measurement is illustrated in Fig. 2. The radial coil extends through the length of the magnet and is rotated around the axis of the magnet. As the coil rotates, it will cut the radial flux lines. A number of flux measurements are done between predefined angles and will permit the precise and simultaneous determination of the strength, quality and geometry of the magnetic field. A Fourier analysis of the measured flux distribution will result in a precise description of the field parameters in terms of the harmonic coefficients:

\[ B_r(r,\varphi) = B_0 \sum_{n=1}^{\infty} \left( \frac{r}{r_0} \right)^{n-1} (b_n \cos n\varphi + a_n \sin n\varphi) \]

where \( B_0 \) is the amplitude of the main harmonic and \( r_0 \) is a reference radius. \( b_n \) and \( a_n \) are the harmonic coefficients. In this notation \( b_1 \) will describe the normal dipole coefficient, \( b_2 \) the
With the advent of modern digital integrators and angular encoders, harmonic coil measurements have improved considerably and are now considered as the best choice for most types of particle accelerator magnets, in particular those designed with cylindrical symmetry [32]. In practice, the coil is rotated one full turn in each angular direction while the electronic integrator is triggered at the defined angles by an angular encoder connected to the axis of the coil. In order to speed up the calculation of the Fourier series, it is an advantage to choose a binary number (e.g. 512) of measurement points. A compensating coil, connected in series and rotated with the main coil may be used to suppress the main field component and thus increase the sensitivity of the system for measurements of field quality. Dynamic fields are measured with a static coil linking to selected harmonics [33]. The harmonic coil measurement principle and its related equipment was described in detail in [34]. A thorough description of the general theory including detailed error analysis can be found in [35]. The practical use of the harmonic coil method for large scale measurements in superconducting magnets was described in [36, 37] and more recent developments in [38–42]

Another induction measurement consists of moving a stretched wire in the magnetic field, thus integrating the flux cut by the wire [43]. It is also possible to measure the flux change while varying the field and keeping the wire in a fixed position. Tungsten is often selected as wire material, if the wire cannot be placed in a vertical position. The accuracy is determined by the mechanical positioning of the wire. Sensitivity is limited, but can be improved by using a multi-wire array. This method is well suited to geometry measurements, to the absolute calibration of quadrupole fields and in particular to measurements in strong magnets with very small aperture.

The choice of geometry and method depends on the useful aperture of the magnet. The sensitivity of the fluxmeter method depends on the coil surface and the quality of the integrator. The coil-integrator assembly can be calibrated to an accuracy of a few tens of ppm in a homogeneous magnetic field by reference to a nuclear magnetic resonance probe, but care must be taken not to introduce thermal voltages in the related cables and connectors. It must also be avoided to induce erratic signals from wire loops exposed to magnetic flux changes. Not only the equivalent surface of the search coil must be measured, but also its median plane which often differs from its geometric plane due to winding imperfections. In the case of long measurement coils, it is important to ensure very tight tolerances on the width of the coil. If the field varies strongly over the length of the coil, it may be necessary to examine the variation of the effective width.

The main advantage of search coil techniques is the possibility of a very flexible design of the coil. The high stability of the effective coil surface is another asset. The linearity and the wide dynamic range also plays an important role. The technique can be easily adapted to measurements at cryogenic temperatures. After calibration of the coils at liquid nitrogen
other hand, the need for relatively large induction coils and their related mechanical apparatus, which is often complex, may be a disadvantage. Finally, measurements with moving coils are relatively slow.

2.3.2 The flux measurement

Induction coils were originally used with ballistic galvanometers and later with more elaborate fluxmeters [44]. The coil method was improved considerably with the development of photoelectric fluxmeters [45] which were used for a long period of time. The measurement accuracy was further improved with the introduction of the classic electronic integrator, the Miller integrator. It remained necessary, however, to employ difference techniques for measurements of high precision [46]. Later, the advent of digital voltmeters made fast absolute measurements possible and the Miller integrator has become the most popular fluxmeter. With the development of solid state d.c. amplifiers, this integrator has become inexpensive and is often used in multi-coil systems.

Figure 3 shows an example of such an integrator. It is based on a d.c. amplifier with a very low input voltage offset and a very high open-loop gain. The thermal variation of the integrating capacitor is the most critical problem. The integrating components are therefore mounted in a temperature-controlled oven. Another problem is the decay of the output signal through the capacitor and the resetting relay. So, careful protection and shielding of these components is essential in order to reduce the voltages across the critical surface resistances.

![Fig. 3 Analog integrator](image-url)

The dielectric absorption of the integrating capacitor sets a limit to the integrator precision. A suitable integrating resistor is much easier to find. Most metal-film resistors have stabilities and temperature characteristics matching those of the capacitor. The sensitivity of the integrator is limited by the d.c. offset and low frequency input noise of the amplifier. A typical value is 0.5 µV which must be multiplied by the measurement time in order to express the sensitivity in terms of flux. Thermally induced voltages may cause a problem, so care must be taken in the choice of cables and connectors. In tests at CERN the overall stability of the integrator time constant proved to be better than 50 ppm over a period of three months. A few electronic fluxmeters have been developed by industry and are commercially available.

In more recent years, a new type of digital fluxmeter has been developed, which is based on a high quality d.c. amplifier connected to a voltage-to-frequency converter (VFC) and a counter. The version shown in Fig. 4 was developed at CERN and is now commercially
of the VFC. Two counters are used in order to measure with continuously moving coils and to provide instant readings of the integrator. One of the counters can then be read and reset while the other is active. In this way no cumulative errors will build up. This fluxmeter has a linearity of 50 ppm. Its sensitivity is limited by the input amplifier, as in the case of the analog amplifier.

This system is well adapted to digital control but imposes limits on the rate of change of the flux since the input signal must never exceed the voltage level of the VFC. The minimum integration period over the full measurement range must be of the order of one second to obtain a reasonable resolution.

![Digital integrator diagram](image)

**Fig. 4** Digital integrator

### 2.4 The Hall generator method

E.H. Hall discovered in 1879 that a very thin metal strip immersed in a transverse magnetic field and carrying a current developed a voltage mutually at right angles to the current and field that opposed the Lorentz force on the electrons [47]. In 1910 the first magnetic measurements were performed using this effect [48]. It was, however, only around 1950 that suitable semiconductor materials were developed [49–51] and since then the method has been used extensively. It is a simple and fast measurement method, providing relatively good accuracy, and therefore the most commonly used in large-scale field mapping [52–54]. The accuracy can be improved at the expense of measurement speed.

#### 2.4.1 Hall-probe measurements

The Hall generator provides an instant measurement, uses very simple electronic measurement equipment and offers a compact probe, suitable for point measurements. A large selection of this type of gaussmeter is now commercially available. The probes can be mounted on relatively light positioning gear [54]. Considerable measurement time may be gained by mounting Hall generators in modular multi-probe arrays and applying multiplexed voltage measurement. Also simultaneous measurements in two or three dimensions may be carried out with suitable probe arrays [55, 56]. The wide dynamic range and the possibility of static operation are other attractive features.

However, several factors set limits on the obtainable accuracy. The most serious is the temperature coefficient of the Hall voltage. Temperature stabilization is usually employed in order to overcome this problem [57], but increases the size of the probe assembly. The temperature coefficient may also be taken into account in the probe calibration by monitoring the temperature during measurements [58]. It depends, however, also on the level of the magnetic field [58], so relatively complex calibration tables are needed. Another complication can be that of the planar Hall effect [59], which makes the measurement of a weak field component normal to the plane of the Hall generator problematic if a strong field component is present parallel to this plane. This effect limits the use in fields of unknown geometry and in particular its use for determination of field geometry.

Last but not least is the problem of the non linearity of the calibration curve, since the Hall
better linearity and has a smaller active surface than the classical rectangular generator. Its magnetic center is, therefore, better defined, so it is particularly well suited for measurements in strongly inhomogeneous fields. Special types, which have a smaller temperature dependence, are available on the market, but these show a lower sensitivity.

The measurement of the Hall voltage sets a limit of about 20 µT on the sensitivity and resolution of the measurement, if conventional d.c. excitation is applied to the probe. This is mainly caused by thermally induced voltages in cables and connectors. The sensitivity can be improved considerably by application of a.c. excitation [61, 62]. A good accuracy at low fields can then be achieved by employing synchronous detection techniques for the measurement of the Hall voltage [63].

Special Hall generators for use at cryogenic temperatures are also commercially available. Although they show a very low temperature coefficient, they unfortunately reveal an additional problem at low temperatures. The so-called "Shubnikov-de Haas effect" [64, 65] shows up as a field dependent oscillatory effect of the Hall coefficient which may amount to about one per cent at high fields, depending on the type of semiconductor used for the Hall generator. This adds a serious complication to the calibration. The problem may be solved by locating the Hall generator in a heated anticryostat [66]. The complications related to the planar Hall effect are less important at cryogenic temperatures and are discussed in detail in [67]. Altogether, the Hall generator has proved very useful for measurements at low temperature [68].

2.4.2 Calibration

Hall generators are usually calibrated in a magnet in which the field is measured simultaneously using the nuclear magnetic resonance technique. The calibration curve is most commonly represented in the form of a polynomial of relatively high order (7 or 9) fitted to a sufficiently large number of calibration points. This representation has the advantage of a simple computation of the magnetic induction from a relatively small table of coefficients.

A physically better representation is the use of a piecewise cubic interpolation through a sufficient number of calibration points which were measured with high precision. This can be done in the form of a simple Lagrange interpolation or even better with a cubic spline function. The advantage of the spline function comes from its minimum curvature and its "best approximation" properties [69]. The function adjusts itself easily to nonanalytic functions and is very well suited to interpolation from tables of experimental data. The function is defined as a piecewise polynomial of third degree passing through the calibration points such that the derivative of the function is continuous at these points. Very efficient algorithms can be found in the literature [70]. The calculation of the polynomial coefficients may be somewhat time-consuming but need only be done once at calibration time. The coefficients (typically about 60 for the bipolar calibration of a cruciform Hall generator) can be easily stored in a microprocessor device [57, 63] and the subsequent field calculations are very fast. The quality of the calibration function can be verified from field values measured between the calibration points. A well designed Hall-probe assembly can be calibrated to a long term accuracy of 100 ppm. The stability may be considerably improved by powering the Hall generator permanently and by keeping its temperature constant.

2.5 Fluxgate magnetometer

The fluxgate magnetometer [71] is based on a thin linear ferromagnetic core on which detection and excitation coils are wound. The measurement principle is illustrated in Fig. 5. In its basic version, it consists of three coils wound around a ferromagnetic core: an a.c. excitation winding A, a detection winding B that indicates the zero field condition and a d.c. bias coil C that creates and maintains the zero field. In practice the coils are wound coaxially in subsequent layers. The core is made up from a fine wire of Mumetal or a similar material that has an almost rectangular hysteresis curve. The method was introduced in the 1930’s and was also named "peaking strip". It is restricted to use with low fields, but has the advantage of offering a linear measurement and is well suited for static operation. As a directional device with very high
Earth magnetic field. Much more complex coil configurations are applied for precision measurements and in cases where the measured field should not be distorted by the probe. The most interesting application is now in space research and important developments of this technique have taken place over the last decades [72–74]. The use of modern materials for magnetic cores has improved the sensitivity to about 20 pT and can assure a wide dynamic range. The upper limit of the measurement range is usually of the order of a few tens of mT, but can be extended by applying water cooling to the bias coil. Fluxgate magnetometers with a typical range of 1 mT and a resolution of 1 nT are commercially available from several sources. They have many other practical applications, for example in navigation equipment.

![Fluxgate magnetometer](image)

**Fig. 5** Fluxgate magnetometer

### 2.6 Magneto-resistivity effect

Magneto-resistivity was discovered by W. Thomson in 1856 [75]. It was exploited quite early and a commercial instrument already existed at the end of last century. Technical problems were, however, important [76]. Dependence on temperature and mechanical stress, combined with difficulties of manufacture and problems with electrical connections, caused a general lack of reliability in this measurement method. As with the Hall generator, it was only when semiconductor materials became available that the method turned into a success. Then inexpensive magneto-resistors came on the market and were used also for magnetic measurements [77]. A more recent application for field monitoring was implemented in one of the large LEP spectrometers [78].

### 2.7 Visual field mapping

The best known visual field mapper is made by spreading iron powder on a horizontal surface placed near a magnetic source, thus providing a simple picture of the distribution of flux lines. Another very classical way of observing flux-line patterns is to place a free-moving compass needle at different points in the volume to be examined and note the direction of the needle. This compass method was applied, long before the discovery of electromagnetism, for studies of the variations in the direction of the earth's magnetic field. Another visual effect may be obtained by observing the light transmission through a colloidal suspension of diamagnetic particles, subject to the field [79, 80].

#### 2.7.1 Faraday effect

The magneto-optical rotation of the plane of polarization of polarized light (Faraday effect) is a classical method for the visualization of magnetic fields. A transparent container filled with a polarizing liquid and placed inside the magnet gap may visualize for example the field pattern in a quadrupole by observation through polarization filters placed at each end of the magnet. The rotation of the plane is proportional to the field strength and the length of the polarizing medium and may give a certain indication of the field geometry. This measurement principle has proved useful for measurements of transient magnetic fields [81, 82]. It is less convincing when applied to the precise determination of magnet geometry, even though modern image
2.7.2 Floating-wire method

Floating-wire measurements were quite popular in the past [83]. If a current-carrying conductor is stretched in a magnetic field, it will curve subject to the electromagnetic force and describe the path of a charged particle with a momentum corresponding to the current and the mechanical tension in the wire. A flexible annealed aluminium wire was used in order to reduce effects of stiffness and gravity. This method has now been entirely replaced by precise field mapping and simulation of particle trajectories by computer programs.

2.8 Measurements based on particle beam observation

A method for the precise measurement of the beam position with respect to the magnetic center of quadrupole magnets installed in particle accelerators has been developed over the last decade [84, 85]. The procedure consists of modulating the field strength in individual lattice quadrupoles while observing the resulting beam orbit oscillations. Local d.c. orbit distortions are applied in the search for the magnetic center. This so-called K-modulation provides a perfect knowledge of the location of the particle beam with respect to the center of a quadrupole. In addition, it may provide other very useful observations for operation and adjustment of the accelerator [86]. This is obviously of particular importance for superconducting accelerators [87]. It is very difficult to provide a superconducting quadrupole magnet with a direct optical reference to its magnetic center, so errors caused by changes of temperature profiles and other phenomena may build up as time passes.

The method may be further improved by synchronous detection of the oscillation, so that its phase can be identified. The sensitivity of the detection is impressive. Experience from LEP [88] showed that an absolute accuracy of 0.05 mm in both the vertical and horizontal plane could be obtained. Furthermore it was observed that a modulation of the quadrupole field of about 300 ppm could be clearly detected, which means that measurements may be carried out on colliding beams while particle physics experiments are taking place. This measurement method also played an important role for adjustments of the so-called Final Focus Beams [89, 90].

3. CONCLUDING REMARKS

Proven measurement methods and powerful equipment is readily available for most of the measurement tasks related to beam-guiding magnets as well as for spectrometer magnets. It is therefore prudent to examine existing possibilities carefully before launching the development of a more exotic measurement method. Many unnecessary costs and unpleasant surprises can be avoided by choosing instruments which are commercially available. The measurement methods described above are complementary and the use of a combination of two or more of these will certainly meet most requirements. Already at an early stage of the system design, particular attention must be drawn to definitions of geometry and the future alignment considerations.

In the field of new technologies, there are two methods which merit consideration. Magnet resonance imaging is a promising technique which could find a lasting application. Also the use of superconducting quantum interference devices (SQUIDS) might in the long run become an interesting alternative as an absolute standard and for measurements of weak fields [91, 92]. The complexity of these methods is still at a level which prevents current laboratory use.
REFERENCES


[35] W.G. Davies, "The theory of the measurement of magnetic multipole fields with rotating


[77] E. Welch, P.R. Mace, "Temperature stabilized magneto-resistor for 0.1 % magnetic field measurement", Proc. 3rd Int. Conf. on Magnet Technology, Hamburg, Germany (1970) 1377-1391.


SEARCH COILS

M.I. Green
Lawrence Berkeley National Laboratory, University of California, Berkeley, California, USA

Abstract
The theory, design, fabrication and calibration of point, area, and line-integral coils used for magnetic measurements is discussed. This includes the various materials used, and the tolerances and techniques of fabrication. Some of the instruments and techniques used to calibrate and utilize search coils are described.

1. INTRODUCTION
Search coils have long been used for measuring magnetic fields [1]. The sensitivities of coils are inherently linear with the strength of magnetic fields. If they are designed appropriately, they have the ability to measure inhomogeneous fields.

The use of search coils is based upon the law of Induction, Faraday’s law, that relates the time rate of change of the flux in a coil to the e.m.f. induced in the coil.

\[ \text{e.m.f.} = -\frac{\partial}{\partial t} \oint_{\text{Coil}} \mathbf{B} \cdot d\mathbf{s} \]  

The change of flux in the coil can be due to:
- the movement of the coil in a static field
- the motion of a magnet in the vicinity of a fixed coil
- the time variation of a magnetic field in the vicinity of a fixed coil

I will be putting my emphasis on those topics with which I am most familiar. Emphasis will also be on techniques and topics that are generally difficult to find in the published literature.

Very little of the following is my own work. I have plagiarized liberally from my colleagues at LBNL, United States and European Laboratories, and commercial vendors.

2. TYPES OF SEARCH COILS
Search coils can be categorized in the following manner:
- Point coils for representing the field at an infinitely small point in space.
- Line integral coils for representing the magnetic induction along a path in space.
- Area coils for representing the flux within an area of space.
- Harmonic coils which are sets of coils configured to be sensitive to specific spatial harmonics of the field.

2.1 Point coils
Point coils are designed to measure/represent the magnetic induction at a point. Note that if the coil was actually a point, it would have no area and consequently would have no sensitivity. Point coils will have an associated vector area, and number of turns in the coil. The vector associated with the area is perpendicular to the plane of the coil windings.

The flux linkage $\Phi$ of the coil is given by the vector dot product of the magnetic induction $\mathbf{B}$ and the vector area $\mathbf{A}$ of the coil:
\[ \Phi = \vec{B} \cdot \vec{A} \]  \hspace{1cm} (2)

where:

\[ \Phi = \text{flux linkage (Tesla-m²)} \]
\[ = \text{(Volt-seconds)} \]
\[ \vec{B} = \text{magnetic induction (Tesla).} \]

![Fig. 1 Area vector](image)

The vector area of the coil \( \vec{A} \) (m²) can be expressed as:

\[ \vec{A} = \sum_{i=1}^{n} \vec{A}_i = A\hat{a} = N\overline{A}_i\hat{a} \]  \hspace{1cm} (3)

where,

\( N = \text{number of turns of coil} \)
\( \vec{A}_i = \text{the vector area of one turn} \)
\( A = \text{effective turns area of the coil (m²)} \)
\( \hat{a} = \text{unit vector in the direction of } \vec{A} \)
\( \overline{A}_i = \text{average area per turn.} \)

Point coils are tightly wound coils usually in cylindrical shape.

Important Parameters

- \( A = \text{effective turns area of the coil (m²)} \)
- Effective location of "point" coil.
- Orientation of coil \( \hat{a} \)

2.1.1 Flux ball [2]

The flux ball is the best approximation to a point coil, i.e. a coil whose e.m.f. best represents the field at a point located at the center of the coil, independent of the spatial harmonics present in the field. Reference [2] shows that a coil wound on a spherical surface with a uniform number of turns per axial length, represents the value of the axial field at its center and is insensitive to higher spatial harmonics. An excellent approximation to the flux ball (Fig. 2) which has a relatively simple construction is fabricated by making the coil of concentric cylinders, each cylinder wound with the identical number of turns per unit length and having the same axial center. The radial distance between cylinders is inversely proportional to their mean radius.

2.1.2 Simple cylindrical point coils

Simple cylindrical coils are the easiest to wind. These coils will have a uniform inside and outside radius, and each layer will be the same length. Herzog and Tischler experimentally determined that a length-to-outside diameter of 0.72 was better than the theoretically derived value of 0.67 [3]. All axially symmetric coils are insensitive to even the harmonics. It is possible to wind a simple cylindrical coil that is also insensitive to the third spatial harmonic. L.J. Laslett presents Table 1 for relative dimensions for simple cylindrical coils. This table is based upon Garett's work (see Bibliography).
Table 1
Relative dimensions of cylindrical "point" coil for vanishing sextupole sensitivity

<table>
<thead>
<tr>
<th>( a_x/a_z )</th>
<th>( \Delta x/a_z )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.67082</td>
</tr>
<tr>
<td>0.1</td>
<td>0.67115</td>
</tr>
<tr>
<td>0.2</td>
<td>0.67341</td>
</tr>
<tr>
<td>0.3</td>
<td>0.67924</td>
</tr>
<tr>
<td>0.4</td>
<td>0.68982</td>
</tr>
<tr>
<td>0.5</td>
<td>0.70584</td>
</tr>
<tr>
<td>0.6</td>
<td>0.72756</td>
</tr>
<tr>
<td>0.7</td>
<td>0.75486</td>
</tr>
<tr>
<td>0.8</td>
<td>0.78738</td>
</tr>
<tr>
<td>0.9</td>
<td>0.82462</td>
</tr>
<tr>
<td>1.0</td>
<td>0.86603</td>
</tr>
</tbody>
</table>

Fig. 2 Quarter section view of flux ball
Fig. 3 Cross section of simple cylindrical point coil

2.1.3 Complex cylindrical point coil (Fig 4)
Laslett also says that Garett has specified the relative dimensions of a coil for which the sensitivities to the third and fifth harmonic vanish and the sensitivity to the seventh harmonic is very small.

\[ x_1 : x_2 : a_3 : a_4 = 0.45735 : 0.82148 : 0.35 : 0.704545 : 1 \]

2.2 Line integral
Line-integral coils are designed to measure the magnetic induction along a path. The path may be straight or curved. The ideal coil has infinitesimal width and height, but is not practical.
\[ BL = \int B \cdot dl = \int \frac{\mathbf{B} \times d\mathbf{A}}{w} = \int \frac{d\Phi}{w} \]  

(4)

Symmetrical rectangular cross section, line-integral coils will be insensitive to the even harmonics. If the dimensions satisfy the following equation, the line-integral coil will also be insensitive to the sextupole component of the field.

\[ \Delta x = \sqrt{\frac{a_1^2 + a_2^2}{2}} \]  

(5)

Important parameters
- \( w \) = turns width, (m)
- \( L \) = length of coil, (m).

![Fig. 4 Complex cylindrical point coil](image1)

![Fig. 5 Cross section of line-integral coil](image2)

2.3 "Area" Coils

I really don't know the proper technical term for this type of coil. I am defining "area" coils as those coils that have significant inside dimensions, for example, a circular or rectangular coil whose inner and outer dimensions are of the same order of magnitude.

Uses:
- Measure ELF (Extra Low Frequency) field generated by power lines.
- One coil of a large Helmholtz coil pair.
- Hysteresis graph coils to measure magnetic properties of materials. These solenoidal type coils enclose the sample.

Important parameters:
- Effective turns area
- Turns
- Any linear dimension

2.4 Harmonic coils

2.4.1 Dipole / Quadrupole / Multipole
2.4.2 Radial / Tangential
2.4.3 Morgan

Harmonic coils could be classified as complex "Area" coils. They are described in a later chapter of this publication.
3. MATERIALS FOR WINDING FORM

3.1 Aluminum
Easy to machine.
Not practical for AC fields.
Lorentz force can be significant.
Coefficient of linear expansion = 25 * 10^-6 / °C.
'On-the-fly' - Not a good idea.

3.2 Alumina
Not easily machinable. Ground. Can be machined in the unfired state—must take into account shrinkage when fired.
Very stiff and stable.
Coefficient of linear expansion = 6.5 * 10^-6 / °C.
Good insulator.
Fragile

3.3 "Glasses" (see Table 2)
MACOR™ Corning [4].
- Easily machinable glass-ceramic.
- High strength
96% Silica (Vycor™ Corning)
- Not easily machinable, usually ground.
Fused silica (Quartz)
- Not easily machinable, usually ground.

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Some Characteristics of &quot;Glasses&quot;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>MACOR</th>
<th>96% Silica (Vycor)</th>
<th>Fused Silica (Quartz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific gravity</td>
<td>2.52</td>
<td>2.18</td>
<td>2.20</td>
</tr>
<tr>
<td>Thermal expansion (°C)</td>
<td>9.4*10^-7</td>
<td>8*10^-7</td>
<td>5.6*10^-7</td>
</tr>
<tr>
<td>Compressive strength (psi)</td>
<td>50,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mod. elasticity (psi)</td>
<td>9.3*10^6</td>
<td>9.6*10^6</td>
<td>10.5*10^6</td>
</tr>
<tr>
<td>Shear modulus (psi)</td>
<td>3.7*10^6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poisson's ratio</td>
<td>0.26</td>
<td>0.18</td>
<td>0.17</td>
</tr>
</tbody>
</table>

3.4 Glass composite - fiberglass reinforced epoxy resin
NEMA G-10 (see Table 3), or 11/FR-4 epoxy fiberglass is a very commonly used material manufactured from continuous filament woven glass fabric impregnated with epoxy resin. They possess low moisture absorption, and excellent electrical characteristics over a wide range of humidities and temperatures. FR-4 is similar to G-10, and is, in addition, flame retardant.
Both are machinable with care.

Table 3

<table>
<thead>
<tr>
<th>Characteristics and properties of NEMA G-10 sheets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tensile strength</strong></td>
</tr>
<tr>
<td>Lengthwise</td>
</tr>
<tr>
<td>Crosswise</td>
</tr>
<tr>
<td><strong>Compressive strength</strong></td>
</tr>
<tr>
<td>Flatwise</td>
</tr>
<tr>
<td>Edgewise</td>
</tr>
<tr>
<td><strong>Flexural strength (1/8&quot; thick)</strong></td>
</tr>
<tr>
<td>Lengthwise</td>
</tr>
<tr>
<td>Crosswise</td>
</tr>
<tr>
<td><strong>Modulus of elasticity</strong></td>
</tr>
<tr>
<td>Lengthwise</td>
</tr>
<tr>
<td>Crosswise</td>
</tr>
<tr>
<td><strong>Specific gravity</strong></td>
</tr>
<tr>
<td><strong>Coeff. of thermal expansion</strong></td>
</tr>
<tr>
<td><strong>Water absorption (1/2 &quot; thick)</strong></td>
</tr>
</tbody>
</table>

3.5 **Graphite composite - graphite reinforced epoxy resin**

   More difficult to machine than glass composite
   Negligible temperature coefficient of thermal expansion at room temperature.
   Has finite conductance.
   Very stiff.

3.6 **Foam board - paper reinforced plastic foam**

   We once had to measure a 3-meter long, 1-meter inside-diameter quadrupole. Paper reinforced plastic foam turned out to be an ideal material for these coils. It is very light and very stiff and quite easy to machine. We obtained a bucking ratio of ~500 with this coil assembly.

4. **CONDUCTORS**

4.1 **Copper**

   Temperature coefficient of resistance = 0.39%/\degree C.
   Work hardens.
   Not very strong.
   Coefficient of linear expansion = $16.6 \times 10^{-6}$/\degree C
   Cross Section
     - Rectangular
     - Square
     - Round

4.2 **Conductor insulation**

   Categorized according to Thermal Class as shown in Table 4.
   Thickness can be single, heavy, triple, quadruple build, etc.
Table 4
Copper wire insulation thermal classes

<table>
<thead>
<tr>
<th>Thermal Class (°C)</th>
<th>Insulation Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>105</td>
<td>Enamel, polyurethane, Formvar, bondable polyurethane, bondable</td>
</tr>
<tr>
<td></td>
<td>Formvar, bondable polyurethane-Nylon</td>
</tr>
<tr>
<td>130</td>
<td>Polyurethane, H. T., polyurethane-Nylon</td>
</tr>
<tr>
<td>155</td>
<td>Polyester</td>
</tr>
<tr>
<td>180</td>
<td>Polyester-imide, polyester-imide-Nylon, solderable polyester, bondable</td>
</tr>
<tr>
<td></td>
<td>Polyester-imide, glass fibers</td>
</tr>
<tr>
<td>200</td>
<td>Polyester-amide-imide, Teflon,</td>
</tr>
<tr>
<td>220</td>
<td>Polyimide, Kapton tape</td>
</tr>
<tr>
<td>500</td>
<td>Aluminum oxide, ceramic coated</td>
</tr>
</tbody>
</table>

Insulation characteristics and general applications can be found in MWS Technical Data Bulletin.

4.3 Conductor bonding

4.3.1 Bondable overcoats (see Table 5)

Heat bonding
- Oven heating
- Resistance heat bonding

Solvent bonding
- Passed through saturated felt wick
- Spray

Table 5
Copper wire bondable overcoats

<table>
<thead>
<tr>
<th>Type</th>
<th>Operating Temp. (°C)</th>
<th>Heat Activation Temp. (°C)</th>
<th>Solvent Activating Agents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polyvinyl Butyral</td>
<td>105</td>
<td>120–140</td>
<td>Alcohol</td>
</tr>
<tr>
<td>Epoxy</td>
<td>155</td>
<td>130–150</td>
<td>Methyl ethyl ketone or acetone</td>
</tr>
<tr>
<td>Polyester</td>
<td>155</td>
<td>130–150</td>
<td>MEK</td>
</tr>
</tbody>
</table>

4.4 Potting

The coil is placed in a mold filled with epoxy, then vacuum pumped and finally oven baked to cure the epoxy. The areas of epoxy encapsulated coils have temperature sensitivities in the range of \(-3 \times 10^{-5} \text{/°C}\).
4.5 Complex conductor configurations

4.5.1 MWS multifilar wire
Multifilar wire (see Fig. 6) is parallel-bonded, color-coded wire. It is available in sizes AWG 24 to 48 with up to twenty strands in some sizes. Its advantages are:
- Geometrically uniform windings
- More precise location of individual strands.
- Bonding may be an art form.

We have used multifilar wire at LBNL in fabricating both line-integral and harmonic coils.

![Fig. 6 MWS color-coded multifilar wire](image)

4.5.1.1 Solder connections
Consider six adjacent windings of 12-strand AWG 44 multifilar wire. Only 11 solder connections are required to obtain 72 turns plus two for leads. The solder connections are made outside of the magnet region of interest but within the calibrating field region. In order to minimize perturbing the coil area, solder connections are made in a plane parallel to the search coil area vector (see Fig. 7).

![Fig. 7 Multifilar wire solder connections](image)
4.5.2 Litz wire

The term 'litz' wire is derived from the German word litzendraht meaning woven wire. As generally defined, it is a wire constructed of individually film-insulated wires bunched or braided together in a uniform pattern of twists and length of lay. It is also essential to position each individual strand in the litz construction in a uniform pattern moving from the center to the outside and back in a given length. Litz wire was developed for use in radio frequency circuits as the multistrand configuration minimizes the power losses found in a solid conductor. The litz wire bundle can be enclosed in a textile wrap for insulation.

MWS has litz wire of 175 strands of AWG 48 wire.
A technicians nightmare!
Wire bundle is an excellent approximation to an infinitesimally-narrow single-strand of wire.
Wire bundle can be laid in a slot or stretched between 'pulleys'.
Spots of epoxy can be used to glue litz wire to the form.
Solder connections are made outside the field region.

4.6 Printed-circuit-board search coils

Can be computer generated.
No machining of form.
No winding of wires.
Density of conductors considerably less than with discrete windings, ~3 conductors/mm OK.
Fiducial holes can be indicated.
Very uniform cross sections obtainable.
Single Layer.
Flexible - i.e. for placement on surface of cylinder.

Multilayer Example

Used for ESRF dipoles.

C. Fougeron at Laboratoire National Saturne specified.

Fabricated by CIRETEC Ref. [5].

Each of the 20 layers has 8.5 turns resulting in a total of 170 turns. The effective width is 6.5 mm. 25- and 40-cm long coils of this geometry were fabricated for ESRF. The full width at half maximum variation was $1 \times 10^{-3}$ and if one was choosy, a coil could be found with $\Delta w/w = 1 \times 10^{-4}$.

Fig. 8 Printed circuit board integral coil

Fig. 9 Variation of width of printed circuit board integral coil
4.7 Wafer technique

Coils have been fabricated using the techniques developed to make integrated circuits. It is possible to place conducting lines on a substrate very accurately. The coils made with this technique have very high resistance turns.

4.8 Copper wire winding tension

Typical winding tensions are shown in Table 6.

<table>
<thead>
<tr>
<th>Size (AWG)</th>
<th>Size (in)</th>
<th>Size (mm)</th>
<th>Tension (gm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>0.0071</td>
<td>0.1803</td>
<td>360</td>
</tr>
<tr>
<td>34</td>
<td>0.0063</td>
<td>0.1601</td>
<td>280</td>
</tr>
<tr>
<td>35</td>
<td>0.0056</td>
<td>0.1422</td>
<td>220</td>
</tr>
<tr>
<td>36</td>
<td>0.0050</td>
<td>0.1270</td>
<td>180</td>
</tr>
<tr>
<td>37</td>
<td>0.0045</td>
<td>0.1143</td>
<td>140</td>
</tr>
<tr>
<td>38</td>
<td>0.0040</td>
<td>0.1016</td>
<td>110</td>
</tr>
<tr>
<td>39</td>
<td>0.0035</td>
<td>0.0889</td>
<td>87</td>
</tr>
<tr>
<td>40</td>
<td>0.0031</td>
<td>0.0787</td>
<td>69</td>
</tr>
<tr>
<td>41</td>
<td>0.0028</td>
<td>0.0711</td>
<td>56</td>
</tr>
<tr>
<td>42</td>
<td>0.0025</td>
<td>0.0635</td>
<td>45</td>
</tr>
<tr>
<td>43</td>
<td>0.0022</td>
<td>0.0559</td>
<td>35</td>
</tr>
<tr>
<td>44</td>
<td>0.0020</td>
<td>0.0508</td>
<td>29</td>
</tr>
<tr>
<td>45</td>
<td>0.00176</td>
<td>0.0457</td>
<td>18.2</td>
</tr>
<tr>
<td>46</td>
<td>0.00157</td>
<td>0.0406</td>
<td>14.5</td>
</tr>
<tr>
<td>47</td>
<td>0.00140</td>
<td>0.0350</td>
<td>11.5</td>
</tr>
<tr>
<td>48</td>
<td>0.00124</td>
<td>0.0305</td>
<td>9.05</td>
</tr>
<tr>
<td>49</td>
<td>0.00111</td>
<td>0.0279</td>
<td>7.30</td>
</tr>
<tr>
<td>50</td>
<td>0.00099</td>
<td>0.0254</td>
<td>5.75</td>
</tr>
<tr>
<td>51</td>
<td>0.00088</td>
<td>0.0224</td>
<td>4.55</td>
</tr>
<tr>
<td>52</td>
<td>0.00078</td>
<td>0.0198</td>
<td>3.58</td>
</tr>
<tr>
<td>53</td>
<td>0.00070</td>
<td>0.0178</td>
<td>2.87</td>
</tr>
<tr>
<td>54</td>
<td>0.00062</td>
<td>0.0158</td>
<td>2.26</td>
</tr>
<tr>
<td>55</td>
<td>0.00055</td>
<td>0.0140</td>
<td>1.78</td>
</tr>
<tr>
<td>56</td>
<td>0.00049</td>
<td>0.0124</td>
<td>1.40</td>
</tr>
<tr>
<td>57</td>
<td>0.000438</td>
<td>0.0111</td>
<td>1.14</td>
</tr>
<tr>
<td>58</td>
<td>0.000390</td>
<td>0.00991</td>
<td>0.90</td>
</tr>
<tr>
<td>59</td>
<td>0.000347</td>
<td>0.00881</td>
<td>0.72</td>
</tr>
<tr>
<td>60</td>
<td>0.000309</td>
<td>0.00785</td>
<td>0.56</td>
</tr>
</tbody>
</table>

4.9 Wiring

4.9.1 Thermal e.m.f.'s

Whenever two dissimilar metals are in contact with each other, a voltage will be generated.
4.10 Thermoelectric potentials

<table>
<thead>
<tr>
<th>Materials</th>
<th>Potential (µV/°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu-Cu</td>
<td>≤ 0.2</td>
</tr>
<tr>
<td>Cu-Ag</td>
<td>0.3</td>
</tr>
<tr>
<td>Cu-Au</td>
<td>0.3</td>
</tr>
<tr>
<td>Cu-Cd/Sn</td>
<td>0.3</td>
</tr>
<tr>
<td>Cu-Pb/Sn</td>
<td>1–3</td>
</tr>
<tr>
<td>Cu-Si</td>
<td>400</td>
</tr>
<tr>
<td>Cu-Kovar</td>
<td>40</td>
</tr>
<tr>
<td>Cu-CuO</td>
<td>1000</td>
</tr>
</tbody>
</table>

4.11 Solder

Low thermal e.m.f. solder is available from Leeds & Northrup. They call it thermal free solder, L & N Part Number 107-1-0-1. It is composed of 70% Cadmium and 30% tin and can be furnished in wire form with a square cross section of 1/16" or a rectangular cross section of .04" x .08". The cost is $35 for a 2 oz kit and instructions are included.

4.11.1 Leads

Leads can cut flux lines and consequently generate e.m.f.'s.

Twisted pair.

Shielded when practical.

4.11.2 Lugs and solder pads

After much effort and expense has been expended in fabricating and calibrating a coil, unrepairable broken leads can be a disaster, especially if there is a restricted time slot available to make measurements.

Interface between coil and leads.

Tie down coil ends—available if leads break.

Highly advisable, not always feasible.

4.11.3 Stress Relief

*ABSOLUTELY ESSENTIAL!*

flexible sleeving—whatever.

4.11.4 Connectors

Non-Magnetic

Don't laugh, a well known manufacturer of Hall-probe Gaussmeters put a magnetic connector on the probe lead of a new instrument.
5. COIL FABRICATION

5.1 Curved line-integral

A pair of integral coils curved to follow the particle trajectory were needed to measure \( \int B \cdot dl \) for the ALS booster dipoles [6]. Curved coils, 1.8-m long, 4-mm effective width, 72 turns were fabricated using 12 strands of AWG #44 multifilar wire.

- Straight forms ground from NEMA G-10.
- Coil winding fixture.
- 6 adjacent windings of multifilar wire.
- Only 11 solder connection for 72 turns.
- Printed circuit board lead tie downs.
- Strongback to provide curvature and rigidity.

This coil had a problem when making AC measurements since one solder joint acted as a rectifier.

CERN in a 1995 internal note also describes the fabrication of coils using multifilar wire [7]. They have designed a special connector for the ends of the multifilar wire coil.

![Diagram of coil winding form and strongback for curved integral coil](image)

Fig. 10 Integral coil winding form  
Fig. 11 Strong back for curved integral coil

5.2 Cryogenic search coil arrays

A cryogenic search coil array has additional problems. It must be able to undergo many cycles from room temperature to cryogenic temperatures without changing its characteristics. In the design and fabrication of cryogenic search coil arrays, one must take into account:

- Generation of thermal e.m.f.’s at connections and junctions.
- The effect of the thermal coefficient of linear expansion on the calibration.
- Bowing introduced by stresses due to lack of mechanical homogeneity or symmetry.
  - Mechanical homogeneity can be improved by fabricating the assembly from laminations laid down with alternating orientations.
  - Mechanical symmetry can be improved by incorporating extra coil windings.
- Stresses on solder joints due to thermal cycling.
- Bearings that operate at cryogenic temperatures.
  - Rulon slides nicely over stainless steel dimples at cryogenic temperatures.
- Heat transfer from room temperature to the magnet.
Several cryogenic dipole arrays and one cryogenic quadrupole search coil array have been fabricated at LBNL, but the latter has been plagued by problems [8]. The quadrupole array consisted of three sets of radial harmonic coils; a center set 10 cm long for measuring the uniform region of quadrupoles, and two sets 71 cm long at each end. Four Rulon bearings were on the search coil, each one sliding on two dimples and each 45° from the bottom center line of the bore tube, gravity holding the coil down on the dimples. At cryogenic temperatures, the coil form was bowing leading to a ~10% change in the gradient transfer function. We rectified this problem by putting in another dimple at the top of the bore tube for each bearing.

Another problem with the cryogenic quadrupole array is that we have lost continuity for two of the nine coils. This problem we have not been able to rectify. Cryogenic coils are much more complicated to fabricate successfully than room temperature coils and are to be avoided if possible.

6. MAPPING EXAMPLE

The block diagram of one of the LBNL Magnetic Measurements Group mapping systems is shown in Fig. 12. Two line-integral coils are in the magnet. The moving mapping coil and an identical coil fixed on the lower pole are connected in series opposition in order to compensate power supply variations. The square loop flux standard is used to calibrate the integrator coil system.

![Block diagram of coil mapping system](image_url)
A map of a B-Factory Low Energy Ring dipole magnet made with the line-integral coils is shown in Fig. 13. Lawrence Berkeley National Laboratory is responsible for the Low Energy Ring which is being built at SLAC (Stanford Linear Accelerator Center). These are the same coils as those described in Section 5, except that new strongbacks were designed so that they held the coils straight instead of curved.

Fig. 13 Mapping signal versus position - BL_{eff} = 0.38 T-m

6.1 Null-mapping technique

This is an example of a null-type map and is an extremely important technique. It should be used whenever feasible to increase sensitivity and resolution. The mapping coil is measuring very small variations of the field with respect to the line-integral field at the center of the magnet. In this particular case there is an "identical" bucking coil, fixed to the bottom pole, connected in series-opposition to the mapping coil, that cancels out signal deviations due to power supply fluctuations.

6.2 Data acquisition procedure

The moving coil is moved to the x = y = 0 position, the integrator is zeroed and then the integrator output and time are recorded by the computer. The coil is then moved to x = -60 mm and from there to x = +60 mm in 2.5 mm increments, and then back to the x = 0 position. At each position the computer records: y-position, x-position, integrator voltage, the time, the output of a hall probe fixed to one pole, the output of a precision current transformer that is monitoring the current, and some various temperatures.

Immediately prior to the map, the computer recorded several fixed parameters such as: date and time, magnet type and serial number, operators names, test current set points, turns width of the line integral coils, the integrator sensitivity, Hall-probe gaussmeter range, precision current-transformer transfer function, etc.

The integrator readings and time at the initial and final x = 0 positions are used to calculate the integrator drift. Knowing the drift vs. time, one corrects the integrator output assuming that the drift was linear. This is a reasonable assumption as long as the time is under a few minutes and the integrator output is large compared to noise such as varying thermal e.m.f.'s.

The result is shown in Fig. 13. Note that there is a small offset of 0.9 μV-seconds at the midpoint x = 0 position. This is an indication of non-linear drift. It is possible to break up the drift correction by using the middle x = 0 position and forcing the output to zero. However, it is
best to let this value be real in order to give an indication of the reality of the data. Note also that
the line integral field in the center is approximately 0.38 Tesla-meter and we are measuring the
deviations of μT·m at 0.38 Tesla-meter as a function of x position.

6.3 Magnitude of line integral field

There are two techniques that can be used to determine the magnitude of the line-integral
field.

6.3.1 Flipping technique

One of the coils is positioned at x = y = 0 and the integrator is zeroed. Time and
integrator output are read, the coil is flipped 180° and data read, flipped back and data read,
flipped 180° and data read, flipped back and data read. The flux linkage for the first three data
points is calculated assuming linear drift and then halved. Similarly for the last three data
points. If the two calculations agree (within some specified accuracy) then the data is accepted,
if the calculations are outside of the necessary accuracy, one redoes the measurement.

6.3.2 Zero Gauss chamber technique

These line-integral coils were initially fabricated with curved strong backs. In this case,
flipping does not work. One must initially put the coils in a “Zero Gauss Chamber” and take
data on the change of flux linkage between the coil in the chamber and the coil in the magnet.

7. CALIBRATION OF SEARCH COILS

The calibration of search coils requires as high a precision and as many methods as
possible or practical without going overboard. Documentation that can be retrieved at some
future date is highly advisable. During the calibration or future use of a coil, discrepancies may
be evident. Good documentation of measurements may make the difference whether a coil can
be used, repaired or thrown out. My experience has also been that if anything can go wrong
during a calibration, it will.

• Make no assumptions!
• Redundant measurements.
• Measure same quantity in different ways.
• Document.
• All instruments with current calibration traceable to a standard.
• Close the loop. Try to make the last measurement of a series the same as the first
measurement.

7.1 Mechanical measurements

• Make mechanical measurements of coil form, prior to winding.
• Check location of fiducial surfaces and/or holes.
  If form is out of tolerance, it is best to discover this prior to winding and calibrating
  coil.
• Integral coils width tolerance.
  Suppose 1 cm width
  If 0.1% accuracy required then, only ±10 μm width tolerance is allowed
  Not easy for long coil! More difficult for pairs!
• After winding
  Attempt to make mechanical measurements on windings.
Width
Thickness
Uniformity
Location with respect to fiducial surfaces and/or holes.

For very large coils, mechanical measurements may be the only data available for use in calibrating the coils!

A combination of mechanical and magnetic measurements may be needed to calibrate a coil. For example, in order to obtain the mean turns-width of a coil, the best method is often to divide the magnetically measured turns-area by the mechanically measured length. Note that for harmonic coils, the location of the coil bundles can be more important than the width of the coil.

7.2 Electrical measurements

![Electrical schematic of search coil](image)

- e = e.m.f. generated by coil.
- L = inductance of coil.
- r = resistance of coil.
- C = distributed capacitance.
- R = resistance coil is loaded with.
- V = externally observed voltage.

Fig. 14 Electrical schematic of search coil

- For measurements of static fields, L and C can be neglected.
- For DC calibrations, we will neglect L and C.

Refer to B. De Raad in the Bibliography for a discussion on transients in coils.

7.2.1 Resistance measurement

- Design specification
  The coil resistance can be calculated from design specifications.

- Tolerances and scatter
  There will be deviations due to copper-wire batch, winding tension, etc.

- Large deviations can be the result of shorted turns or improper number of turns.

Four-wire resistance measurements should be made of the coil, and if the leads are a significant fraction of the coil resistance, separate resistance measurements of the coil and leads should be made. As copper has a significant (0.4% /°C) temperature coefficient of resistance it is advisable to record temperature during these measurements.

- Resistance measurements are very useful to detect a change in coil parameters.
- Inductance measurements may be necessary for coils designed for high frequency or pulsed magnetic field use.
- Capacitance measurements may be necessary for coils designed for high frequency or pulsed magnetic field use.
7.3 Magnetic measurements

![Diagram of analog integrator coil calibration block diagram](image)

![Diagram of digital integrator coil calibration block diagram](image)

Fig. 15 Analog integrator coil calibration block diagram

Fig. 16 Digital integrator coil calibration block diagram

7.3.1 Turns Area

\[ \Delta \Phi (\text{Volt} - \text{seconds}) = A \times \Delta B (\text{Tesla} - \text{meter}^2) \]  

(6)

where:

\( \Delta \Phi \) = change in flux linkage.

\( A \) = effective turns area of coil.

\( \Delta B \) = change in magnetic induction within coil.

The change in flux linkage or magnetic induction is obtained either by flipping the coil 180°, or by bringing the coil from a zero Gauss chamber into the field region.

7.3.2 Things to be concerned about

- Are \( B \) and \( A \) parallel to each other? Poles parallel and symmetrical.
- Field Uniformity. Map field prior to calibration with second NMR probe.
- Field variation during calibration. All measurements referenced to a fixed NMR probe.
- Coil resistance vs. input resistance of integrator.
- Close the loop.
- Is there a discrepancy between the design, mechanically measured, and the magnetically measured value of the turns area, \( A \)?

8. PRECISION AND ACCURACY

*Magnetic field* \( B \).

With an NMR, the magnitude of \( B \) can easily be measured to better than ±0.001% at 1 Tesla.

*Field vs. coil orientation*

The accuracy with which the coil has to be aligned has a cosine dependence.

**Table 8**

Cosine dependence of angular tolerance

<table>
<thead>
<tr>
<th>Precision (%)</th>
<th>Angular tolerance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>8.11°</td>
</tr>
<tr>
<td>0.1</td>
<td>2.56°</td>
</tr>
<tr>
<td>0.01</td>
<td>0.81°</td>
</tr>
</tbody>
</table>
Flux linkage
How good is your integrator and calibration technique?
• 0.1% - easy.
• 0.01% - need to be careful.

Complex coil assemblies
Quadrupole coils. Location of coil bundles more important than the coil area.

8.1 Matching coils
Often, it is more important that two or more coils be matched to each other, than that they have a precise effective turns area. Coils can be matched to a very high precision using a null technique. The coils are connected in series-opposition and flipped together 180° in a uniform field. If the coils are identical, the flux linkage will be zero.

Whenever possible we magnetically measure the turns-area of a coil by putting the coil in a static uniform field and measuring the flux linkage. Other quantities such as the turns-width, etc. may be derived from the turns-area. The procedure that has worked for us with analog integrators is as follows:

• Map the volume that the coil will occupy with an NMR probe while monitoring the field with another fixed position NMR probe.
• Calculate the flux density as a function of position if necessary
• Calibrate integrator
• Insert coil in magnet, zero integrator
• Record integrator output and NMR field
• Flip coil 180°, record integrator output and NMR field
• Flip coil back 180°, record integrator output and NMR field
• Flip coil 180°, record integrator output and NMR field
• Flip coil back 180°, record integrator output and NMR field

Assuming that the five measurements are separated by equal intervals of time, then the second data set can be corrected for integrator drift by the difference between the integrator output of the third and first measurement. Similarly the fourth measurement can be corrected for integrator drift using the third and fifth measurement. This technique incorporates redundant measurements of drift and data enabling one to check the calibration.

8.2 Coil width
The width of short line integral coils can be obtained by dividing the magnetically measured turns area of the coil by the mechanically measured length of the coil.

\[ w = \frac{A}{L} \]  

(7)

Width variation
Figure 17 displays a technique for checking the width variation of an integral coil. A jig should keep the coil at a constant position as the coil is moved longitudinally through the magnet. The output of the coil is fed to an integrator and the integrator output is monitored as the coil traverses the magnet.
8.3 Coil orientation

The coil axis may be found by observing the orientation which has the maximum flux linkage. This is not a very sensitive procedure as it is dependent upon the sine of the angle between the field and the coil axis.

A more precise method, but somewhat more complicated, is to bring the coil from a zero gauss chamber into a uniform field region and look for two mutually perpendicular orientations for which the change in flux linkage is zero.

9. INTEGRATORS

Why integrators? A voltage is generated when a coil of wire is in a changing magnetic field, i.e. the coil's flux linkage is changing. The flux-linkage change may be due to a time varying field or to a change in the position or orientation of the coil with respect to the field.

Let us assume that the coil area vector \( \vec{A} \) and the field vector \( \vec{B} \) are parallel to each other, then the vector dot product becomes

\[
\vec{A} \cdot \vec{B} = AB
\]  

The voltage generated is related to the change in flux linkage \( d\Phi \), the field vector and the coil area vector as:

\[
V(t) = -\frac{d\Phi}{dt} = -\frac{d}{dt} \left( \vec{A} \cdot \vec{B} \right) = A \frac{dB}{dt}
\]  

where

- \( \Phi \) = flux linkage, (Volt-seconds = Tesla-m²)
- \( A \) = the effective coil area, (m²) and \( B \) = magnetic induction, (Tesla).

The effective coil area can also be written

\[
\vec{A} = \sum_{i=1}^{N} \vec{A}_i = A\hat{a} = N\bar{A}_i\hat{a}
\]  

where

- \( A_i \) = the area of each individual turn,
- \( N \) = the number of turns of the coil, and
- \( \bar{A}_i \) = the mean area per turn.

Integrating Eq. (9), the flux linkage is given by;

\[
\Phi = \int V \cdot dt, \text{(Volt - seconds)}
\]  

or

\[ \Phi = A \cdot B \cdot (\text{Tesla} - \text{m}^2) \]  

\[ \text{(12)} \]

or

\[ \int V \cdot dt (\text{Volt} - \text{seconds}) = A \cdot \Delta B (\text{Tesla} - \text{m}^2) \]  

\[ \text{(13)} \]

and where we note that the change in magnetic induction \( B \) is given by the integral over time of the coil output voltage:

\[ \Delta B = \frac{1}{A} \int V(t) \cdot dt \]  

\[ \text{(14)} \]

9.1 Simple RC integrator

![Simple RC integrator](image)

Fig. 18 Simple RC integrator

\[ \Delta V = \frac{1}{RC} \int_{t_1}^{t_2} V_c \cdot dt, \text{ (for } t_2 - t_1 \ll RC) \]  

\[ \text{(15)} \]

Since

\[ V_c = A \cdot \frac{dB}{dt} \]  

\[ \text{(16)} \]

the field change at the coil is

\[ B_2 - B_1 = \frac{RC}{A} \cdot \Delta V \]  

\[ \text{(17)} \]

The simple RC integrator is limited by the rapid decay of the integrator voltage. It is only useful for \( t_2 - t_1 \ll RC \), where, \( RC \) is typically less than 10 seconds.

9.2 Electronic analog integrator (Miller)

The voltage decay time of the Miller integrator is \( GRC \) where \( G \) is the gain of the amplifier. \( G \) is typically greater than \( 10^6 \).

\[ \Delta V = \frac{G}{G+1} \cdot \frac{1}{RC} \int_{t_1}^{t_2} V_c \cdot dt, \text{ (for } t_2 - t_1 \ll GRC) \]  

\[ \text{(18)} \]

since \( G \gg 10^6 \), and
\[
\int V \cdot dt (\text{Volt - seconds}) = A \cdot \Delta B (\text{Tesla - m}^2)
\]  (19)

we get:

\[
\Delta V = \frac{A}{\tau} \cdot \Delta B
\]  (20)

where \(\tau = RC\), time constant of integrator.

Decreasing \(\tau\) will make integrator more sensitive, however, since electronic noise is typically \(\sim 1 \mu V\)-second, the accuracy may be independent of \(\tau\).

![Simplified schematic of an analog integrator](image)

**Fig. 19 Simplified schematic of an analog integrator**

**9.2.1 High frequency response**

- **AC - 100 kHz very reasonable.**

Note that the coil is in series with the integrator resistance.

\[
R_{tot} \rightarrow R + r
\]  (21)

where \(r\) is coil resistance.

- Temperature coefficient of resistance of \(R\) can be very good.
- Temperature coefficient of resistance of copper is NOT! The temperature coefficient of resistance of copper is 0.4% / °C
- Needs to be taken into account for precise measurements.

**9.2.2 Capacitors**

The choice of the type of capacitor is critical.

- **Low leakage**
- **Low dielectric absorption.** Sometimes capacitors don't discharge as readily as theory predicts. A voltage can be observed across a capacitor even after it has been fully discharged. This effect is known as soakage, recovery voltage, or dielectric absorption. (DA).
- **Low Temperature coefficient**
Table 9
Characteristics of Various types of capacitors, All values shown as representative figures.

<table>
<thead>
<tr>
<th>Type</th>
<th>Temperature coefficient (ppm/°C)</th>
<th>Dissipation factor (% at 1 kHz)</th>
<th>Dielectric absorption (%)</th>
<th>Leakage (Ohm-μF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mylar</td>
<td>+400 ± 200</td>
<td>0.5-1</td>
<td>0.5</td>
<td>5x10^9–10^11</td>
</tr>
<tr>
<td>Polystyrene</td>
<td>-120 ± 30</td>
<td>0.01–0.1</td>
<td>0.02</td>
<td>10^{11–10^{12}}</td>
</tr>
<tr>
<td>Polycarbonate</td>
<td>0 ± 100</td>
<td>0.1–0.5</td>
<td>0.2</td>
<td>10^{10–10^{11}}</td>
</tr>
<tr>
<td>Polypropylene</td>
<td>-450 ± 300</td>
<td>0.01–0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parylene</td>
<td>0 ± 50, -200</td>
<td>0.1–0.3</td>
<td>0.1-1</td>
<td>10^{10–10^{12}}</td>
</tr>
<tr>
<td>Teflon</td>
<td>-200</td>
<td>0.1–0.2</td>
<td>0.2</td>
<td>10^{11–10^{12}}</td>
</tr>
<tr>
<td>Glass</td>
<td>+140 ± 25</td>
<td>0.03–0.1</td>
<td></td>
<td>10^9</td>
</tr>
</tbody>
</table>

9.2.3 Bells and whistles

Drift adjust
- Discharge capacitor
  - Momentary switch
  - Standby switch

Gain adjust
- R & C choice

Scale-factor adjust
- R & C choice
- Output divider

Peak reading
Digital meter
Analog meter

9.2.4 Commercial vendors
- Walker Scientific, USA [9].
- Dowty RFL Industries, USA [10].

Many laboratories have their own versions.

9.3 Digital integration
Digital integrators have the following in common:
- A device that transforms a voltage into a frequency, commonly called V/f, (Voltage to frequency converter). This device typically produces rectangular pulses, the pulse rate proportional to the voltage input. The linearity of the devices used for integrators is typically better than 0.01%. Note that commercially available V/f modules are normally
10 V unipolar devices. Modules can be obtained with frequencies up to 10 MHz. The highest linearity devices are 100 kHz.

- A counter to sum the pulses from the V/f.
- Some method of determining the polarity of the input voltage and coding the output correspondingly.

![Diagram](image)

**Fig. 20** Very simplified block diagram of a digital integrator

The V/f works in the following way:

\[ B = \frac{1}{A} \int V(t) \cdot dt \]

(22)

Let

\[ V \rightarrow C \cdot f \]

(23)

\[ \int V \cdot dt \rightarrow \frac{\Delta n_i}{\Delta t_i}, \left( \frac{\text{counts}}{\text{seconds}} \right) \]

(24)

where \( C \) is the V/f transfer function, (Volts/Hz). Then

\[ \int V \cdot dt \rightarrow \sum C \cdot \frac{\Delta n_i}{\Delta t_i} = C \sum \Delta n_i \]

(25)

and hence

\[ B = \frac{C}{A} \sum \Delta n_i \]

(26)

9.3.1 Some positive attributes

- High resolution 1:10⁶
- High precision < 0.01% on higher ranges
- Good for On-The-Fly measurements
- No processing time
- Only latch counter time which is in the sub microsecond region.
- Large dynamic range is possible

9.3.2 A negative attribute

- Performance deteriorates as input frequency increases.

\[ V = \frac{d\Phi}{dt} \]

(27)
For AC signal with amplitude $\Phi_o$:

$$F = F^o \sin wt$$ (28)

$$V = F^o w \cos wt$$ (29)

as frequency increases, the voltage range that the V/f has to go up to increases, and hence the transfer function, Volts/Hz increases.

- We get less counts for signal $\Phi_o$
- Can run into BIT noise.

9.3.3 Polarity algorithms
- Bias input to mid V/f range.

![Fig. 21 Simplified block diagram of biased V/f digital integrator](image)

- Dual unipolar V/f's, one for positive, other for negative voltages.
- Absolute value amplifier.
- Detect input signal polarity.
- Inverts negative signals.
- Shifts V/f output according to polarity.

![Fig. 22 Simplified block diagram of absolute value amplifier V/f digital integrator](image)

9.3.4 V/f bells and whistles
- Input amplifier to increase range.
- Over-range detection.
• Manual and/or remote control.
• Front panel configuration indicators.
• Drift adjust.
• Front panel count rate indicators.

Metrolab—based on CERN design [12].
Uses bias input technique for polarity.
\[ \pm 5 \text{ mV} \rightarrow \pm 5 \text{ V} \text{ in a 1, 2, 5 sequence.} \]

MHz option has resolution of \(10^{-8} \text{ V-s}\).
MHz option has maximum range of 40 V-s.
Motor control
Optical encoder interface
G64, RS-232, GPIB versions, VME soon

9.3.5 LBNL, HERA, FERMILAB, field effects

• Detect input signal polarity.
• Absolute value amplifier—inverts negative signals.
• Shifts V/f output according to polarity. LBNL [13]
• 0.01% accuracy on higher ranges
0.1% \(\rightarrow\) 0.5% on lowest range, a preamplifier in a search coil switching module
improves performance on the lowest ranges.

Input voltage ranges from \(\pm 1 \text{ mV} \rightarrow \pm 10 \text{ V, 1, 2, 5 sequence.}\)

MHz \(\Rightarrow 10^{-9} \text{ V-s resolution}\) the 1 mV range with SCSM input, (100 gain preamp,
atenuators)
\[ \pm 0.1 \text{ mV} \rightarrow \pm 1000 \text{ V,} \]
\[ \Rightarrow 10^{-10} \text{ V-s resolution.} \]

Maximum V-s range dependent upon counter.
The LBNL instrument is packaged as a NIM module with TTL logic.
Control and monitor are independent, manual and remote.

9.3.6 Saclay

• C. Fougéron’s group used discrete component electronics V/f for ESRF series dipole
measurements.
• Made by BARRAS PROVENCE.
• Designed by J. Faure, CNRS in Grenoble.

9.3.7 Schlumberger Model 7061 & 7062 System Voltmeters [14]

• 4 1/2 to 7 1/2 Digit DVM
• 0.1 \(\mu\text{V-s}\) resolution
• At least 10,000 \(\mu\text{V-s}\) capability.
• No real data acquisition on-the-fly. Each trigger resets DVM and has a time dead band
associated with the trigger.
• 0.01% V-s accuracy with option 70616C (fixed frequency clock).
• GPIB interface.
• Up to 8000 readings into memory.
• 1500 readings/second capability into memory.
• Programmable integration time.
• 18 channel multiplexer option.
• Can be used as a Calibrator.

9.3.8 DVM or ADC (Analog to Digital Converter) & Computer

![Diagram of computer style integrator](image)

Fig. 23 Computer style integrator

\[ \int V \cdot dt \rightarrow \sum V_i \Delta t_i \]  \hspace{1cm} (30)

• \( \Delta t \) has to be known very precisely, and

\[ V = \frac{d\Phi}{dt}, \text{ rate of change of flux.} \]  \hspace{1cm} (31)

• BNL uses for harmonic analysis measurements.
• LBNL ALS Insertion Device Group uses computer integration for permanent magnet block measurements.

9.3.9 Analog vs digital

An advantage of digital integration is that there is very high correspondence between a trigger pulse and digital readout.

Digital not too good for AC measurements.

10. SEARCH COIL SWITCHING MODULES (SCSM)

10.1 Multiplexer/scanner

• One choice of input, one output.
• Low thermal e.m.f. relays.

Hewlett Packard 3495A went to extreme of putting mercury wetted reed relays on a separate board from actuators in an attempt to reduce thermal e.m.f.'s.

10.2 Search coil switching module

• Multiplexer
• Sum coils.
Aiding
Opposing
- Preamplifier.
- Attenuator.
- Reconfigurable.
- Computer interface.
- Low thermal e.m.f. relays.
- No commercial vendor that I know of!

Fig. 25 LBNL SCSM

Fig. 26 LBNL SCSM using two integrators

Fig. 27 Universal type search coil switching module
11. MAGNETIC FIELD CALIBRATORS

Zero Field

- Zero Gauss Chambers.
  
  May be cylinders of mu-metal or some other high permeability alloy.

Permanent Magnets

Usually some temperature dependence.

Solenoids and Helmholtz coils

Calibrated theoretically based upon geometry and current.

11.1 Field direction

The Magnaprobe [15].

This is a little device that I feel is indispensable. It gives the field direction and detects very low permeable materials. It consists of a small bar magnet suspended in a gimbal mounting. The Mark I has jeweled bearings and is sensitive enough to align itself with the dip of the earth's magnetic field. The Mark II does not have jeweled bearings and is not quite as sensitive as the Mark I.

11.2 NMR

Nuclear magnetic resonance magnetometers.

The theory of these instruments is covered in the chapter by Claude Reymond. The accuracy of these instruments depends upon how well one can measure frequency. Since frequencies can be measured to very high precision, $10^{-6}$ is quite easy, these instrument are our calibration standard for magnetic fields.

They also have some limitations. They can not be used in fields that have moderate gradients as the signal-to-noise ratio deteriorates rapidly with field gradient. There ability to track field changes automatically is also limited. They are best suited for mapping magnets which have very homogeneous fields and which do not vary in time.

12. VOLTAGE

- For most purposes, modern DVM's are adequate.
- At least once a year, DVM calibrations are checked.

13. VOLT-SECONDS CALIBRATORS

13.1 MTC-1 Fluxmeter Calibrator, Walker Scientific

Puts out rectangular voltage pulse - Volt-seconds & 0.1 V

0.5, 1, 5, 10 seconds

±0.05 % specs

13.2 LBNL Square Loop Flux Standard, "SLUFUS" [16]
Fig. 28  Simplified schematic of square loop flux standard

This instrument is based upon a temperature controlled, temperature compensated, toroidal-core transformer. The toroidal core transformer is kept saturated. A switch reverses the current in the primary which generates constant $\pm V dt$ pulses. Let us remember that 1 Volt-second is the same as 1 Tesla-meter$^2$. A simplified schematic diagram of SLUFUS is shown in Fig. 28. The diagram portrays a primary winding on the core of $N_p$ turns with resistance $R_p$ and a secondary winding of $N_s$ turns with resistance $R_s$. Actually $R_c$ has seven taps for outputs of 0.01 → 1.0 V·s, in a 1, 2, 5, sequence. The core is fabricated from DELTAMAX with a variation of $\Phi_{sat}$ of 0.01%/yr. The core is held in saturation at $H_{op}$. The voltages in the primary circuit $V_p$ and secondary circuit $V_s$ are given by,

$$V_p = N_p \frac{d\Phi}{dt}$$  \hspace{1cm} (32)

$$V_s = N_s \frac{d\Phi}{dt}$$  \hspace{1cm} (33)

where $\Phi$ is the flux in the core.

When the polarity is switched,

$$V_p \tau = \int_0^\tau V_p \cdot dt = 2N_p \Phi_{op}$$  \hspace{1cm} (34)

$$V_s \tau = \int_0^\tau V_s \cdot dt = 2N_s \Phi_{op}$$  \hspace{1cm} (35)

where $\tau$ is the switching time.

The SLUFUS is in the coil-integrator circuit both while calibrating and while making measurements. This ensures that there is no change in the circuit that may effect the calibration. Up to four coils can be connected in series. One has the option of reversing the polarity of a coil. The reversing switch is a spring-loaded push-button switch. One of the neat things about this instrument is that a positive Volt-second pulse is generated when the button is depressed, and an equal amplitude negative pulse is generated when the button is released. This allows one to correct the integrator output voltages for any drift that may occur during the calibrating pulses.

The long term accuracy of the LBNL square loop flux standard is $\pm 0.03\%$.

13.3 Schlumberger system voltmeter

$\pm 0.01\%$
13.4 Custom system rectangular voltage pulse

±0.005% or better

ACKNOWLEDGMENTS

Many of the figures have been created by Dave Van Dyke. Don Nelson and Paul Barale were kind enough to review earlier versions of this work and make valuable suggestions. Kelly Gonzalez was helpful in reformatting material that was published five years ago. Klaus Halbach always made himself available to me for theoretical consultations. I would like to thank LBNL for helping support the effort involved in producing this chapter.

REFERENCES


[4] Corning Glass works, Corning, N.Y., USA.


BIBLIOGRAPHY


Proceedings of Particle Accelerator Conferences.


Low Level Measurements, For Effective Low Current, Low Voltage, and High Impedance Measurements, Keithley Instruments, Inc., 28775 Aurora Road, Cleveland, Ohio 44139, USA.

MWS Wire Industries, 312 Cedar Valley drive, Westlake Village, CA 91362, USA.

Many commercial vendors publish application notes and "How it Works" pamphlets that can be very useful.

Leeds & Northrup, (field offices throughout the USA and the world).


HARMONIC COILS

Animesh K. Jain
RHIC Project, Brookhaven National Laboratory, Upton, New York 11973-5000, USA

Abstract
The basic theory of radial and tangential coils for harmonic analysis is described. A $2m$-pole harmonic coil is discussed as a special case. The expression for flux is generalized for a coil of arbitrary geometry. The effects of transverse and torsional vibrations in the rotational motion are derived, leading to the need for bucking. A generalized bucking algorithm is presented for measuring magnets of different multipolarities with a tangential coil design used at RHIC. Simple, analytical estimates are provided for the errors in harmonic measurements resulting from a variety of random and systematic errors in the construction of measuring coils. Estimates are also provided for errors resulting from imperfect placement of the coil in the magnet. Such errors include a radial offset, sag and tilt of the measuring coil axis relative to the magnet axis. A procedure is given for accurate calibration of the geometric parameters of a five-winding tangential coil.

1. INTRODUCTION

The field in the aperture of a long, relatively straight accelerator magnet can be considered to be two dimensional for most practical purposes. Such a field is generally described in terms of a harmonic expansion, as discussed in detail in another chapter in these proceedings [1]. The field quality in accelerator magnets is expressed in terms of the magnitude of undesirable harmonic terms in this expansion. Rotating coils, in which a loop of wire is rotated in the aperture of the magnet, provide the most convenient means to accurately measure these harmonic terms. Since these coils are used for harmonic measurements, such rotating coils are also referred to as harmonic coils.

The radial and azimuthal components of a two dimensional field in a current free region can be written as:

$$B_r(r, \theta) = \sum_{n=1}^{\infty} C(n) \left( \frac{r}{R_{\text{ref}}} \right)^{n-1} \sin[n(\theta - \alpha_n)]$$  \hspace{1cm} (1)

$$B_\theta(r, \theta) = \sum_{n=1}^{\infty} C(n) \left( \frac{r}{R_{\text{ref}}} \right)^{n-1} \cos[n(\theta - \alpha_n)]$$  \hspace{1cm} (2)

where $C(n)$ and $\alpha_n$ are constants and $R_{\text{ref}}$ is an arbitrary reference radius, typically chosen to be 50-70\% of the magnet aperture. The corresponding Cartesian components are most
conveniently described in terms of a complex field, \( B(z) \), defined as a function of the complex variable, \( z = x + iy = r \cdot \exp(i\theta) \), as:

\[
B(z) = B_y(x, y) + iB_x(x, y) = \sum_{n=1}^{\infty} [C(n)\exp(-in\alpha_n)]\left(\frac{z}{R_{\text{ref}}}\right)^{n-1} 
\]

(3)

It should be noted that all complex quantities are denoted by a bold and italic font in this paper. The real and the imaginary parts of the expansion coefficients in Eq. (3) are the normal and skew components respectively of the \( 2n \)-pole term. In view of the different notations for the normal and skew components currently in use, we shall refrain from explicitly using these components. Accordingly, all equations in this chapter are written in terms of the parameters \( C(n) \) and \( \alpha_n \). The two dimensional field is completely characterized in the aperture of the magnet by a set of the field parameters \( C(n) \) and \( \alpha_n \) (or equivalently, the normal and the skew components). The purpose of harmonic coils is to measure these field parameters in a magnet.

The basic principle of a harmonic coil is very simple. As a loop of wire is rotated in the field, a voltage signal is induced in the loop. The nature of this signal provides information on the angular dependence of the field, and hence its harmonic components according to Eqs. (1) and (2). Based on the geometry of the loop, there are two main types of harmonic coils—radial coils and tangential coils. The fundamental equations governing the use of such coils are described in Secs. 2 and 3. Special coil geometries, designed to be sensitive to selected harmonics, are treated in Sec. 4. Subsequent sections deal with the effects of imperfections in the rotational motion, random and systematic errors in coil construction and errors in aligning the coil axis with the magnet axis during measurements. Much of the material presented here follows closely the treatment in references [2] and [3]. The subject of harmonic coils is also covered in detail in references [4] to [6].

2. THE RADIAL COIL

A cross section of a radial coil is shown schematically in Fig. 1. A radial coil has a flat loop of wire whose plane coincides with a radial plane of the rotating cylinder. More specifically, the two sides of the loop and the rotation axis lie in the same plane. The two sides of the loop are located at radii \( R_1 \) and \( R_2 \), as shown in Fig. 1. Such a coil is typically constructed by winding the loop on a flat bobbin, and then sandwiching it between the two halves of the rotating cylinder [6,7]. A radial coil is sensitive to the azimuthal component of the field, \( B_\theta \), as shown in Fig. 1. The flux, \( \Phi_{\text{radial}}(\theta) \), through the coil at any angular orientation, \( \theta \), can be obtained by integrating the \( B_\theta \) component given by Eq. (2) over the region covered by the radial coil:

\[
\Phi_{\text{radial}}(\theta) = NL \int_{R_1}^{R_2} B_\theta(r, \theta) \, dr = \sum_{n=1}^{\infty} \frac{NLR_{\text{ref}}}{n} \left[ \left( \frac{R_2}{R_{\text{ref}}} \right)^n \right] \left[ \left( \frac{R_1}{R_{\text{ref}}} \right)^n \right] C(n) \cos(n\theta - n\alpha_n) 
\]

(4)

where \( N \) is the number of turns and \( L \) is the length of the coil along the magnet axis. The above derivation assumes that the two sides of the coil loop are located on the same side of the rotation axis, as shown in Fig. 1. If the two sides are located on opposite sides of the
rotation axis, as is true for many practical coils, then one should replace $R_1$ by $-R_1$. If the coil rotates with an angular velocity $\omega$ and $\theta = \delta$ is the angular position at time $t = 0$, then the flux as a function of time is given by:

$$
\Phi_{\text{radial}}(t) = \sum_{n=1}^{\infty} \frac{NLR_{\text{ref}}}{n} \left[ \left( \frac{R_2}{R_{\text{ref}}} \right)^n - \left( \frac{\pm R_1}{R_{\text{ref}}} \right)^n \right] C(n) \cos(n \omega t + n \delta - n \alpha_n) \quad (5)
$$

where the sign of $R_1$ depends on whether the rotation axis is outside the region of the coil or inside it. Assuming a constant angular rotation speed, $\omega$, the voltage signal induced in a radial coil is given by:

$$
V_{\text{radial}}(t) = -\left( \frac{\partial \Phi_{\text{radial}}}{\partial t} \right) = \sum_{n=1}^{\infty} NLR_{\text{ref}} \omega \left[ \left( \frac{R_2}{R_{\text{ref}}} \right)^n - \left( \frac{\pm R_1}{R_{\text{ref}}} \right)^n \right] C(n) \sin(n \omega t + n \delta - n \alpha_n) \quad (6)
$$

A Fourier analysis of the voltage signal can give the amplitude, $C(n)$, and phase, $\alpha_n$, of the $2n$-pole component of the field, provided the geometric parameters of the coil are known. The amplitude of the voltage signal is proportional to the angular velocity. If the angular velocity fluctuates during the rotation, the signal will be distorted and will give rise to spurious harmonics. For analysis based on voltage signals, it is essential to control the angular velocity and make corrections for any speed fluctuations. One way to overcome the dependence on angular speed is to integrate the voltage signal to get a signal proportional to the flux, which is independent of the angular speed:

$$
\int V_{\text{radial}}(t) \, dt \equiv -\Phi_{\text{radial}}(\theta) = -\sum_{n=1}^{\infty} \frac{NLR_{\text{ref}}}{n} \left[ \left( \frac{R_2}{R_{\text{ref}}} \right)^n - \left( \frac{\pm R_1}{R_{\text{ref}}} \right)^n \right] C(n) \cos(n \theta - n \alpha_n) \quad (7)
$$

In the above equation, we have ignored the constant of integration, since any constant term is of no significance from the point of view of a Fourier analysis. As the coil rotates, the voltage

---

**Fig. 1** Cross section of a radial coil.
signal is integrated using an integrator, which is reset at \( \theta = 0 \). The output of the integrator is read out at several angular positions during the rotation. Typically, 128 or 256 angular positions are used to carry out fast, accurate Fourier transforms to obtain harmonics of interest in accelerator physics \( n \leq 15 \). A more detailed account of the techniques can be found in an earlier review of the subject [5].

3. THE TANGENTIAL COIL

A cross section of a tangential coil is shown schematically in Fig. 2. A tangential coil also has a loop of wire, similar to the radial coil. However, the plane of this loop is arranged to be normal to the radial vector drawn from the rotation axis to the center of the coil, as shown in Fig. 2. The two sides of the loop are equidistant from the rotation axis, and are located at a radius of \( R_c \). Such a coil is sensitive to the radial component, \( B_r \), of the magnetic induction. The flux, \( \Phi_{\text{tang.}}(\theta) \), through such a coil at any angular position, \( \theta \), is:

\[
\Phi_{\text{tang.}}(\theta) = NL \int_{\theta-\Delta/2}^{\theta+\Delta/2} B_r(R_c, \theta) R_c d\theta = \sum_{n=1}^{\infty} \frac{2NL_{\text{ref}}}{n} \left( \frac{R_c}{R_{\text{ref}}} \right)^n \sin \left( \frac{n\Delta}{2} \right) C(n) \sin(n\theta - n\alpha_n) \tag{8}
\]

where \( N \) is the number of turns, \( L \) is the length of the coil along the magnet axis and \( \Delta \) is the angle subtended by the coil at the rotation axis, commonly referred to as the opening angle. If the coil rotates with an angular velocity \( \omega \) and \( \theta = \delta \) is the angular position at time \( t = 0 \), then the flux as a function of time is given by:

\[
\Phi_{\text{tang.}}(t) = \sum_{n=1}^{\infty} \frac{2NL_{\text{ref}}}{n} \left( \frac{R_c}{R_{\text{ref}}} \right)^n \sin \left( \frac{n\Delta}{2} \right) C(n) \cos(n\omega t + n\delta - n\alpha_n) \tag{9}
\]

Assuming a constant angular rotation speed, \( \omega \), the voltage signal induced in a tangential coil is given by:

\[
V_{\text{tang.}}(t) = -\left( \frac{\partial \Phi_{\text{tang.}}}{\partial t} \right) = -\sum_{n=1}^{\infty} 2NL_{\text{ref}} \omega \left( \frac{R_c}{R_{\text{ref}}} \right)^n \sin \left( \frac{n\Delta}{2} \right) C(n) \cos(n\omega t + n\delta - n\alpha_n) \tag{10}
\]

Similar to the case of the radial coil discussed in Sec. 2, a Fourier analysis of the voltage signal can give the amplitude, \( C(n) \), and phase, \( \alpha_n \), of the \( 2n \)-pole component of the field, provided the geometric parameters of the coil are known. The remarks made in Sec. 2 concerning the angular speed fluctuations are equally applicable to tangential coils. The integrated voltage signal is proportional to the flux, and is independent of the angular speed:

\[
\int V_{\text{tang.}}(t) dt \equiv -\Phi_{\text{tang.}}(\theta) = -\sum_{n=1}^{\infty} 2NL_{\text{ref}} \left( \frac{R_c}{R_{\text{ref}}} \right)^n \sin \left( \frac{n\Delta}{2} \right) C(n) \sin(n\theta - n\alpha_n) \tag{11}
\]

The radius, \( R_c \), of the coil should be maximized to get good signal strength for higher harmonics. The opening angle, \( \Delta \), should be large enough to give good signal and small
enough so that \( \sin(n\Delta/2) \) does not vanish for higher harmonics of interest \((\Delta \ll 2\pi/n_{max})\). For most practical purposes, \( \Delta \sim 15 \) degrees can be considered optimum. A coil with an opening angle of only 10 degrees is more sensitive to harmonics higher than 28-pole as compared to a coil with an opening angle of 15 degrees, but at the expense of sensitivity for lower order terms. However, such high order multipoles are of limited interest for most accelerator physics studies. On the other hand, a coil with \( \Delta = 20 \) degrees is more sensitive than a coil with \( \Delta = 15 \) degrees for lower order harmonics, but the sensitivity begins to decline after only the 18-pole term.

4. MULTIPOLE COILS

In general, the radial and the tangential coils described in the previous sections are sensitive to all harmonics. Sometimes, it may be advantageous to design a coil which is sensitive to only some multipoles. The most common examples of multipole coils are the dipole and the quadrupole coils used for bucking (see Sec. 7).

4.1 A dipole coil

Let us first consider the special case of the simplest multipole coil—the dipole coil. A dipole coil has a flat loop of wire arranged in such a way that the rotation axis passes through the center of the loop, as shown in Fig. 3(a). Such a coil has a Dipole Symmetry, namely an antisymmetry under rotation by \( \pi \) radians. The flux through this coil can be calculated by treating it as a radial coil with \( R_1 = -R_c \) and \( R_2 = +R_c \), oriented at an angle \( \theta \), as illustrated in Fig. 3(a). The flux through the coil can also be calculated by treating it as a tangential coil with an opening angle of \( \pi \) radians, oriented at an angle of \( \theta' = \theta + \pi/2 \). Both approaches give the same expression for the flux:

\[
\Phi_{\text{Dipole}}(\theta) = \sum_{n=1}^{\infty} \frac{2NLR_{\text{ref}}}{n} \left( \frac{R_c}{R_{\text{ref}}} \right)^n C(n) \cos(n\theta - n\alpha_n) \tag{12}
\]
It should be noted that the terms for even multipoles vanish in this particular geometry. A dipole coil is therefore sensitive to only the odd harmonics, i.e., dipole, sextupole, decapole, etc. Such a coil is almost universally used in both radial and tangential coil systems for bucking the main dipole field in a dipole magnet.

### 4.2 A multipole coil of order \( m \)

A multipole coil of order \( m \), or a \( 2m \)-pole coil, is a generalization of the dipole coil described above. Such a coil, shown schematically in Fig. 3(b), has \( m \) loops connected in series. For any angular position characterized by the angle \( \theta \), the loops span the angular regions of \( \theta \) to \( \theta+\pi/m \), \( \theta+2\pi/m \) to \( \theta+3\pi/m \), \( \theta+4\pi/m \) to \( \theta+5\pi/m \), and so on. The flux through such a coil as a function of \( \theta \) can be easily calculated by treating it as an array of \( m \) identical tangential coils with opening angle of \( \Delta = \pi/m \) and having angular positions of \( \theta' = \theta+(\pi/2m) \), \( \theta' + 2\pi/m \), \( \theta' + 4\pi/m \) and so on.

The flux through the first segment of the coil, covering the angular region \( \theta \) to \( \theta + \pi/m \), is:

\[
\Phi_1(\theta) = \Phi_{\text{tang}}(\theta') = \sum_{n=1}^{\infty} \frac{2NLR_{\text{ref}}}{n} \left( \frac{R_c}{R_{\text{ref}}} \right)^n \sin \left( \frac{n\pi}{2m} \right) C(n) \sin(n\theta' - n\alpha_n) = \text{Im} \sum_{n=1}^{\infty} \frac{2NLR_{\text{ref}}}{n} \left( \frac{R_c}{R_{\text{ref}}} \right)^n \sin \left( \frac{n\pi}{2m} \right) C(n) \exp[i(n\theta' - n\alpha_n)] \tag{13}
\]

The total flux through the array of loops is obtained by summing the contributions from all the segments located at angular intervals of \( 2\pi/m \):
\[
\Phi_{2m\text{-pole}}(\theta) = \text{Im} \sum_{n=1}^{\infty} X_n e^{in\theta} \left[ 1 + e^{i2\pi n/m} + e^{i4\pi n/m} + \ldots (m \text{ terms}) \right] \\
= \text{Im} \sum_{n=1}^{\infty} X_n e^{in\theta} \left[ \frac{1 - \exp(i2n\pi)}{1 - \exp(i2n\pi / m)} \right]
\]  
(14)

where,

\[
X_n = \frac{2NLR_{\text{ref}}}{n} \left( \frac{R_c}{R_{\text{ref}}} \right)^n \sin\left( \frac{n\pi}{2m} \right) C(n) \exp(-in\alpha_n)
\]  
(15)

Since \( n \) is always an integer, the numerator of the quantity in square brackets in Eq. (14) is always zero. For integral values of \((n/m)\), the denominator also vanishes and the quantity in square brackets has a limiting value of \( m \). However, for even values of \((n/m)\), the contribution again vanishes as the quantity \( X_n \) in Eq. (15) becomes zero due to the \( \sin(n\pi/2m) \) factor. Therefore, all the harmonic terms in the summation vanish, except for those values of \( n \) which are odd multiples of \( m \). For example, a quadrupole coil \((m = 2)\) will only be sensitive to the quadrupole \((n = 2)\), the dodecapole \((n = 6)\), the 20-pole \((n = 10)\), etc. terms. The total flux for the \(2m\)-pole coil can be written as:

\[
\Phi_{2m\text{-pole}}(\theta) = \sum_{n=m}^{\infty} \frac{2mNLR_{\text{ref}}}{n} \left( \frac{R_c}{R_{\text{ref}}} \right)^n \frac{C(n)\cos(n\theta - n\alpha_n)}{n} 
\]  
(16)

where \( k \) is any integer, including zero. If the coil rotates with an angular velocity \( \omega \) and \( \theta = \delta \) is the initial angular position of the coil, then \( \theta = \omega t + \delta \). The flux and the voltage at any time are given by:

\[
\Phi_{2m\text{-pole}}(t) = \sum_{n=m}^{\infty} \frac{2mNLR_{\text{ref}}}{n} \left( \frac{R_c}{R_{\text{ref}}} \right)^n \frac{C(n)\cos(n\omega t + n\delta - n\alpha_n)}{n} 
\]  
(17)

\[
V_{2m\text{-pole}}(t) = \sum_{n=m}^{\infty} \frac{2m\omega NLR_{\text{ref}}}{n} \left( \frac{R_c}{R_{\text{ref}}} \right)^n \frac{C(n)\sin(n\omega t + n\delta - n\alpha_n)}{n} 
\]  
(18)

The results for a dipole coil are obtained by putting \( m = 1 \). Quadrupole coils \((m = 2)\) are commonly used for bucking in tangential coil systems. Sextupole and other higher order coils may be used for special applications where it may be necessary to monitor or measure specific harmonics.
5. GENERAL TREATMENT OF ROTATING COILS

So far we have considered rotating coils of a specific geometry, such as the radial and the tangential coils. Even though the construction methods of these coils differ significantly due to the nature of their geometries, the underlying physics is quite similar for both types of coils. Consequently, it is advantageous to develop a formalism for a generalized coil shape. The radial and the tangential coils can then be treated as special cases of the general coil.

5.1 Flux through a coil of arbitrary shape

We consider a coil made up of a loop of wire running parallel to the Z-axis as shown in Fig. 4. The shape of the coil is defined by a path from the point \( z_1 \) to \( z_2 \) in the complex plane. The length of the loop is \( L \) along the negative Z-axis, as shown in Fig. 4.

Let \( dr \) be an element along the path from \( z_1 \) to \( z_2 \). The element of area, \( ds \), defined by this line element is given by

\[
ds = |dr|L = (\hat{k} \times dr)L = \hat{k} \times (\hat{x} \, dx + \hat{y} \, dy) \, L = (-dy \, \hat{x} + dx \, \hat{y}) \, L
\]  

(19)

The flux through this area element \( ds \) is given by

\[
d\Phi = B \cdot ds = (B_x \hat{x} + B_y \hat{y}).(-dy \, \hat{x} + dx \, \hat{y})L = (B_y \, dx - B_x \, dy)L = \text{Re}[B(z) \, dz]
\]  

(20)

where \( B(z) = B_1 + iB_z \) is the complex field defined in Eq. (3). This leads us to the general result for the flux through a loop with \( N \) turns:

\[
\Phi = NL \text{Re} \left[ \int_{z_1}^{z_2} B(z) \, dz \right] = \text{Re} \sum_{n=1}^{\infty} C(n) \exp(-in\alpha_n) \left( \frac{NLR_{\text{ref}}}{n} \right) \left( \frac{z_2}{R_{\text{ref}}} \right)^n - \left( \frac{z_1}{R_{\text{ref}}} \right)^n \right] \]  

(21)

![Fig. 4 Calculation of flux through a coil of arbitrary shape.](image-url)
In writing Eq. (21), use has been made of the multipole expansion of the complex field, \( B(z) \), given by Eq. (3). For two-dimensional fields, the total flux depends only on the end points, \( z_1 \) and \( z_2 \), and not on the actual path followed by the loop between these points. This does not hold true for three dimensional fields in general, where \( B(z) = B_y + iB_z \) is no longer an analytic function of \( z \).

### 5.2 A rotating coil of arbitrary shape

Let us consider a rotating coil of arbitrary shape formed by a loop of wire passing through two arbitrary points in the X-Y plane, as shown in Fig. 5. In general, both the radial and the azimuthal coordinates of these two points will be different. The *radial coil* is a special case where the azimuthal coordinates of both the points are either the same, or differ by \( \pi \). Similarly, the *tangential coil* is a special case where the radial coordinates of the two points are the same. Any angular position of the coil is characterized by an angle \( \theta \) measured from the “initial position” \((z_{1,0}, z_{2,0})\) of the wires. If \( z_1 \) and \( z_2 \) denote the locations of the two points in the complex plane at any instant, then the flux through the coil of length \( L \) and with \( N \) turns is given by Eq. (21). From Fig. 5,

\[
z_1 = z_{1,0} \exp(i\theta); \quad z_2 = z_{2,0} \exp(i\theta)
\]

Substituting in Eq. (21), we get,

\[
\Phi(\theta) = \text{Re} \left[ \sum_{n=1}^{\infty} K_n \exp(in\theta) C(n) \exp(-in\alpha_n) \right]
\]

where \( K_n \) is the *sensitivity factor* of the coil for the harmonic of order \( n \) and is given by:

\[
K_n = \left( \frac{NLR_{\text{ref}}}{n} \right) \left\{ \left( \frac{z_{2,0}}{R_{\text{ref}}} \right)^n - \left( \frac{z_{1,0}}{R_{\text{ref}}} \right)^n \right\}
\]

![Fig. 5 A rotating coil of arbitrary shape](image-url)
5.3 Radial and tangential coils as special cases

For a radial coil, when $\theta$ is measured from the X-axis, as shown in Fig. 1, $z_{1,0} = \pm R_1$ and $z_{2,0} = R_2$. It should be noted that $z_{1,0} = +R_1$ when $R_1$ and $R_2$ are on the same side of the center and $z_{1,0} = -R_1$ when $R_1$ and $R_2$ are on the opposite sides of the center. This gives,

$$K_{n_{radial}}^r = \frac{NLR_{ref}}{n} \left[ \left( \frac{R_2}{R_{ref}} \right)^n - \left( \frac{\pm R_1}{R_{ref}} \right)^n \right]$$  \hspace{1cm} (25)

The sensitivity factor for a radial coil is purely real in this case. Substituting in Eq. (23), we get the same expression for flux through a radial coil as was derived earlier in Eq. (4).

For a tangential coil, where $\theta$ denotes the angular position of the coil center measured from the X-axis (see Fig. 2), we have,

$$z_{1,0} = R_c \exp(i\Delta/2); \quad z_{2,0} = R_c \exp(-i\Delta/2) \hspace{1cm} (26)$$

$$K_{n_{tangential}}^t = -i \frac{2NLR_{ref}}{n} \left( \frac{R_c}{R_{ref}} \right)^n \sin \left( \frac{n\Delta}{2} \right)$$  \hspace{1cm} (27)

The sensitivity factor for a tangential coil is purely imaginary. It can be easily verified that the expression for flux in Eq. (23) reduces in this case to that derived earlier in Eq. (8).

5.4 An array of rotating coils

Let us consider an array of $M$ coils mounted on the same rotating system. Let the sensitivity factor of the $j$-th coil for the $n$-th harmonic be denoted by $K_{n_{(j)}}^{(j)}$, $j = 1,2,3,\cdots M$. Let all these coils be connected either in series, or in opposition, to generate a combined signal. The total flux through this array of coils is the algebraic sum of the fluxes through individual coils:

$$\Phi_{Array}(\theta) = \sum_{j=1}^{M} \beta_j \Phi_j(\theta)$$  \hspace{1cm} (28)

where $\beta_j = +1$ if the $j$-th coil is connected in series, and $\beta_j = -1$ if the $j$-th coil is connected in opposition. In fact, one may consider a somewhat more general situation where the signals from the various coils are summed using a resistor/amplifier system. In this case, the flux corresponding to the resultant signal is still given by Eq. (28), where $\beta_j$ are the appropriate amplification factors. From the general formula for the flux through an individual coil, we obtain:

$$\Phi_{Array}(\theta) = \sum_{j=1}^{M} \beta_j \left[ \text{Re} \sum_{n=1}^{\infty} K_{n_{(j)}}^{(j)} \exp(i\theta) C(n) \exp(-i\pi n) \right]$$  \hspace{1cm} (29)
If all the amplification factors, $\beta_j$, are purely real, then the expression for the total flux can be rearranged as:

$$\Phi_{\text{Array}}(\theta) = \sum_{j=1}^{M} \beta_j \left[ \Re \sum_{n=1}^{\infty} K_n^{(j)} C(n) \exp(-in\alpha_n) \exp(in\theta) \right]$$

$$= \Re \sum_{n=1}^{\infty} \left( \sum_{j=1}^{M} \beta_j K_n^{(j)} \right) C(n) \exp(-in\alpha_n) \exp(in\theta)$$

$$\equiv \Re \sum_{n=1}^{\infty} K_n^{(\text{Array})} C(n) \exp(-in\alpha_n) \exp(in\theta)$$

From Eq. (30), the sensitivity factor of the overall array to the $2n$-pole term is:

$$K_n^{(\text{Array})} = \sum_{j=1}^{M} \beta_j K_n^{(j)}, \quad (\text{All } \beta_j \text{ must be real})$$

Thus, the overall sensitivity factor of the array is given by the algebraic sum of the sensitivities of individual coils in the array, provided the weight factors are purely real. In other words, any amplifier system used with the coil signals should not introduce a phase shift in the signals. If the individual coils are properly designed and $\beta_j$ are appropriately chosen, the overall sensitivity of an array of coils to a particular harmonic (or a set of several harmonics) can be made zero. This is the principle used in the design of bucking windings discussed later in Sec. 7.

6. IMPERFECTIONS IN COIL ROTATION: THE NEED FOR BUCKING

The expressions derived so far for the coil signals have assumed a perfect rotational motion. Such a perfect motion implies a rotation axis which is fixed and steady in space (no transverse vibrations), and no error in the angular positions at which the coil signals are to be sampled (no torsional vibrations). Since small amounts of such vibrations will invariably be present in any practical system, it is important to understand their effect on the measurement of harmonics. Typically, the undesired harmonics in an accelerator magnet are at the level of $\leq 10^{-4}$ of the fundamental field. Clearly, accurate measurement of these harmonics requires utmost care in eliminating the effects of imperfections in a practical system. It should be noted that while an irregular rotation speed is also an imperfection (and generally undesirable), one can, in principle, eliminate its effect either by integrating the voltage signal [see Eqs. (7) and (11)], or by measuring the instantaneous rotation speed and scaling the observed signal to a fixed rotation speed.

6.1 Transverse vibrations of the rotation axis

Let us consider a rotating coil of an arbitrary shape whose rotation axis has a displacement as the coil rotates. Such a transverse motion of the rotation axis is depicted schematically in Fig. 6. In the figure, the origin is assumed to be at the position of the rotation axis when the loop of the wire is at its initial location given by the points ($z_{1,0}, z_{2,0}$). The origin of the coordinate system is assumed fixed in space. As the coil rotates through an angle $\theta$, the
Fig. 6 Transverse vibrations of the rotation axis. As the coil rotates by an angle θ from the initial position given by \((z_{1,0}, z_{2,0})\), it gets displaced from the ideal position (shown by the thick dashed line) by \(D(θ)\).

axis of the coil undergoes a displacement \(D(θ)\) in the complex plane. Consequently, the positions of the two sides of the wire loop in the fixed coordinate system are given by

\[
z_1 = z_{1,0} \exp(iθ) + D(θ); \quad z_2 = z_{2,0} \exp(iθ) + D(θ)
\]

instead of the expressions in Eq. (22). The angular dependence of the flux through the coil can be obtained by substituting from Eq. (32) into Eq. (21). Let us examine the implications of the displacement \(D(θ)\) on the harmonic content of the measured flux through the coil. We shall assume that there is only one dominant harmonic in the magnetic field, as is indeed the case for most accelerator magnets.

### 6.1.1 A nearly pure dipole field

In a nearly pure dipole field, the expression for flux can be approximated by the \(n = 1\) term in Eq. (21). The flux at any angular position, \(θ\), is given by

\[
Φ(θ) ≈ \text{Re}[NLC(1)\exp(-iα_1)(z_2 - z_1)] = \text{Re}[NLC(1)\exp(-iα_1)(z_{2,0} - z_{1,0})\exp(iθ)]
\]

The flux in this case depends only on the quantity \((z_2 - z_1)\), which is independent of the displacement, \(D(θ)\). Thus, the flux linked through a coil in a pure dipole field is unaffected by transverse displacements of the rotation axis. This result is not too surprising because a pure dipole field is uniform in space and displacements in such a field do not produce any feed down harmonics.

### 6.1.2 A nearly pure \(2n\)-pole field

Let us now consider the measurement of harmonics in a nearly pure \(2n\)-pole field magnet \((n ≥ 2)\), such as a quadrupole, sextupole, etc. Once again, only one harmonic term in
the summation in Eq. (21) needs to be retained. An approximate expression for the flux in a pure 2n-pole field can be obtained by using a binomial expansion and neglecting terms of the second and higher orders in \[|D(\theta)/R_{ref}|\), assuming that \(|D(\theta)| \ll R_{ref}\). We get:

\[
\Phi_n(\theta) \approx \text{Re}\left[K_n e^{in\theta} C(n)e^{-\imath n\alpha_n}\right] + \text{Re}\left[K_{n-1} e^{i(n-1)\theta} \left\{\frac{(n-1)D(\theta)}{R_{ref}}\right\} C(n)e^{-\imath n\alpha_n}\right] + \cdots \tag{34}
\]

The first term in Eq. (34) represents the flux in the absence of a displacement. The flux picked up by a rotating coil in a pure 2n-pole field is, in general, affected by transverse displacements of the rotation axis. To a first approximation, the effect is proportional to the amplitude of the displacement, \(|D(\theta)|\), and the sensitivity, \(K_{n-1}\), of the coil to the 2(n-1)-pole term. It should be noted that the highest exponent of \(D(\theta)\) in the expression for the flux from a 2n-pole field is \((n - 1)\). Thus, only the first term in the above expression survives for a pure dipole field, whereas the first two terms represent the complete expression for flux in a pure quadrupole field. For fields of higher multiplicities, other higher order terms are also present, but can be neglected since it is expected that the condition \(|D(\theta)| \ll R_{ref}\) will be satisfied for any decent measuring system.

The effect of transverse vibrations can be practically eliminated by making the sensitivity to the 2(n-1)-pole term, \(K_{n-1}\), equal to zero when only the 2n-pole term is the dominant term. This is the basis for bucking the 2(n-1)-pole term in the measurement of a 2n-pole magnet. This bucking can be achieved, for example, by an array of coils. It should be noted that transverse vibrations make precise measurement of the 2(n-1)-pole term difficult in a 2n-pole magnet because the sensitivity, \(K_{n-1}\), cannot be made zero. Such measurements are of interest, for example, in the determination of the magnetic axis. One must minimize the amplitude, \(D(\theta)\), of the transverse vibrations, or use a non-rotating coil system for this purpose.

### 6.1.3 Spurious harmonics due to periodic transverse motion

Let us assume that the displacement amplitude, \(D(\theta)\), is a periodic function of \(\theta\). This means that the coil follows the same motion in every rotation cycle. Such a periodic displacement may be expressed in terms of its Fourier components as,

\[
D(\theta) = \sum_{p=-\infty}^{\infty} D_p \exp(i p \theta) \tag{35}
\]

The expression for flux in a pure 2n-pole field is given by Eq. (34). Substituting for \(D(\theta)\) from Eq. (35) into Eq. (34) and using the identity

\[
\text{Re} \sum_{p=n}^{\infty} D_{p} e^{-i(p-n+1)\theta} K_{n-1} C(n)e^{-\imath n\alpha_n} = \text{Re} \sum_{p=n}^{\infty} D_{p}^* e^{i(p-n+1)\theta} K_{n-1}^* C(n)e^{\imath n\alpha_n} \tag{36}
\]

the expression for flux in this case can be written as
\[ \Phi_n(\theta) = \text{Re} \left[ K_n e^{i n \theta} C(n) e^{-i n \alpha_n} \right] + \text{Re} \left[ \sum_{p=-n+2}^{n-1} K_{n-1} e^{i(p+n-1)\theta} \frac{(n-1)D_p}{R_{\text{ref}}} C(n) e^{-i n \alpha_n} \right] \]

\[ + \text{Re} \left[ K_{n-1} \frac{(n-1)D_{n+1}}{R_{\text{ref}}} C(n) e^{-i n \alpha_n} \right] + \text{Re} \left[ \sum_{p=n}^{\infty} K_{n-1} e^{i(p-n+1)\theta} \frac{(n-1)D^*_p}{R_{\text{ref}}} C(n) e^{i n \alpha_n} \right] \]  \hspace{1cm} (37)

The first term on the right hand side of Eq. (37) is the 2n-pole term in the absence of vibrations. The other terms are spurious terms of other multipoles. The amount of spurious 2m-pole harmonic in the measured flux is given by:

\[ C'(m)e^{-i m \alpha_m} \approx (n-1) \left[ \frac{K_{n-1}}{K_m} \left( \frac{D_{m-(n-1)}}{R_{\text{ref}}} \right) C(n)e^{-i n \alpha_n} + \frac{K_{n-1}^*}{K_m} \left( \frac{D_{m-(n-1)}^*}{R_{\text{ref}}} \right) C(n)e^{i n \alpha_n} \right] \]  \hspace{1cm} (38)

For a pure sin(p \theta) displacement, only \( D_p \) and \( D_{-p} \) are non-zero. In a nearly pure 2n-pole field, it follows from Eq. (38) that only the harmonics \( m = (p+n-1), (p-n+1), \) and \( (n-p-1) \) are affected by such vibrations.

### 6.2 Torsional vibrations of the rotation axis

Let us now consider a type of rotational imperfection where the position of the rotating coil corresponding to angular position \( \theta \) is not at \( \theta \), but at an angle of \( \theta + T(\theta) \), as shown in Fig. 7. Such an imperfection may result either from an actual torsional vibration of the rotating coil, or it could be due to errors in the triggering of the data acquisition by the angle encoder. The position of the coil as a function of the angular parameter, \( \theta \), is given by:

\[ z_1 = z_{1,0} \exp[i \theta + iT(\theta)]; \quad z_2 = z_{2,0} \exp[i \theta + iT(\theta)] \]  \hspace{1cm} (39)

Fig. 7  Torsional vibrations of the rotation axis. As the coil supposedly rotates by an angle \( \theta \) from the initial position given by \((z_{1,0}, z_{2,0})\), it is actually off from the ideal position (shown by the thick dashed line) by an angle \( T(\theta) \).
6.2.1 Case of a nearly pure $2n$-pole field

Substituting in the general expression for flux given by Eq. (21) and considering the case of small torsional amplitudes, we can write for the flux in a nearly pure $2n$-pole field:

$$\Phi_n(\theta) = \text{Re}\left[K_n e^{in\theta} C(n) e^{-in\alpha_n}\right] + \text{Re}\left[inK_n T(\theta) e^{in\theta} C(n) e^{-in\alpha_n}\right]$$  \hspace{1cm} (40)

The first term on the right hand side is the expected $2n$-pole term in the absence of torsional vibrations. The second term will give rise to other spurious harmonics depending on the magnitude and angular dependence of $T(\theta)$. Terms of second and other higher orders in $T(\theta)$ are neglected in writing Eq. (40). To a good approximation, the amplitude of distortion in the flux seen by the coil is proportional to the amplitude of the torsional vibration as well as the sensitivity, $K_n$, of the coil to the $2n$-pole field. If the magnet has only one dominant harmonic, then the effect of torsional vibrations can be minimized by making the sensitivity of the coil (or an array of coils) zero for that particular harmonic. This is the basis for bucking out the most dominant harmonic term from the pick up signal. It should be noted that if the magnet has large allowed or unallowed multipoles, the effect of torsional vibrations is not completely cancelled by this bucking.

6.2.2 Spurious harmonics due to periodic torsional vibrations

Let us assume that the torsional vibration amplitude, $T(\theta)$, is a periodic function of $\theta$. Such a periodic vibration can be expressed in terms of its Fourier components as

$$T(\theta) = \sum_{p=-\infty}^{\infty} T_p \exp(ip\theta)$$  \hspace{1cm} (41)

Substituting for $T(\theta)$ from Eq. (41) into Eq. (40) and using the identity

$$\text{Re}\left[inT_p K_n e^{-i(p-n)\theta} C(n) e^{-in\alpha_n}\right] = \text{Re}\left[-inT_p^* K_n^* e^{i(p-n)\theta} C(n) e^{in\alpha_n}\right]$$  \hspace{1cm} (42)

the expression for flux in a nearly pure $2n$-pole field can be written as

$$\Phi_n(\theta) = \text{Re}\left[K_n e^{in\theta} C(n) e^{-in\alpha_n}\right] + \text{Re}\left[\sum_{p=-(n-1)}^{\infty} inT_p K_n e^{i(p+n)\theta} C(n) e^{-in\alpha_n}\right]$$

$$+ \text{Re}\left[inT_{-n} K_n C(n) e^{-in\alpha_n}\right] + \text{Re}\left[\sum_{p=n+1}^{\infty} -inT_{-p}^* K_n^* e^{i(p-n)\theta} C(n) e^{in\alpha_n}\right]$$  \hspace{1cm} (43)

The amount of spurious $2m$-pole harmonics in the measured flux is given by:

$$C'(m) e^{-im\alpha_m} \approx \text{Im}\left[\left(\frac{K_n}{K_m}\right)T_{m-n} C(n) e^{-in\alpha_n} - \left(\frac{K_n^*}{K_m^*}\right)T_{-m-n}^* C(n) e^{in\alpha_n}\right]$$  \hspace{1cm} (44)
If $T(\theta)$ has a simple angular dependence of the form $T_0\cos(p\theta)$ or $T_0\sin(p\theta)$ then only $T_p$ and $T_{-p}$ are non-zero. We can see from Eq. (44) that such a motion will produce spurious harmonics corresponding to $2(n+p)$-pole and $2[(n-p)]$-pole field in a $2n$-pole magnet.

7. IMPLEMENTATION OF BUCKING IN PRACTICAL COIL DESIGNS

It was shown in the previous section that existence of transverse and torsional vibrations in the rotational motion of a coil can produce spurious harmonics. To achieve a level of error at $\sim 10^{-5}$ of the fundamental field, the tolerable amplitudes of such vibrations are impractically small. In other words, it is essential, in practice, to buck out at least the most dominant term and the next lower order feed down term in order to obtain accurate harmonics from the observed signal. This is usually achieved by an array of coils. Examples of such coils used in recent years are presented in this section. Since dipoles and quadrupoles are the most important magnets from the point of view of field quality in a high energy accelerator, most coil designs incorporate the ability to buck the dipole and the quadrupole components in a simple way. The same coils could also be used for measuring magnets of other multipolarities by incorporating signal amplifiers/attenuators or by using digital bucking (see Sec. 7.4)

7.1 Coils for HERA dipoles and quadrupoles

The magnets for the HERA accelerator at DESY were measured with radial coils [6]. For measuring dipole magnets, one needs to buck out only the dipole term. This can be achieved with a relatively simple design, as shown in Fig. 8(a), with only two windings. The outer winding (coil A) is the main winding and the central winding is used for bucking out the dipole term. In order to do this with a simple addition of the two signals, the following condition must be satisfied

$$N_A(r_2 - r_1) = N_B(r_3 + r_4)$$

(45)

where $N_A$ and $N_B$ are the number of turns in the two windings and the radii are as shown in Fig. 8(a). The two windings, A and B, had the same number of turns and the same width. In

![Fig. 8 Design of radial coils used for measuring (a) the dipoles and (b) the quadrupole magnets for the HERA collider at DESY (from reference [6])](image)

Page 16
addition to satisfying Eq. (45) for matching the magnitude of the signals, it is also essential that the two windings be coplanar, so that the two signals have exactly the same (or opposite) phase. In practice, there may be an angular misalignment between the two windings due to construction errors. In order to correct for any resulting phase error, a third winding (coil C), orthogonal to the central winding, is also used.

In the case of the quadrupole magnets, one needs to buck out the quadrupole as well as the dipole terms. This requires a slightly more complicated geometry, as shown in Fig. 8(b). The system consists of three main coils, coils A-C, and a fourth coil (D) to compensate for any angular misalignments. Referring to Fig. 8(b), the conditions for bucking the dipole and the quadrupole components can be written as

\[
N_A(r_2 + r_1) - N_B(r_3 + r_4) - N_C(r_5 + r_6) = 0 \tag{46}
\]

\[
N_A(r_2^2 - r_1^2) - N_B(r_4^2 - r_3^2) - N_C(r_6^2 - r_5^2) = 0 \tag{47}
\]

where the bucked signal is given by

\[
V_{\text{bucked}} = V_A - V_B - V_C \tag{48}
\]

### 7.2 Tangential coils for the RHIC magnets

All magnets for the Relativistic Heavy Ion Collider (RHIC), currently nearing completion at the Brookhaven National Laboratory, are measured with a common tangential coil design [8,9] shown in Fig. 9. All windings are placed inside precisely machined grooves on the surface of an insulating cylinder. The system consists of a tangential winding (T1) with 30 turns and an opening angle of 15 degrees. A pair of dipole windings (D1 and D2) and a pair of quadrupole windings (Q1 and Q2) are used for bucking. The bucking windings have

![Fig. 9](image)

A tangential coil design used to measure all the RHIC magnets. The pair of grooves opposite to the T1 winding are not used.
three turns each, and are of the type described in Sec. 4. The angular positions of the various bucking windings are chosen in such a way that a simple addition of all the five signals completely cancels the dipole and the quadrupole terms. While measuring dipole magnets with small harmonic content, it is essential only to buck the dipole term. The use of the quadrupole windings in this case is optional. Although the design of the coil system is common for all magnets, the outer radius of the cylinder on which the windings are placed is chosen to match the maximum available aperture in each magnet type. This provides the maximum accuracy in the measurement of higher harmonics specified at a reference radius which is approximately a fixed fraction (~ 0.625) of the magnet coil inner radius.

7.3 Measuring coils for the Large Hadron Collider (LHC)

At present, work is underway for construction of a Large Hadron Collider (LHC) at CERN. Several measuring coil systems have been designed [7,10] to measure various magnets in this accelerator. The radial coils to be used for measuring the dipole and the quadrupole magnets are shown in Figs. 10(a) and (b). The dipole radial coil, to be used for warm measurements, is a simple design with an external, main winding, and a central bucking winding. Both windings have 400 turns and are of the same width. The second external coil shown in Fig. 10(a) is mainly for mechanical symmetry, but it can also be used as a spare or for diagnostic purposes. The coils are fabricated very precisely and their alignment is very carefully controlled [7]. The phases of the two signals are therefore the same, making it unnecessary to use an orthogonal bucking winding. A simple addition (or subtraction) of the signals from the outer winding and the central winding results in a complete cancellation of the dipole term.

The radial coil system for measuring quadrupoles has five windings and is shown in Fig. 10(b). There is a pair of “outer” coils, a pair of “intermediate radius” coils, and a “central” coil. Only one of the “outer” coils is actually used in the harmonic measurements, while the other outer coil is a spare. The central coil is sensitive to the dipole term and insensitive to the quadrupole term. The two intermediate radius coils have equal and opposite sensitivities for the quadrupole term, but identical sensitivities for the dipole term. The outer coil is twice as sensitive as the intermediate radius coils for the quadrupole term. Thus, a

Fig. 10  Coil designs planned for measuring the magnets for LHC – (a) radial coil for dipoles, (b) radial coil for quadrupoles and (c) tangential coil for dipoles. [Courtesy J. Billan, CERN ]
simple addition of the four signals with appropriate signs results in a cancellation of the dipole and the quadrupole terms. The same system can be used to buck out other harmonics when measuring magnets of a different multipolarity if the signals are added with weight factors other than ±1. This would normally require use of amplifiers with adjustable gains to suit magnets of different multipolarities.

For measurements of LHC dipoles in the superconducting state, it is planned to use a long chain of 13 simultaneously rotating coil assemblies to cover a length of 15 meters at the same time. Simultaneous coverage of the entire magnet is desirable due to time dependent effects in the superconductor. Each of these rotating coils consists of a tangential coil of 36 turns mounted on a ceramic form and a centrally located dipole bucking coil, as shown in Fig. 10(c). The tangential coil has a rather large opening angle of about 30 degrees for practical reasons, resulting in some loss of sensitivity for very high order harmonics. This is not expected to be a problem since this system will be used only for high field measurements, where the signals are sufficiently strong. The second tangential coil on the opposite side is for mechanical symmetry, but it can also be used as a spare or for diagnostic purposes.

7.4 Analog and digital bucking

In the preceding subsections, we have seen several examples of coil designs used for measuring dipoles and quadrupoles. A common feature of these designs is that all of them provide a cancellation of the dipole and the quadrupole terms with a simple addition of the signals. In practice, an exact cancellation of the signals is rarely achieved due to various construction errors. This necessitates either a resistor/amplifier network for precise compensation, or a very careful selection and adjustment of the coils to avoid significant construction errors. Implementation of bucking by actually combining the signals in this way is commonly referred to as analog bucking. Such a scheme is shown in Fig. 11. The most dominant component is obtained from the main winding directly, whereas all the other harmonic information is obtained from an analysis of the bucked signal. The analog bucking scheme has the advantage that it requires fewer data recording channels and a smaller dynamic range in the voltmeters or integrators. On the other hand, it is relatively difficult or cumbersome to change the weight factors of various signals in order to buck out different harmonics in magnets of different multipolarities.
The tangential coils are fabricated at Brookhaven National Laboratory (BNL) by winding the coils directly into grooves machined in the surface of an insulating cylinder. In this technique, it is not possible to produce a large number of coils, and then match those to the bucking coils. In order to accommodate reasonable construction errors and still achieve good bucking, a technique known as digital bucking has been used. In this technique, all the individual coil signals are recorded using digital voltmeters, as shown in Fig. 12. The bucked signal is generated digitally by adding the signals with appropriate weight factors in the analysis software. This technique allows one to use coils with significant construction errors (with proper calibration, of course). Also, the weight factors can be easily changed in the software to measure magnets of different multipolarities. A disadvantage of this method is that more recording channels are needed than in analog bucking, and the voltmeters must have good linearity over a large dynamic range.

The technique of digital bucking has been used successfully for measurement of the magnets for RHIC at BNL. An automated algorithm was developed that senses the multipolarity of the magnet being measured from a FFT of the tangential coil signal. The software then calculates the bucked signal, which is a superposition of signals from all the five windings (see Sec. 7.2 for the RHIC coil design) according to the relation

\[ V_{\text{bucked}} = V_{\text{tangential}} + f_1 V_{D1} + f_2 V_{D2} + f_4 V_{Q1} + f_5 V_{Q2} \]  

(49)

where the real coefficients \( f_1 \) to \( f_5 \) are calculated to buck out the most dominant, and the next lower order harmonic for any magnet. The dipole windings can be used to buck out the dipole, the sextupole, or the decapole term. The quadrupole windings can be used to buck out the quadrupole or the 12-pole terms. The RHIC coil design does not allow bucking of the octupole harmonic, since neither the dipole nor the quadrupole windings have any sensitivity for this term (see Sec. 4.2). If \( n_1 \) is the harmonic to be cancelled with the D1 and D2 windings, then the design values of \( f_1 \) and \( f_2 \) are given by:

\[ f_1 = \left( \frac{N_3}{N_1} \right) \left( \frac{R_3}{R_1} \right)^{n_1} \left( \frac{\sin \{n_1(\delta_3 - \delta_2)\}}{\sin \{n_1(\delta_2 - \delta_1)\}} \right) \sin \left( \frac{n_1 \Delta}{2} \right) \sin \left( \frac{n_1 \pi}{2} \right) \]  

\[ f_2 = \left( \frac{N_3}{N_2} \right) \left( \frac{R_3}{R_2} \right)^{n_1} \left( \frac{\sin \{n_1(\delta_1 - \delta_3)\}}{\sin \{n_1(\delta_2 - \delta_1)\}} \right) \sin \left( \frac{n_1 \Delta}{2} \right) \sin \left( \frac{n_1 \pi}{2} \right) \]  

(50)

Fig. 12  Schematic of digital bucking.
where the \( \delta \)'s are the angular positions of the windings at time \( t = 0 \). Similarly, if \( n_2 \) is the harmonic to be cancelled using the Q1 and Q2 windings:

\[
\begin{align*}
f_4 &= \left( \frac{N_3}{2N_4} \right) \left( \frac{R_3}{R_4} \right)^{n_2} \frac{\sin\{n_2(\delta_3 - \delta_5)\}}{\sin\{n_2(\delta_5 - \delta_4)\}} \sin \left( \frac{n_2 \Delta}{2} \right) \sin \left( \frac{n_2 \pi}{4} \right) \\
f_5 &= \left( \frac{N_3}{2N_5} \right) \left( \frac{R_3}{R_5} \right)^{n_2} \frac{\sin\{n_2(\delta_4 - \delta_3)\}}{\sin\{n_2(\delta_5 - \delta_4)\}} \sin \left( \frac{n_2 \Delta}{2} \right) \sin \left( \frac{n_2 \pi}{4} \right)
\end{align*}
\]

(51)

The design values of these weight factors are given in Table 1 for various harmonics. In practice, these weight factors are determined from the FFT of the actual signals to account for any construction errors. As can be seen from the table, all the weight factors are unity for bucking the dipole and the quadrupole components. Thus, a properly built RHIC coil can also be used with analog bucking for dipoles and quadrupoles.

Table 1  Design values of weight factors to cancel various harmonics

<table>
<thead>
<tr>
<th>( n_1 )</th>
<th>( f_1 )</th>
<th>( f_2 )</th>
<th>( n_2 )</th>
<th>( f_4 )</th>
<th>( f_5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.00</td>
<td>-1.00</td>
<td>2</td>
<td>-1.00</td>
<td>-1.00</td>
</tr>
<tr>
<td>3</td>
<td>-2.26</td>
<td>-2.26</td>
<td>6</td>
<td>-2.06</td>
<td>-2.06</td>
</tr>
<tr>
<td>5</td>
<td>+7.57</td>
<td>+7.57</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

8. **EFFECT OF COIL CONSTRUCTION ERRORS**

So far, we have considered the effect of imperfections in the rotational motion of the coil, and the ways to overcome these effects using bucking. The measuring coil itself was assumed to be built perfectly as per design. In practice, a variety of systematic and random errors may be present due to limitations of the construction method. A systematic error, for example in the radius or the angular position of the coil, may not affect the measurements at all provided the error is accounted for in the analysis by a proper calibration. However, if one is using analog bucking, such systematic errors will result in poor bucking ratios, and are undesirable. In this section, we shall consider the effect of some of the coil construction errors on the measured harmonics.

8.1 **Effect of a finite size of the coil windings**

In practice, the coil windings are not point-like. To accommodate the necessary number of turns (which may be as high as several hundred), the winding must have a finite cross section. This could introduce errors in the measurement of the amplitudes of the harmonics, since different turns of the winding are at different radii.

Although a typical winding cross section is rectangular, it is convenient to approximate it with a sector of an annulus, as shown in Fig. 13(a). The winding is assumed to have an angular width of \( (2\alpha) \) and thickness \( (2\delta) \). The mean position of the winding is denoted by \( z_0 = R.\exp(i\phi) \). The sensitivity factor of the winding to the \( 2n \)-pole term [see Eq. (24)] involves the quantity \( z^n \). The effective value of \( z^n \) averaged over the cross section of the winding can be obtained by integration as follows:
If the winding is assumed to be point like and located at the geometric center, $z_0$, the above formula gives the error in estimating the amplitude of the $2n$-pole term. Expanding in a power series, it can be shown that the leading correction terms are of the second order in $\alpha$ and $(\delta/R)$. In deriving the above formula, it has been assumed that the density of turns varies inversely with the radius. In other words, a layer of the winding at a larger radius has the same number of turns as a layer at a smaller radius. If one assumes a constant density of turns, then the layer at a larger radius will have a slightly higher number of turns. The quantity $(n+1)$ in the last factor on the right hand side of Eq. (52) should be replaced with $(n+2)$ in this case.

The calculated errors in the amplitudes of various harmonics are shown in Fig. 13(b) for windings of 1 mm x 1 mm cross section placed at mean radii of 10 mm and 25 mm. For the case of a 25 mm radius, the error is negligible for all harmonics of interest. The errors are more pronounced for a smaller radius coil. However, even for a 10 mm radius coil, the errors may not be serious since the higher harmonics in most accelerator magnets are very small, at the level of a unit ($10^{-4}$ of the most dominant term) or less and a 1% error in their values is generally negligible. The finite size may, however, have a serious effect while measuring the

![Graph](image-url)

**Fig. 13** (a) A coil winding of a finite size, and (b) errors in the amplitudes of various harmonics with a 1 mm x 1 mm winding for different coil radii. In calculating the errors, the winding is assumed to have a mean position $z_0$, as shown in (a).
transfer function of higher multipole magnets (such as sextupoles, octupoles, etc.) with a small diameter measuring coil. In such instances, it may be desirable to apply a correction to the observed amplitudes.

In calculating the errors shown in Fig. 13(b), it has been assumed that the effective radius of the winding is given by the geometric center, as shown in Fig. 13(a). It should be noted that Eq. (52) gives a correction factor equal to \( \sin \alpha / \alpha \) even for the dipole term. This is because the geometric center is not the center of gravity in the case of a sector of an annulus. If the effective radius is taken to be \( z_0 (\sin \alpha / \alpha) \) instead of \( z_0 \), which is equivalent to determining the radius from a calibration in a reference dipole magnet, then the errors are considerably reduced from those given by Eq. (52). For a winding with the same width as height (\( \alpha = \delta / R \)), it can be shown in this case that the second order correction terms vanish, and the leading correction terms are of the fourth order in \( (\delta / R) \).

In practice, the windings are generally of a rectangular shape, as shown in Fig. 14(a). It can be shown that for a rectangular winding of height \( h \) and width \( w \), the average value of \( z^n \) is given by

\[
(z^n)_{\text{avg.}} = z^n_0 \cdot \left[ \frac{\xi_{1,2}^{n+2} \sin((n+2)\lambda_1) - \xi_{2}^{n+2} \sin((n+2)\lambda_2)}{2(h/2R)(w/2R)(n+1)(n+2)} \right]
\]

(53)

where

\[
\xi_{1,2} = \sqrt{(1 \pm \frac{h}{2R})^2 + \left(\frac{w}{2R}\right)^2} \quad \lambda_{1,2} = \tan^{-1}\left[ \frac{(w/2R)}{1 \pm (h/2R)} \right]
\]

(54)

Fig. 14(b) shows the errors calculated for 1 mm\(^2\) rectangular windings of 10 mm radius and different aspect ratios. It is seen that the errors are significantly smaller for a square shape (\( w/h = 1 \)) due to cancellation of the second order terms. Also, the errors for the 1 mm \( \times \) 1 mm case are much less than those in Fig. 13(a) for a similar case. This is a result of the fact that the geometric center correctly represents the center of gravity of a rectangular winding.

Fig. 14  (a) a winding with a rectangular cross section, and (b) errors in harmonic amplitudes for a 1 mm\(^2\) winding size with different aspect ratios for \( R = 10 \) mm.
8.2 Random variation in the winding radius

For a well constructed coil, it is essential that all the geometric parameters of the coil be maintained uniform along the length of the coil. In practice, there may be random variations in these geometric parameters. We shall study this, and the next few subsections, the effect of such random variations along the length.

Let us first consider the case of a random variation in the coil radius. We consider one segment of the coil loop in either a radial or a tangential coil. The radius is assumed to vary randomly along the length $L$ of the coil, as shown in Fig. 15, with a mean value of $R_c$ and a standard deviation of $\sigma_R$. The effective sensitivity factor of the coil for the $n$-th order harmonic is proportional to the average value of $R^n$ over the coil length. If $\varepsilon(z)$ is the deviation in the radius, $R(z)$, at axial position, $z$, from the mean radius, $R_c$, then we may write:

$$ R(z) = R_c + \varepsilon(z); \quad \frac{1}{L} \int_0^L R(z) \, dz = R_c; \quad \int_0^L \varepsilon(z) \, dz = 0; \quad \sigma_R^2 = \frac{1}{L} \left[ [R(z) - R_c]^2 \right] dz \quad (55) $$

Assuming small errors, and using Eq. (55), we can write the average value of $R^n$ as

$$ \frac{1}{L} \int_0^L [R(z)]^n \, dz = \frac{R_c^n}{L} \left[ 1 + \frac{n}{R_c} \varepsilon(z) + \frac{n(n-1)}{2R_c^2} \varepsilon^2(z) + \cdots \right] dz \approx R_c^n \left[ 1 + \frac{n}{2} \left( \frac{\sigma_R}{R_c} \right)^2 \right] \quad (56) $$

The sensitivity factor for the $n$-th harmonic is given by:

$$ K_n \approx K_n^{\text{ideal}} \left[ 1 + \frac{n(n-1)}{2} \left( \frac{\sigma_R}{R_c} \right)^2 \right] \quad (57) $$

In order to keep the error in the amplitude of the $n = 15$ term less than 1%, we should have $\sigma_R \leq 10^{-2} R_c$. A somewhat tighter tolerance may be required if such a coil is to be used to determine the transfer function of a magnet of higher multipolarity, such as a dodecapole corrector magnet. Practical coils of radius $\sim 20$ mm or more are generally built with $\sigma_R \leq 10^{-3} R_c$, for which the errors calculated from Eq. (57) are negligible.

![Fig. 15](image-url) Random variation in the winding radius along the length of the coil.
8.3 Random variation in the angular position (twist)

Let us consider the case of random variation in the angular position of the coil. The error will be derived for the specific case of a tangential coil in which the radii and opening angle are uniform along the length, but the angular position, $\delta$, is assumed to vary randomly along the length $L$ of the coil, as shown in Fig. 16, with a mean value of $\delta_c$ and a standard deviation of $\sigma_\delta$. The case of a radial coil can be handled in an identical manner, and also yields an identical result. If $\varepsilon(z)$ is the deviation in the angular position at axial position, $z$, we may write

$$\delta(z) = \delta_c + \varepsilon(z); \quad \frac{1}{L} \int_0^L \delta(z) \, dz = \delta_c; \quad \frac{1}{L} \int_0^L \varepsilon(z) \, dz = 0; \quad \sigma_\delta^2 = \frac{1}{L} \left[ \int_0^L [\delta(z) - \delta_c]^2 \, dz \right]$$

The flux seen by the tangential coil for the $2n$-pole field is [see Eq. (9)]:

$$\Phi_n(t) \propto \frac{1}{L} \int_0^L C(n) \sin(n \omega t + n \delta_c + n \varepsilon(z) - n \alpha_n) \, dz$$

Expanding $\sin[n \varepsilon(z)]$ and $\cos[n \varepsilon(z)]$ in power series and retaining only the terms up to the second order, it is easy to show that:

$$\Phi_n(t) \propto C(n) \sin(n \omega t + n \delta_c - n \alpha_n) \left[ 1 - \frac{n^2}{2} \sigma_\delta^2 \right]$$

The sensitivity factor for the $n$-th harmonic is given by:

$$K_n \approx K_n^{\text{ideal}} \left[ 1 - \frac{n^2}{2} \sigma_\delta^2 \right]$$

It should be noted from Eq. (61) that to a good approximation, the correction factor is a real quantity. This implies that the variation in angular position primarily leads to an error in the amplitude, and not the phase of a harmonic term, provided the mean angular position, $\delta_c$, is accurately determined. The same result can be obtained starting from Eq. (5) for a radial coil.

Fig. 16  Random variation in the angular position of a tangential coil.
### 8.4 Random variation in the opening angle of a tangential coil

While a radial coil is characterized by only the radii and the angular position, characterization of a tangential coil involves the opening angle also. For a long coil, this opening angle may vary along the length. Let us consider a tangential coil in which the radii and the angular position are uniform along the length. However, the opening angle, \( \Delta \), is assumed to vary randomly along the length \( L \) of the coil, as shown in Fig. 17, with a mean value of \( \Delta_c \) and a standard deviation of \( \sigma_\Delta \). If \( \varepsilon(z) \) is the deviation in the opening angle at axial position, \( z \), we may write

\[
\Delta(z) = \Delta_c + \varepsilon(z); \quad \frac{1}{L} \int_0^L \Delta(z) \, dz = \Delta_c; \quad \int_0^L \varepsilon(z) \, dz = 0; \quad \sigma^2_\Delta = \frac{1}{L} \int_0^L [\Delta(z) - \Delta_c]^2 \, dz \tag{62}
\]

The flux seen by the coil for the \( 2n \)-pole field is [see Eq. (9)]:

\[
\Phi_n(\theta) \propto \frac{1}{L} \int_0^L \sin \left( \frac{n\Delta(z)}{2} \right) \, dz \tag{63}
\]

Substituting for \( \Delta(z) \) from Eq. (62), expanding \( \sin[n\varepsilon(z)/2] \) and \( \cos[n\varepsilon(z)/2] \) in power series and retaining only up to the second order terms, it is easy to show that:

\[
\Phi_n(\theta) \propto \sin \left( \frac{n\Delta_c}{2} \right) \left[ 1 - \frac{n^2}{8} \sigma^2_\Delta \right] \tag{64}
\]

The sensitivity factor for the \( n \)-th harmonic is given by:

\[
K_n \approx K^\text{ideal}_n \left[ 1 - \frac{n^2}{8} \sigma^2_\Delta \right] \tag{65}
\]

It is seen from Eqs. (61) and (65) that the measured amplitude of a given harmonic is more sensitive to random variations in the angular position than to random variations in the opening angle.

*Fig. 17 Random variation in the opening angle of a tangential coil. The radii and the angular position are assumed to be constant along the length.*
8.5 A tilt in the plane of a tangential coil

Ideally, the plane of a tangential coil should be perpendicular to the radial vector through the center of the coil. In practice, the two grooves used for winding the coil may not be at the same radius (e.g., in the method used at BNL), or the coil may be imperfectly mounted (e.g., in the method used at CERN), resulting in a slight tilt of the plane of the coil, as shown in Fig. 18. Even with a perfectly built coil, such an imperfection will be apparent if the rotation axis does not exactly coincide with the geometric center of the windings. Let us assume that the wires are located at radii of $R_c - \varepsilon$ and $R_c + \varepsilon$, as shown in Fig. 18. The general expression for the sensitivity factor of any rotating coil was derived in Sec. 5, and is given by Eq. (24). In the case of an imperfect tangential coil with unequal radii, we may write,

$$z_{1,0} = (R_c - \varepsilon) \exp(i\Delta / 2); \quad z_{2,0} = (R_c + \varepsilon) \exp(-i\Delta / 2)$$  

$$z_{2,0}^n - z_{1,0}^n = (R_c + \varepsilon)^n \exp(-in\Delta / 2) - (R_c - \varepsilon)^n \exp(in\Delta / 2)$$

$$\approx 2R_c^n \left[ -i\sin\left(\frac{n\Delta}{2}\right) + \left(\frac{n\varepsilon}{R_c}\right) \cos\left(\frac{n\Delta}{2}\right) \right]$$  

The first term on the right hand side in Eq. (67) is related to the sensitivity of the perfect tangential coil [see Eq. (27)]. The second term implies that both amplitude and phase errors are introduced in the sensitivity factor. Also, for coils such as the dipole coil with $\Delta = \pi$, the flux for a perfect coil is zero for even harmonics. This is no longer the case with an imperfect coil. However, the allowed terms for a dipole coil ($\Delta = \pi$) are not affected since $\cos(n\pi/2) = 0$ for odd values of $n$. Assuming that $\sin(n\Delta/2)$ is not zero, as is the case for the harmonics of interest in a practical tangential coil ($\Delta \approx 15^\circ$), we may write:

$$K_n = K_n^{ideal} A_n \exp(in\lambda_n)$$  

Fig. 18 A tilt in the plane of a tangential coil, resulting in (or resulting from) unequal radii for the two grooves of the winding.
where \( A_n \) is an amplitude correction factor and \( \lambda_n \) is a phase error given by:

\[
A_n = \sqrt{1 + \left(\frac{n \varepsilon}{R_c}\right)^2 \cot^2 \left(\frac{n \Delta}{2}\right)} \approx 1 + \frac{n^2}{2} \left(\frac{\varepsilon}{R_c}\right)^2 \cot^2 \left(\frac{n \Delta}{2}\right)
\]  
(69)

\[
\lambda_n = \left(\frac{1}{n}\right) \tan^{-1} \left[ \left(\frac{n \varepsilon}{R_c}\right) \cot \left(\frac{n \Delta}{2}\right) \right] \approx \left(\frac{\varepsilon}{R_c}\right) \cot \left(\frac{n \Delta}{2}\right)
\]  
(70)

The amplitude error is of the second order in \((\varepsilon/R_c)\) and can generally be neglected. For typical values of \((\varepsilon/R_c) \sim 10^{-3}\), the phase error could be several milli-radians for the lowest order harmonics. The phase error reduces with the order of the harmonic as roughly \((1/n)\). This harmonic dependent phase error makes a calibration of the angular position difficult, because the apparent initial angular position of the coil measured in a dipole field will be different from that measured in a quadrupole field. In fact, such a discrepancy in the two calibrations can be used to estimate the tilt in the plane of the coil (see Sec. 10.1.2) as part of the coil calibration procedure. One can then use the sensitivity factor given by Eq. (68) to account for this tilt. Such a calibration and analysis procedure allows one to tolerate much larger construction errors than would be acceptable otherwise. An interesting observation from Eqs. (69) and (70) is that allowed harmonics in multipole coils (of the type discussed in Sec. 4) are insensitive to this type of error because \(\cot(n\Delta/2)\) is zero.

8.6 A tilt in the plane of a radial coil

Similar to the case of a tangential coil discussed in the previous subsection, there can be a tilt in the plane of a radial coil from the ideal position. Such a tilt for a radial coil is shown in Fig. 19. The two ends of the coil are located at radii \(R_1\) and \(R_2\), and at two different angular positions, \(\theta - \varepsilon\) and \(\theta + \varepsilon\), instead of at a fixed angle \(\theta\). In this case, we may write,

\[
z_{1,0} = R_1 \exp(-i\varepsilon); \quad z_{2,0} = R_2 \exp(i\varepsilon)
\]  
(71)

The sensitivity factor for such a coil is given by

\[
K_n = \frac{NLR_{ref}}{n} \left[ \left( \frac{R_2}{R_{ref}} \right)^n - \left( \frac{R_1}{R_{ref}} \right)^n \right] \cos(n\varepsilon) + i \sin(n\varepsilon) \left[ \left( \frac{R_2}{R_{ref}} \right)^n + \left( \frac{R_1}{R_{ref}} \right)^n \right]
\]  
(72)

The sensitivity factor is no longer a purely real quantity. Assuming a non-zero sensitivity factor for the ideal coil (i.e. \(R_2^n - R_1^n \neq 0\)) and small values of \(\varepsilon\), we can write

\[
K_n \approx K_{n,ideal}^\text{ideal} A_n \exp(in\lambda_n)
\]  
(73)

where \(A_n\) is an amplitude correction factor given by:

\[
A_n = \sqrt{\cos^2(n\varepsilon) + \sin^2(n\varepsilon) \left( \frac{R_2^n + R_1^n}{R_2^n - R_1^n} \right)^2} \approx 1 + \frac{(n\varepsilon)^2}{2} \left[ \frac{R_2^n + R_1^n}{R_2^n - R_1^n} \right]^2 - 1
\]  
(74)
and $\lambda_n$ is the phase error for the $2n$-pole harmonic given by

$$\lambda_n = \left(\frac{1}{n}\right)\tan^{-1}\left[\tan(n\theta) \frac{R_2^n + R_1^n}{R_2^n - R_1^n}\right] = \varepsilon \left(\frac{R_2^n + R_1^n}{R_2^n - R_1^n}\right)$$

The amplitude error is of the second order in $\varepsilon$, and can be neglected in practice. For $|R_1| \sim |R_2|$, the phase error may be several times $\varepsilon$ for the lower order terms and has a limiting value of $\varepsilon$ for sufficiently higher harmonics. As was mentioned in the previous subsection, a tilt in the plane of a radial coil will make the calibration of the initial angular position dependent on the multipolarity of the field used. Once again, this fact can be used to calibrate such a tilt and then correct the analysis for this effect. For the special case of a dipole coil with $R_1 = -R_2$, it can be seen that the errors are zero for all the allowed harmonics. This result is similar to that for a tilt in the plane of tangential coils.

### 8.7 An offset in the rotation axis

A coil assembly may be built perfectly, but the drive mechanism that rotates the coil may be misaligned with the geometric center of the coil. As was mentioned earlier, the effect of such an offset in the rotation axis is equivalent to a tilt in the plane of the coil, which has been discussed in the previous two subsections. If the radii and angular positions are calibrated with the rotating system already in place, then one need not consider the offset of the rotation axis as a separate imperfection. However, if one is using the design values of the coil parameters, then the effect of the offset in the rotation axis should be estimated.

Fig. 20(a) shows a general type of rotating coil with the initial positions of the wires given by $z_{1,0}$ and $z_{2,0}$ with respect to the geometric center as the origin. The rotation axis is assumed to be offset from the geometric center by a vector $\Delta z_0$ in the complex plane. We have

$$K_n \propto \left[ (z_{2,0} + \Delta z_0)^n - (z_{1,0} + \Delta z_0)^n \right], \quad K_{n}^{ideal} \propto \left[ z_{2,0}^n - z_{1,0}^n \right]$$
If $\Delta K_n = (K_n - K_n^{ideal})$ is the error in the sensitivity factor for the $2n$-pole term, then it can be shown that

$$\frac{\Delta K_n}{K_n} \text{tangential} = \sum_{k=1}^{n-1} \left[ \frac{n!}{k!(n-k)!} \right] (\Delta z_0) R_c \frac{\Delta z_0}{R_c} \frac{n}{2} \sin \left( \frac{(n-k)\Delta}{2} \right) \sin \left( \frac{n\Delta}{2} \right)$$

(77)

$$\frac{\Delta K_n}{K_n} \text{radial} = \sum_{k=1}^{n-1} \left[ \frac{n!}{k!(n-k)!} \right] (\Delta z_0) R_c \frac{R_2^{n-k} - R_1^{n-k}}{R_2^n - R_1^n}$$

(78)

Usually, the offset is expected to be only a very small fraction of the coil radii. Consequently, a first order approximation is adequate:

$$\frac{\Delta K_n}{K_n} \text{tangential} \approx n \left( \frac{\Delta z_0}{R_c} \right) \sin \left( \frac{(n-1)\Delta}{2} \right) ; \quad \frac{\Delta K_n}{K_n} \text{radial} \approx n (\Delta z_0) \left[ \frac{R_2^{n-1} - R_1^{n-1}}{R_2^n - R_1^n} \right]$$

(79)

Eq. (79) may be used to estimate the tolerance on the alignment of the rotation axis with the geometric axis for a desired degree of accuracy in the measurement of harmonics. These results may also be used to estimate the effect of a “bow” or a bend in the measuring coil. Different sections of such a coil will rotate about an axis which is offset from the geometric
center by different amounts. An approximate upper bound on the resulting effect can be obtained by equating $\Delta z_0$ to the total bend in the measuring coil, while an integration of Eq. (79) over the actual profile of the measuring coil should be used for a more accurate estimate. The estimated errors due to a 0.1 mm offset in the rotation axis are shown in Fig. 20(b) for a tangential and a radial coil with typical dimensions.

8.8 Systematic errors in the estimation of coil parameters

The coil parameters of primary interest are the radius ($R$), the angular position at the start of the data acquisition ($\delta$), and in the case of a tangential coil, the opening angle ($\Delta$). A systematic error in the knowledge of these parameters will result in systematic errors in the calculation of the field parameters, namely the amplitudes $C(n)$ and the phase angles $\alpha_n$. Such errors would arise, for example, when the as-built coil has dimensions different from the design values, but the design values are used in the analysis. With a good calibration procedure, the effect of systematic errors can be practicallly eliminated. The only exception is in the case of analog bucking (see Sec. 7.4), where systematic errors can result in very poor bucking ratios unless amplifiers/resistors are used.

8.8.1 Systematic error in the radius

The sensitivity factor for the $2n$-pole field is proportional to $R^n$. Therefore, the error in the sensitivity factor due to a systematic error, $\Delta R$, in the radius is given by

$$\frac{\Delta K_n}{K_n} = n\left(\frac{\Delta R}{R}\right)$$

(80)

For a $(\Delta R/R) \sim 10^{-3}$, the systematic error in the amplitude of the 20-pole term will be $\sim 1\%$. Such an error is quite acceptable for the higher harmonics. The tolerance on the error in the knowledge of the radius arises primarily from considerations of the accuracy required in the measurement of the transfer function in the main dipole and quadrupole magnets in an accelerator and the desirability of consistency between different measuring systems.

8.8.2 Systematic error in the angular position

A systematic error, $\epsilon_\delta$, in the initial angular position, $\delta$, leads to the same error in the determination of all the phase angles. This would give rise to skew terms in a purely normal magnet, and vice versa.

$$\alpha'_n = \alpha_n - \epsilon_\delta; \quad C'(n) \exp(-in\alpha'_n) = C(n) \exp(-in\alpha_n)[\cos(n\epsilon_\delta) + i\sin(n\epsilon_\delta)]$$

(81)

where the primed quantities refer to the measured values and the unprimed quantities refer to the true values. The normal and skew multipoles in a magnet are generally expressed in a reference frame where the main field component has a zero phase angle. In this case, there will be no systematic error in the multipoles, since the phase angles relative to the main field still remain the same. However, when accurate determination of the field direction of the main component is required, such a systematic error is unacceptable. Efforts must be made to periodically check the calibration, and/or correct for the errors by making measurements from the lead and the non-lead ends of the magnet. A $2m$-pole magnet with a true phase angle of $\alpha_m$ when viewed from the lead end, has a phase angle of $[1+(-1)^m](\pi/2m) - \alpha_m$ when viewed from
the non-lead end. The measured phase angles from the lead and the non-lead ends in the presence of a systematic error in the angular position are:

\[ \alpha_{\text{lead}} = \alpha_m - \varepsilon_\delta; \quad \alpha_{\text{non-lead}} = \left[1 + (-1)^m\right]\left(\frac{\pi}{2m}\right) - \alpha_m - \varepsilon_\delta \tag{82} \]

From these two measured values, it is easy to see that the true phase angle, \( \alpha_m \), is given by

\[ \alpha_m = \left(\frac{1}{2}\right)\left[\alpha_{\text{lead}} - \alpha_{\text{non-lead}} + \left(1 + (-1)^m\right)\left(\frac{\pi}{2m}\right)\right] \tag{83} \]

and the systematic error in the angle is given by

\[ \varepsilon_\delta = \left(\frac{1}{2}\right)\left(1 + (-1)^m\right)\left(\frac{\pi}{2m}\right) - \alpha_{\text{lead}} - \alpha_{\text{non-lead}} \tag{84} \]

Having determined the systematic error in the angle from Eq. (84), the correction can be incorporated by adjusting the parameter \( \delta \) accordingly in all the future analysis. It is generally a good idea to carry out periodic checks of this calibration whenever a reliable measurement of the field angle is required.

8.8.3 Systematic error in the opening angle of a tangential coil

The sensitivity factor, \( K_n \), of a tangential coil [see Eq. (27)] depends on the opening angle, \( \Delta \), through the \( \sin(n\Delta/2) \) factor. For a systematic error \( \varepsilon_\Delta \) in the opening angle,

\[ \frac{\Delta K_n}{K_n} = \left(\frac{n}{2}\right)\cot\left(\frac{n\Delta}{2}\right)\varepsilon_\Delta \tag{85} \]

This error is generally insignificant for higher order harmonics. In a typical tangential coil system, the dipole and the quadrupole terms are obtained from the dipole (\( \Delta = 180 \) degrees) and the quadrupole (\( \Delta = 90 \) degrees) windings respectively. It is seen from Eq. (85) that such multipole coils (\( \Delta = \pi/n \)) are insensitive to small errors in the opening angle.

8.9 Effect of a finite signal averaging time

In the acquisition of voltage data from the RHIC tangential coils, the signals are averaged over one power line cycle to get rid of any AC noise on the signals. At a typical angular speed of one revolution every 3.5 seconds and a power line frequency of 60 Hz, the coil rotates about 1.7 degrees during this time. This motion during data integration can cause errors. Even though this does not strictly fall under construction errors, a finite integration time manifests itself in some sense as a calibration error, as we shall see. Accordingly, this topic is discussed in this section as part of the general discussion on coil errors.

We shall consider the case of a tangential coil here, although a similar result can be derived for radial coils also. If \( \Delta t \) is the averaging time, the \( n \)-th harmonic component in the measured voltage signal is [see Eq. (10)]:

\[ V_n(t) \propto \frac{1}{\Delta t} \int_{t}^{t+\Delta t} \cos(n\omega t + n\delta - n\alpha_n)dt = \frac{\sin(n\omega \Delta t / 2)}{n\omega \Delta t / 2} \cos\left[n\omega t + n\left(\delta + \frac{\omega \Delta t}{2}\right) - n\alpha_n\right] \tag{86} \]
It is seen that both the amplitude and the phase of the $2n$-pole term are affected. While the amplitude correction is harmonic dependent, the phase correction is independent of the harmonic. The constant phase error can be absorbed into the calibration of the initial angular position, $\delta$. From Eq. (86), we can write,

$$\text{Amplitude Correction Factor} = \sin\left(\frac{n\omega \Delta t}{2}\right) / n = 1 - \frac{1}{2}n^2 \left(\frac{\Delta t}{T}\right)^2$$  \hspace{2cm} (87)

$$\text{Effective Angle Calibration} = \delta' = \delta + \omega \Delta t / 2 = \delta + \pi \left(\frac{\Delta t}{T}\right)$$  \hspace{2cm} (88)

where $T$ is the period of rotation. For $\Delta t = 1/60$ sec. and $T = 3.5$ sec, ($\Delta t/T \sim 4.8 \times 10^{-3}$, the amplitude error is 0.004% for the dipole term and is 0.84% for the 30-pole term. This effect is negligible. However, the angle calibration is affected by roughly 0.86 degrees (15 mrad). Fortunately, the error is harmonic independent, and can be absorbed in the calibration of the coil, as long as the angular velocity is kept the same. Considerable error will result, for example, if the coil were to rotate in the opposite direction ($\omega \rightarrow -\omega$).

In practice, the coil rotation period may not be exactly the same during calibration and measurements. It is necessary, therefore, to specify the rotation speed of the measuring coil along with the angular parameters. Corrections must be applied to the calibration values based on the actual rotation speed during the measurements according to the above equations if accuracy in the range of 0.1 mrad is desired in the field angle.

9. DEVIATION OF THE ROTATION AXIS FROM THE MAGNETIC AXIS

In the previous section, we considered the effects of various random and systematic errors in the construction of a rotating coil. The harmonic expansion parameters in Eq. (1)-(3) are dependent on the choice of the coordinate axis [1]. In the analysis of harmonic coil data, it is assumed that the origin is at the rotation axis. In the measurement of an accelerator magnet, one is generally interested in the harmonics at the magnetic center. Unless special effort is made, the rotation axis of the coil may not coincide with the magnetic axis. In this section, we shall study the effect of a misalignment of the rotation axis with the magnetic axis. Three different forms of misalignment will be considered – a simple offset, a sag of the measuring coil due to its own weight, and a tilt of the rotation axis with respect to the magnet axis.

9.1 Rotation axis offset from the magnetic axis

Let us first assume that the rotation axis is parallel to the magnetic axis, but is not coincident with it. This is the most common form of misalignment in practice. This is also the most important form of misalignment because the measured harmonic coefficients are affected by feed down. Figure 21 shows the measuring coil reference frame, $X$-$Y$, with the origin at the rotation axis passing through $O$, and the desired reference frame, $X' - Y'$, with the origin at the magnetic axis passing through $O'$, located at $(x_0, y_0)$ in the $X$-$Y$ frame.

If $C(n)$ and $\alpha_n$ are the measured parameters in the measuring coil frame, then the parameters $C'(n)$ and $\alpha'_n$ in the magnet’s frame are given by [1]:

Page 33
If the offset \( z_0 = (x_0 + i y_0) \) is somehow known, one can account for the misalignment by calculating the field parameters in the magnet’s frame using Eq. (89).

9.1.1 Magnets other than dipole magnets

For a \( 2m \)-pole magnet other than a dipole, it is quite natural to define the magnetic center as the location where the \( 2(m-1) \)-pole feed down term is zero. Thus, the offset itself can be calculated from the measured harmonics using the following equation derived from Eq. (89):

\[
C'(m-1)e^{-i(m-1)\alpha_{m-1}} = C(m-1)e^{-i(m-1)\alpha_{m-1}} + (m-1)C(m)e^{i m \alpha_m} \left( \frac{z_0}{R_{\text{ref}}} \right) + \frac{m(m-1)}{2} C(m+1)e^{-i(m+1)\alpha_{m+1}} \left( \frac{z_0}{R_{\text{ref}}} \right)^2 + \frac{m(m-1)(m+1)}{6} C(m+2)e^{-i(m+2)\alpha_{m+2}} \left( \frac{z_0}{R_{\text{ref}}} \right)^3 + \cdots = 0 \tag{90}
\]

For most \( 2m \)-pole magnets, amplitudes other than \( C(m) \) are small. For small offsets, we may neglect higher order terms in \( (z_0/R_{\text{ref}}) \) and use a simplified equation:

\[
\left( \frac{z_0}{R_{\text{ref}}} \right) = - \frac{1}{(m-1)} \frac{C(m-1)\exp\{-i(m-1)\alpha_{m-1}\}}{C(m)\exp(-i m \alpha_m)} ; \quad m \neq 1 ; |z_0| \ll R_{\text{ref}} \tag{91}
\]

Once the offset is calculated from Eq. (90) or (91), the measured data can be transformed using Eq. (89) to obtain the field parameters in the desired reference frame. This procedure of transforming the measured data to the magnet’s reference frame is commonly referred to as the centering of data.
9.1.2 Dipole magnets

For dipole magnets, the field is practically uniform in the entire magnet aperture and no "natural" definition of a magnetic center can be used. Various strategies are used to define the center of a dipole magnet. For example, one could argue that the very high order unallowed terms are not sensitive to small construction errors, and hence must be zero. In practice, this certainly appears to be the case for any carefully built accelerator dipole magnet. If so, one could pick \( m \) in Eq. (90) to be a sufficiently high order allowed term and calculate the center by requiring the amplitude of the next lower unallowed term, \( C' (m-1) \), to be zero. Of course, this works well only if \( C(m) \) itself has sufficient strength. Because of the large values of \( m \), the measured coefficients are of comparable strengths for both the allowed and the unallowed harmonics, even with relatively small offsets. As a result, it is often necessary to include terms of higher orders in \( (z_0/R_{ref}) \) to calculate the offset. In many cases, ambiguous, yet physically meaningful, results may be obtained because of the non-linear nature of the equations. To resolve this, it is best to find an offset that will simultaneously minimize several unallowed harmonics, rather than just one.

Other strategies for dipoles have exploited a current dependence of the sextupole term. Such a current dependence may arise from non-uniform saturation of the iron yoke at high fields, or from the magnetization of the superconductor at relatively low fields in a superconducting magnet. The quadrupole term at the magnetic center is not expected to show any current dependence. However, if the measuring coil axis does not coincide with the magnetic axis, a current dependence will be seen due to feed down from the sextupole term. Thus, the offset of the measuring coil can be obtained by requiring that the current dependence of the quadrupole multipole be minimized. In practice, one may find that the calculated center depends on the range of the excitation curve used in the minimization. In some cases, the current dependence in the quadrupole term may not be entirely due to feed down. For example, the critical currents of the superconducting cables used for the upper and the lower coils may be slightly different, leading to a current dependence in the quadrupole multipole at low fields. Such effects may limit the accuracy achievable with this technique. Furthermore, this technique can not be used for warm measurements at low fields.

9.1.3 The quadrupole configured dipole ("ugly quad") method for dipole magnets

The methods discussed so far for dipole magnets have inherent limitations and are not universally applicable. A different approach, known as the quadrupole configured dipole (QCD) or an “ugly quad” method, was proposed for centering in the SSC dipoles [11]. Subsequently, the method was adopted for all the RHIC dipoles and has been used with much success for measurements of magnets in both warm and superconducting state.

The quadrupole configured dipole method relies on powering the two coil halves of a dipole magnet with equal and opposite currents, as shown in Fig. 22, to produce a strong skew quadrupole field, instead of a dipole field. This requires a center tap connection on the magnet. The allowed harmonics are now the skew quadrupole, skew octupole, skew dodecapole, and so on. Several of these allowed harmonics are quite strong, and feed down from any one of them could be used to calculate the center. Since the quadrupole field produced in this way has large octupole and higher harmonics, the method is sometimes referred to as the “ugly quad” method.
Since two separate power supplies are required in this mode, it is important to balance the current in the two halves with great accuracy, otherwise a spurious dipole field will also be generated that would affect centering calculations. The sensitivity to such a current mismatch is reduced by several orders of magnitude if one uses feed down from the skew octupole term, or any other higher order term, rather than the dominant skew quadrupole term. Matching the currents in the two halves within 0.1% will give a centering error of less than a few micrometers in this case. The offsets calculated from the QCD method have very little noise (typically only a few micro-meters) and are in very good agreement with the centers calculated by making high order unallowed terms zero.

9.2 Sag of the measuring coil due to its own weight

In order to measure precisely the integral field in long magnets, one has to build long rotating coils. For such long measuring coils, the weight of the cylindrical tube containing the coil may be enough to cause a sagitta in the coil. In this case, each subsection of the coil rotates about its local geometric center. However, the location of this center varies along the length of the magnet, as shown by the dashed line in Fig. 23. This is different from a “bow” or bend in the coil, in which all sections of the coil rotate about the same straight line axis, but the axis is offset from the geometric center by different amounts along the length of the coil. The case of a bend in the coil was discussed briefly in Sec. 8.7. In the case of a sag of the coil, various subsections of the coil see harmonics that are in a frame which is slightly displaced from the adjacent subsections. If \( r_0(Z) \) is the vertically downward offset at axial position \( Z \), the measured coefficients are given by:

\[
C'(n) \exp(-in\alpha'_n) = \sum_{k=n}^{\infty} C(k) \exp(-ik\alpha_k) \frac{(k-1)!\exp\left\{-i(k-n)\frac{\pi}{2}\right\}}{(n-1)!(k-n)!} \int_{-L/2}^{L/2} \left(\frac{r_0(Z)}{R_{\text{ref}}}\right)^{k-n} dZ \quad (92)
\]
where $C(k)$ and $\alpha_k$ are the true field parameters. Equation (90) has been used in writing Eq. (92) where the offset $z_0$ is vertically downward, and is a function of the axial position $Z$:

$$z_0(Z) = r_0(Z) \exp(-i\pi / 2)$$  \hspace{1cm} (93)

If the profile of the coil is known, the integral in Eq. (92) can be evaluated to estimate the effect of the sag on various harmonics. Assuming a parabolic approximation to the profile given by

$$r_0(Z) = h \left[ 1 - \frac{Z^2}{(L/2)^2} \right]; \quad -\frac{L}{2} \leq Z \leq \frac{L}{2}$$  \hspace{1cm} (94)

where $h$ is the amount of sag at the center of the coil ($Z = 0$), it can be shown [12] that:

$$C'(n) \exp(-i\alpha'_n) = \sum_{k=n}^{\infty} C(k) \exp(-ik\alpha_k) \frac{(k-1)! \exp\left\{-i(k-n)\frac{\pi}{2}\right\}}{(n-1)!(k-n)!} \frac{[2(k-n)]!!}{[2(k-n)+1]!!} \left( \frac{h}{R_{ref}} \right)^{k-n}$$  \hspace{1cm} (95)

where $(2k)!! = 2.4.6.8. \ldots 2k$, $(2k+1)!! = 1.3.5. \ldots (2k+1)$ and $0!! = 1$.

The error in certain harmonics may be significant, even with small amounts of sag. However, for small values of $h$ ($<<R_{ref}$), the effect of sag from Eq. (95) is nearly the same as a uniform displacement of the coil by an amount $(2h/3)$, which is the average displacement of the sagging coil. If the measured data are corrected for any offset of the measuring coil axis, most of the errors due to sag are also subtracted out [3], except for terms of second and higher orders in $(h/R_{ref})$ which can be neglected for practical values of $h$. Thus, within reasonable limits, the sag of a measuring coil is not a problem for the measurement of harmonics, provided the data are centered using a reliable centering technique based on the measured harmonics, as described in Sec. 9.1. However, if the harmonic coil is being used specifically to measure the magnetic axis, for example in a quadrupole, even a relatively small sag in the coil may be unacceptable, because one can not distinguish between a true offset of the magnetic axis and an apparent offset due to the sag.
9.3 Tilt of the measuring coil axis relative to the magnet axis

If sufficient care is not taken in aligning the coil, the measuring coil axis can be tilted with respect to the magnetic axis. Generally, the clearance between the inner diameter of the magnet bore and the outer diameter of the coil assembly is kept sufficiently small, so that such a tilt can not be very large. In most cases, this tilt is not of much concern. However, in some situations, even a small tilt can have a significant impact on the measurement of certain harmonics. As a result, it is important to study this type of misalignment of the coil axis.

We shall consider a simple case in which the magnet has uniform field quality along the length, \( L \), covered by the measuring coil. This would apply, for example, to measurements with a relatively short coil in the straight section of a magnet. We also assume a pure tilt, with the average offset of the measuring coil to be zero, as shown in Fig. 24. The displacements of the coil-axis at the two ends are assumed to be given by \((r_0, \xi)\) and \((r_0, \xi+\pi)\) in polar coordinates. The measured harmonics can be calculated by integrating along the length using a procedure similar to that used for a sag of the coil in Sec. 9.2. If the field quality is uniform along the length, the odd orders of feed down from one half of the magnet will be cancelled by the corresponding feed down from the other half of the magnet. The even orders of feed down from the two halves will add to each other. The measured coefficients in this case are given by [3]:

\[
C'(n) \exp(-in\alpha'_{n'}) = \sum_{k=0}^{\infty} \frac{C(k) \exp(-ik\alpha_k)}{(k-n+1)(k-1)!} \frac{(k-1)!}{(n-1)!(k-n)!} \left( \frac{r_0 \exp(i\xi)}{R_{\text{ref}}} \right)^{k-n} \tag{96}
\]

It should be noted that the summation in Eq. (96) includes only those values of \( k \) for which \((k-n)\) is even. The lowest order correction term is of the second order in \((r_0/R_{\text{ref}})\), and can generally be neglected for dipole and quadrupole magnets, since there can not be any second order feed down from the main field component. The only errors will be due to second and higher order feed down from the higher harmonics, which are expected to be small themselves. However, for sextupoles and magnets of higher multipolarity, there can be a second order feed down from the main harmonic, leading to large errors even with a relatively small tilt. For example, the dipole field component will be incorrectly measured in a sextupole magnet, the quadrupole component in an octupole magnet, and so on.

![Fig. 24 Tilt of the measuring coil axis with respect to the magnet axis. The cylinder represents the section of the magnet covered by the length of the measuring coil](image-url)
As a demonstration of the effect of tilt, Fig. 25 shows the correlation between harmonics measured in RHIC sextupoles with a well aligned coil on a granite table and with a possibly tilted coil in a vertical dewar. As explained above, the dipole term is expected to be affected by a tilt in the measuring coil axis. The correlation is indeed poor for the dipole term, as can be seen from Fig. 25(a). The difference between the two measurements is not a systematic one resulting from calibration errors, but is random depending upon the amount of tilt in the coil, and could be up to ±25 units. On the other hand, for the octupole term, which is not expected to be affected by such a tilt, the correlation between the two measurements is quite good, as shown in Fig. 25(b).

The above analysis assumed that the field quality of the magnet is uniform over the length of the measuring coil. Often, magnets may have rather large harmonics in the lead end region which are absent in the non-lead end region. In this case, if one is making an integral measurement with a long coil, even the first order terms from the two halves will not cancel each other, causing large errors. Examples of such errors are in the measurement of integral decapole terms in a quadrupole magnet having large dodecapole terms in the lead end region.

10. CALIBRATION OF HARMONIC COILS

In the previous two sections, we studied the effects of various imperfections in the coil construction and placement. Analysis of such errors can be helpful in determining various tolerances. On the other hand, once a coil is constructed, it is important to characterize the geometric parameters of the various windings as precisely as possible. Most coil systems in use have special windings sensitive to the dipole and the quadrupole terms. The radii and angular positions of such windings can be determined if reference dipole and quadrupole magnets with known field strengths and directions are available. Having such magnets with apertures and lengths suitable for the coil being calibrated greatly simplifies the process of
calibration. In an accelerator, magnets of various lengths and apertures are used. Correspondingly, one needs to use a variety of measuring coils to measure these magnets. It may not always be practical to provide reference dipole and quadrupole magnets for all coil types. In such a situation, one has to obtain as much information about the coil geometry as possible, without requiring an explicit knowledge of the field strength or direction. The number of calibration parameters can be greatly reduced simply by comparing the outputs of various windings in a given field type. One must devise a scheme appropriate to the specific coil design at hand. The basic concept will be illustrated here with the example of the five-winding tangential coil design used at RHIC.

10.1 Calibration of a five-winding tangential coil

The cross-section of a five winding tangential coil design used at RHIC is shown in Fig. 9. There are two dipole buck windings (D1 and D2), two quadrupole buck windings (Q1 and Q2) and a tangential winding (T1) with an opening angle of 15 degrees. The radii of these windings are assumed to be given by $R_1$ through $R_3$. Similarly, the angular positions of the windings are assumed to be given by $\delta_1$ through $\delta_5$. In addition, one needs to know the exact opening angle, $\Delta$, of the tangential winding. The nominal values of these eleven parameters are known from the design. However, the precise values of these parameters in the as-built coil may differ slightly from the design and must be obtained by a calibration procedure. In addition to these eleven parameters, it has been found necessary to account for a small tilt in the plane of the tangential coil, characterized by a parameter $\varepsilon$, which can give rise to a harmonic dependent correction in the angular position (see Sec. 8.5). Thus, there are a total of 12 parameters to be determined in this case. It will be shown here that one can determine all the angles relative to each other, all the radii relative to each other, the absolute value of the opening angle, as well as the tilt in the plane of the tangential coil — all by using dipole, quadrupole and sextupole fields whose strengths and field directions are not necessarily known. The absolute values of all the parameters can be obtained if just one of the field strengths and one of the field directions are known.

10.1.1 Calibration of the radii

Let the $n$-th harmonic amplitude in the voltage signal from the $j$-th winding be denoted by $V_{j,n}$. In a dipole field, assuming a measuring coil longer than the magnet, the dipole amplitudes in the signals from D1, D2 and T1 windings are given by [see Eq. (10)]:

$$V_1(1) \propto N_1 R_1; \quad V_2(1) \propto N_2 R_2; \quad V_3(1) \propto N_3 R_3 \sin(\Delta/2)$$

(97)

where $N_1$, $N_2$ and $N_3$ are the number of turns in the three windings, and are assumed to be known. For measuring coils shorter than the magnet, one has to include the lengths of the windings also in the above equations. In general, the windings have lengths much larger than their radii. Thus, small errors in the lengths are not as critical as similar errors in the radii. Except for ultra short coils, one can assume the lengths to be given by the design value. Similarly, the opening angles of the two dipole buck windings can be safely assumed to be 180 degrees, since the quantity $\sin(n\Delta/2)$ is not very sensitive to small errors in $\Delta$ in this case. The quadrupole windings are not sensitive to the dipole field, and are expected to give practically zero signal in a dipole field. From Eq. (97), we have,
\[
\frac{R_2}{R_1} = \left( \frac{V_2(1)}{V_1(1)} \right) \left( \frac{N_1}{N_2} \right); \quad \frac{R_3}{R_1} \sin \left( \frac{\Delta}{2} \right) = \left( \frac{V_3(1)}{V_1(1)} \right) \left( \frac{N_1}{N_3} \right) \tag{98}
\]

Similarly, in a quadrupole field, the amplitudes of the quadrupole terms in the voltage signals from the Q1, Q2 and T1 windings are given by [see Eqs. (18) and (10)]:

\[
V_4(2) \propto 2N_4R_4^2; \quad V_5(2) \propto 2N_5R_5^2; \quad V_3(2) \propto N_3R_3^2 \sin(\Delta) \tag{99}
\]

which leads to the ratios of radii,

\[
\left( \frac{R_5}{R_4} \right) = \left[ \left( \frac{V_5(2)}{V_4(2)} \right) \left( \frac{N_4}{N_5} \right) \right]^{1/2}; \quad \left( \frac{R_3}{R_4} \right)^2 \sin(\Delta) = \left( \frac{V_3(2)}{V_4(2)} \right) \frac{2N_4}{N_3} \tag{100}
\]

The constants of proportionality in Eqs. (97) and (99) involve the field strengths and the angular velocity. The dependence on the angular velocity can be removed, in principle, by looking at the integral of the voltage signal. If the field strengths in the dipole and the quadrupole magnets are also known, one can obtain the absolute values of $R_1$, $R_2$, and $R_4$, $R_5$. From these values, the absolute values of $R_3$ and $\Delta$ can also be determined using the above equations. If the absolute strengths are not known, one can only calculate the ratios of radii. Even then, we have only four equations in five unknowns (four ratios of the radii and the opening angle) and need more information. If we use a sextupole field, then the D1, D2 and T1 windings are sensitive to this field. The amplitudes of the sextupole term in the signals are given by [see Eqs. (18) and (10)]:

\[
V_1(3) \propto N_1R_1^3; \quad V_2(3) \propto N_2R_2^3; \quad V_3(3) \propto N_3R_3^3 \sin(3\Delta / 2) \tag{101}
\]

which lead to the ratios

\[
\left( \frac{R_2}{R_1} \right) = \left[ \left( \frac{V_2(3)}{V_1(3)} \right) \left( \frac{N_1}{N_2} \right) \right]^{1/3}; \quad \left( \frac{R_3}{R_1} \right)^3 \sin(3\Delta / 2) = \left( \frac{V_3(3)}{V_1(3)} \right) \left( \frac{N_1}{N_3} \right) \tag{102}
\]

Eqs. (98), (100) and (102) provide enough information to calculate the five unknowns, namely, $(R_2/R_1)$, $(R_3/R_1)$, $(R_4/R_1)$, $(R_5/R_1)$ and $\Delta$. We get one redundant equation, which could be used for consistency check on $(R_3/R_1)$, or could be used for estimating the length errors in the case of a short coil. To know the absolute values of the radii, we need to independently determine just one radius. This could be obtained easily if a reference field is available. Otherwise, the radius of one of the windings has to be estimated from mechanical measurements. It is generally a good idea to choose the winding with the minimum number of layers of wire for such an estimation. If another well calibrated coil is available, the radii can also be calibrated against this coil.

10.1.2 Calibration of the angular positions

If reference dipole and quadrupole magnets are available with precisely known phase angles, the angular positions $\delta_1$, $\delta_2$, $\delta_4$, and $\delta_5$ can be easily determined from the phases of the dipole and the quadrupole components of the measured signals. Due to a possible tilt in
the plane of the tangential coil, the apparent angular position of the tangential winding, $\delta_3(1)$, determined in a dipole magnet will not be the same as the apparent angular position, $\delta_3(2)$, determined in a quadrupole magnet. These values, however, can be used to obtain the true angular position, $\delta_3^0$, and the tilt parameter, $(\varepsilon/R_3)$, using the relations (see Sec. 8.5):

$$\delta_3(n) = \delta_3^0 + \lambda_n; \quad \lambda_n = \left( \frac{1}{n} \right) \tan^{-1} \left( \frac{n\varepsilon}{R_c} \right) \cot \left( \frac{n\Delta}{2} \right) \quad (103)$$

If the absolute phase angles of both the dipole and the quadrupole fields are not known, we cannot use Eq. (103) for calibration. In this case, we again make use of a sextupole field. In a dipole field, we can determine the quantities $\delta_2 - \delta_1$ and $\delta_3(1) - \delta_1$ from the observed phases of the dipole components of the voltage signals. Similarly, in a quadrupole field, we can determine the quantities $\delta_3 - \delta_4$ and $\delta_3(2) - \delta_4$. Finally, $\delta_2 - \delta_1$ and $\delta_3(3) - \delta_1$ are determined in a sextupole field.

Combining the data from the dipole and the sextupole fields, we get the quantity $\delta_3(3) - \delta_3(1)$, which depends on $(\varepsilon/R_3)$ and the opening angle, $\Delta$. Since $\Delta$ is obtained along with the calibration of the radii, the parameter $(\varepsilon/R_3)$ is determined. Knowing $(\varepsilon/R_3)$, one can calculate $\delta_3(2) - \delta_1$, which can be combined with the data in a quadrupole field to get $\delta_4$ and $\delta_5$ also relative to $\delta_1$. All angles are thus known relative to one of the windings. For coils equipped with a gravity sensor, the absolute angles can be obtained by making measurements from the lead and the non-lead ends of a magnet (see Sec. 8.8.2). For other systems, such as for measurements in a vertical dewar, absolute values of angles are often unnecessary. The design value of $\delta_1$ can be used in this case and all other angles can be determined from the procedure described here. All the parameters of interest for carrying out harmonic analysis using such a tangential coil are thus determined. As shown in this section, all the parameters relative to one of the windings can be obtained without any knowledge of the field strengths or directions of the magnets used for the calibration.

**ACKNOWLEDGEMENTS**

The author would like to thank J. Billan of CERN for providing Fig. 10 and details of the measuring coils planned for the LHC magnets and P. Schmüser of DESY for permission to use Fig. 8. Many helpful discussions and suggestions from P. Wanderer and J. Herrera are gratefully acknowledged. This work was supported by the U.S. Department of Energy under contract no. DE-AC02-76CH00016.

***

**REFERENCES**


Abstract
First the main characteristics of the NMR magnetic measurements, such as accuracy, independence of field direction and zero temperature coefficient are recalled. Then some different magnetic field measuring techniques using NMR, and the conditions to achieve such measurements, are described. Finally, recent NMR applications, used in various domains such as MRI and accelerator-magnet alignment, are described.

1. BASIC NMR REMINDER
1.1 Classical theory
If we consider the proton or nucleus which has kinetic and magnetic moments in an external magnetic field, we have a precessing movement; the spinning mass of the nucleus will act as a gyroscope and the forces between the magnetic moment and the external field tend to align the magnetic moment with the field; the proton will process around the main field direction as shown in Fig. 1.

The precession frequency varies essentially linearly with the field amplitude and is not affected by the temperature of the environment. It is very easy to measure since frequency measurements can be made with a precision of about $10^{-10}$ to $10^{-12}$. This frequency measurement gives the value of the magnetic field modulus, independently of its direction. The field direction has influence only on the signal amplitude.

For a better understanding of this theory, please refer to Abragam’s book mentioned in the Bibliography, at the end of this paper.
1.2 Quantum theory

The levels of energy of the nucleus separate in the field and the energy between the two states is proportional to the field amplitude. In pulsed NMR, when a RF magnetic field pulse is applied for a given time perpendiculary to the main field, the energy levels are pumped to the upper state. When the magnetic field is removed, they will decay giving the frequency corresponding to the difference of energy of the levels. In both cases, we have a relation between the field amplitude and the frequency; this relation is $\gamma$.

\[ \Delta E = h\nu = \mu B / I \quad \gamma = \mu / hI = \nu / B \]

Fig. 2 Energy separation in a magnetic field

1.3 Nuclei commonly used in NMR magnetometers

Different nuclei are used in NMR magnetometry:

The proton is mainly used; it is the nucleus for which we have the more precise evaluation of $\gamma$, so it is used as a primary standard. Probably the best measurement of $\gamma$ has been made by NIST (The National Institute for Standards and Technology in US) [1]. In Metrolab magnetometers, we use samples in the form of natural rubber pieces which are easier to handle for probe fabrication.

Deuterium is used to measure higher magnetic fields, and we use it in our NMR magnetometer which is derived from the CERN development by Borer and Frémont [2]. For frequencies up to 100 MHz, we can measure magnetic fields up to 14 T. We use the deuterium as heavy water contained in sealed glass ampoules.

Fluor is also used as a measuring nucleus in MRI equipments since its $\gamma$ differs by 5% from the proton value, allowing field measurements simultaneous with imaging, with no interference with the imaging system frequency.

Aluminium powder is also used, mainly in cryogenic equipments

He3 has also been proposed by Jim Clark from UCLA for use in the LHC.

Electron Paramagnetic Resonance (EPR) has also been used in our standard NMR apparatus, to make a probe measuring low fields in the range 5 to 30 gauss, with the same controller and the same frequency range as the standard NMR measurement system.

Other nuclei such as Li, Na, Cs etc. can also be used.

A summary of $\gamma$ values can be seen in the following table:
1.4 Proton gyromagnetic ratio

The gyromagnetic ratio $\gamma$ of proton has been measured by different means and the official value was given by NIST in 1986. In 1990, a small correction of 1.5 ppm was applied, with a relative uncertainty of 70 ppb. For rubber samples, we recorded a 3 ppm shift in the $\gamma$ value, compared to the official water sample value. Our NMR instruments equipped with rubber samples have therefore been corrected accordingly. For an absolute estimation of uncertainty, we give a conservative value of $\pm$ 5 ppm and propose a $\pm$ 2 ppm error on the value of the magnetic field measured by our instruments. However, it is very hard to prove this because of the difficulty of transporting a field standard. It is a challenge for the G-2 experiment people in Brookhaven where they need to insure the traceability with the NIST standard up to the precision of $10^{-7}$, since it is a real problem to get magnets as stable in time as in space at this level of precision.

The practical $\gamma$ value used in the Metrolab magnetometer is given in the following table:

<table>
<thead>
<tr>
<th>Source</th>
<th>$\gamma$ value (MHz/T)</th>
<th>Relative uncertainty (ppm)</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Codata 1986</td>
<td>42.576 375(13)</td>
<td>0.3</td>
<td>NIST Journal, Vol 92, Nb 2, p 85, March/April 87</td>
</tr>
<tr>
<td>1990 correction</td>
<td>42.576 396(3)</td>
<td>0.07</td>
<td>NIST Journal, Vol 95, Nb 5, p 521, Sept/Oct 90</td>
</tr>
<tr>
<td>Metrolab</td>
<td>42.576 268(3)</td>
<td></td>
<td>Theoretical value: -3 ppm due to para rubber/Brucker 200 MHz spectrometer</td>
</tr>
<tr>
<td>Metrolab</td>
<td>42.576 255</td>
<td></td>
<td>Practical value: -0.3 ppm accepted error due to technical reason</td>
</tr>
</tbody>
</table>

2. NMR MAGNETOMETER PRINCIPLE

2.1 Continuous Wave (CW)

2.1.1 Q meter

Our standard NMR magnetometer, developed 20 years ago at CERN by Borer and Frémont [2], uses a water sample in a RF coil placed in the magnetic field, and is low-frequency modulated by an additional pair of modulation coils (see Fig. 3). It is a tuned-coil and capacitor system, fed with continuous RF, using the Q-meter principle for which the peak-to-peak RF level on the coil is detected by a diode system. The signal is very small so it is amplified in an AC amplifier and allows the DC component and its drift problems to be rejected.
The B modulation used is generated by a simple circuit (Fig. 3). The constant frequency RF source is applied by a small coil around the sample. This coil detects the NMR signal only, in contrast to a frequency modulation where the signal can lock on the beating with the external disturbing frequency. Systems with frequency modulation (F modulation) don’t need additional modulation coils, leading to simpler probe design. However, some drawbacks may appear such as modulation noise due to the frequency excursion on the tuning curve.

2.1.2 Closed-loop principle

In our standard magnetometer system, we use a Variable Controlled Oscillator (VCO) which is locked on the NMR signal (Fig. 4).

When the NMR signal crosses a threshold, it generates a pulse freezing the modulation voltage, proportional to the current in the modulation coils. This voltage is used as an error signal and is sent to the integrator generating the control voltage of the VCO to lock onto the NMR frequency (Fig. 5).
2.1.3 Open-loop principle

We have developed, for MRI magnet measurement, an open-loop principle. It uses a Direct Digital Synthetizer generator (DDS) where frequency is controlled by a computer, and need not be measured a posteriori. This RF is modulated with exactly known parameters. We use, for MRI measurements, probe arrays (up to 30 probes) that can be fed by the same frequency modulated RF signal, so there is no danger of beating between probes (Fig. 6).

The offset from a reference frequency \( F_0 \) is loaded into a counter for each probe. In each counter, we record the instant the signal appears, going up and down. We take the mean value of these two numbers to get the value of the field (Fig. 7). So, without the need of a multiplexer system, we can measure simultaneously all the values of all probes in a couple of seconds. Then, for each probe, we analyse the mean value and standard deviation, for a number of periods (about 50 for example) with the microcomputer. If we have a gross error, we discard this measure from the mean. The RMS value of the measures also gives information on the confidence level of that measurement. The resolution we achieve with this principle is better than expected, getting below \( 10^{-7} \). Therefore, we proposed this system for the G-2 experiment in Brookhaven.

![Fig. 6 Metrolab multiprobe system used in MRI probe arrays](image)

2.1.4 Z meter

Another technique, for future magnetometers using higher frequencies, up to 1 GHz, is in development at Metrolab. The principle is not new [3]; it consists of measuring the impedance change in the sample coil. The coil containing the sample is placed at the end of a perfectly adapted coaxial cable; the impedance imbalance at resonance is detected with a hybrid T as in Fig. 8.

![Fig. 7 Principle of detection of the multiprobe system](image)
When the proton perturbs this balance, there will be a signal that can be detected with a phase detector, and used to produce a feedback signal to control the VCO and lock onto the magnetic field. This principle uses higher modulation frequency, allowing the magnetic field to be tracked more quickly. The use of a lock-in detector also permits a good recovery of weak signals, even with a signal-to-noise ratio lower than 1. The high frequency range also allows protons to be used to measure field amplitude up to 20 T. Probe design is also simplified, since it consists only of an adapted coil placed at the end of a coaxial cable. Using appropriate material, we should be able to make cryogenic probes using this principle.

2.2 Pulsed NMR

The pulsed NMR technique (Fig. 9) is a very clever method for NMR spectroscopy, chemical analysis and analysis of nuclei in molecules [4, 5]. In fact, chemical bondings in molecules split energy levels of nuclei with magnetic moment; so they get different $\gamma$ values in resonance, displaying a spectrum of frequencies instead of the original frequency of the free nucleus.

To obtain a spectrum, a RF pulse is applied to the sample to tilt the magnetic moment of the nucleus, the tilt of the nucleus spin being determined by the magnitude of the RF pulse. After a given time, the pulse is switched off, and the signal induced by the precessing moments is observed and processed by FFT until they are completely aligned with the external magnetic field (Fig. 10).
2.3 Flowing-liquid principle

The principle has been used in Bratislava by L. Jansak and S. Kvitkovic (and also by J. Tatarczuk in Mainz, Ye Sheng in Wuhan, China, etc...), to measure the field by water transportation. This water transports spin direction, after passage into a polarizing magnet (Fig. 11).

The water proton takes several seconds to align with the field and several seconds to return to a random direction. Once the proton is aligned, water is transported to the field to be measured. A frequency synthesizer emits RF energy which depolarises the nucleus if the frequency matches the resonance. With the auxiliary analyser, only the strength of the signal is observed, not the frequency. When the signal becomes smaller, it indicates that the frequency of the synthesiser is exactly the resonance frequency of the proton.

FID is the Free Induced Decay.

\( T_2 \) is the energy relaxation time between the nucleus and the surrounding network [6].

\( T_1 \) is the relaxation time between the nucleus and field \( B_0 \) (about 3,6 S for pure water).
3. MEASURING RANGE AND ACCURACY

3.1 Range

Field measurements can be made, with NMR methods and with the different techniques described above, down to the earth magnetic field (0.05 mT). Current magnet measurements, in MRI and particle accelerators, are in the ranges:

- 0.04 T to 2 T with protons as the sample in the probes
- 2 T to 14 T with deuterium
- 0.5 to 3.2 mT with EPR sample.

3.2 Accuracy

The accuracy we can obtain with the NMR principle can be affected by some parameters, but the gyrometric ratio $\gamma$ is not a source of error in the measurements. Some years ago, Brucker gave an estimation of the temperature coefficient of $\gamma$ inferior to $10^{-7} / ^\circ C$. Later, we did some measurements on this effect and essentially found a $T_c$ value of zero, between 25°C and 60°C with an uncertainty of $\pm 50 \text{ ppb} / ^\circ C$. This interpretation can be discussed in the case of the presence of paramagnetic material in the surrounding of the probe, since this can have a magnetic $T_c$ different from zero and affect the accuracy.

With a pulsed NMR magnetometer, we can achieve even better accuracy, thanks to the long relaxation time. For example, with a very homogeneous field, it is possible to have an uncertainty of a couple of hertz on a 600 MHz frequency, corresponding to a few ppb.

In the CW NMR magnetometer, currently, for a 1 second measurement, we get a relative accuracy of $0.5 \text{ ppm} \text{ RMS}$. If we do a mean of several measurements, we can even reach $0.1 \text{ ppm}$ for a 10 second measurement duration.

4. MEASURED-FIELD PROPERTY

4.1 In time

The field to be measured must be stable, especially if fields are to be measured with a precision up to 1 ppm. In the Borer system from CERN, we have a tracking speed not exceeding 1% of field variation per second, which is quite slow. With the RF-principle magnetometer, we expect to be able to follow the field more quickly since the modulation frequency is much higher.

4.2 In space

4.2.1 Definition

The magnetic field must also be stable in space. Of course, measurements with an accuracy of 0.1 ppm and if homogeneity is about 1000 ppm / cm, you will see that 0.1 ppm corresponds to a dimension of 1 µm. So, to use NMR magnetometers the field must be homogeneous for two reasons:
(i) probe positioning must be achieved with a spatial position precision proportional to field gradient, which can make impossible positioning with normal means.

(ii) the signal deteriorates in high gradient fields proportionally to the sample dimensions. The effect of the gradient is to spread and weaken the resonance line making it undetectable. So, in a non-homogeneous field, it becomes difficult to detect the NMR signal and to lock onto it.

4.2.2 Example

In our probes, the sample dimension is about 4 mm, giving a base of estimation of maximum gradient. The limit for good measurements is about 1000 ppm/cm for protons and •100 ppm/cm for deuterium. The table below gives the limits of field homogeneity for the standard Metrolab probes.

<table>
<thead>
<tr>
<th>Probe N°</th>
<th>Field range (Tesla)</th>
<th>Probe type</th>
<th>Active volume Ø x L mm</th>
<th>Part of the probe range</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.043 – 0.13</td>
<td>¹H</td>
<td>7 x 4.5</td>
<td>600 900 600</td>
</tr>
<tr>
<td>2</td>
<td>0.09 – 0.26</td>
<td>¹H</td>
<td>5 x 4.5</td>
<td>1200 1600 1200</td>
</tr>
<tr>
<td>3</td>
<td>0.17 – 0.52</td>
<td>¹H</td>
<td>4 x 4.5</td>
<td>1200 1400 1400</td>
</tr>
<tr>
<td>4</td>
<td>0.35 – 1.05</td>
<td>¹H</td>
<td>4 x 4.5</td>
<td>1500 900 800</td>
</tr>
<tr>
<td>5</td>
<td>0.70 – 2.1</td>
<td>¹H</td>
<td>4 x 4.5</td>
<td>250 600 350</td>
</tr>
<tr>
<td>6</td>
<td>1.5 – 3.4</td>
<td>²H</td>
<td>4 x 4.5</td>
<td>240 280 280</td>
</tr>
<tr>
<td>7</td>
<td>3.0 – 6.8</td>
<td>²H</td>
<td>4 x 4.5</td>
<td>300 180 160</td>
</tr>
<tr>
<td>8</td>
<td>6 – 13.7</td>
<td>²H</td>
<td>4 x 4.5</td>
<td>50 120 70</td>
</tr>
</tbody>
</table>

4.2.3 Local compensation of the field gradient

A method to make measurements in non-homogeneous fields is to use a set of compensating coils, intended to locally compensate the gradient (Fig. 12). The coil set is designed to create a gradient inverse to the local field gradient without modifying the measurement. For this, it is important to place the geometrical centre of the coil set at exactly the centre of the sample of the NMR probe.

Fig. 12 Principle of compensating coils
5. SOME SPECIAL MEASUREMENT CONDITIONS

5.1 On-the-fly measurement of fast ramping magnet

A special use of NMR is as a stand-by system when there are pulses of magnetic field or fast-programmed field variations as in the CERN PS. It gives some reference points in the cycle of the magnetic field. The NMR magnetometer is fed with a given frequency and we wait for the signal to appear when the corresponding value of the field is reached. The difficulty is to manage a proper field variation speed, in order to get a good NMR signal (Fig. 14). Another difficulty is to evaluate the delay between the signal detection and the real NMR resonance. This can be done experimentally and you can get then an accuracy of about 2 ppm.

Fig. 14 Choice of field measurement points according to the slope

5.2 Measurement of Tokamak magnets

At the Princeton TFTR Tokamak fusion magnet, measurements have been made to control the magnetic alignment of the toroidal coils shown in Figs. 15 and 16. In this arrangement the field is proportional to 1/R, the main radius of the annulus, and it is highly inhomogeneous, so the compensating coils presented previously have been used to allow NMR measurement.

Fig. 15 Description of geometry
Fig. 16 Assembly of the coils
Such measurements have also been successfully made in Cadarache (France) on a similar magnet with the same type of gradient-compensation-coil system. An optical positioning system has been used for this purpose (Fig. 17).

5.3 Field cycle of the CERN SPS (Fluxmeter measurement backed with NMR)

In order to measure with high precision and good time resolution the field in the CERN SPS cycle (Fig. 18), showing plateaus at 600 gauss (1 second) and 2 T (2 seconds), magnetic measurements have been made [7] by a system (Fig. 19) including:

A Precision Digital Integrator (PDI) recording the complete shape of the cycle, with a high resolution in amplitude and time, but possible slow fluctuations in offset and gain.

Two or three NMR probes controlled by a multiplexer and PT 2025 magnetometer chosen to measure, with NMR precision, at least two points of the cycle, if possible on a plateau. This system gives high precision points to the profile of the magnetic field cycle, allowing the curve recorded by the PDI to be calibrated exactly.
The overall precision obtained is about $10^{-5}$, with 1 µS response time. The comparison of the fixed NMR points to the corresponding output of the PDI, at each cycle of the field, gives a measure of the PDI drift and allows to verify whether this drift is linear and then permits it to be compensated continuously, producing a PDI accuracy of $10^{-5}$. This system has given good results over the last 3 years.

5.4 CERN PS deflection magnet

We have also carried out measurements on the inhomogeneous fields in a magnet of the CERN PS. There is a rule of thumb that states that if you have a long cylindrical-symmetry-gap magnet, you can expect to find an appropriate zone to place an NMR probe in the middle plane of the gap. Closer to the end of the poles, the signal would be poor. If an inner finger is needed in a vacuum chamber, its material must be chosen with extreme care. If you use welded stainless steel, there will be an alteration of the metallurgy in the welding area and ferromagnetic properties may appear, perturbing the homogeneity of the field or even destroying the signal. It seems best to use, for an assembly, stainless steel for machined pieces and brass for welded pieces. For brass however, the material should be tested before use, since some samples show unwanted magnetic properties.

5.5 Measurements on MRI magnets

This is achieved with probe arrays. These arrays are half-circle shaped and equipped with 12 to about 30 probes, depending on the models (Fig. 22). By turning these arrays
around their diametrical axis, it is possible to achieve, on a spherical surface, a great deal (up to 500) of measurement points in less than 5 minutes. This permits the field of MRI machines to be controlled and shimming of the superconducting magnet.

Fig. 22 Example of a Metrolab 12# probe array

With such a probe array, it is important to have a good calibration from one probe to another. A difference may appear if pieces of paramagnetic material or small electronic components are near the probes. These may slightly distort the flux lines of the magnetic field and then create small discrepancies with adjacent probes. This requires a calibration of each probe array, in a very homogeneous field, and a correction table for each probe of the array. Thus, we achieve a relative accuracy of 0.1 ppm between each probe of an array so that a MRI magnet can be mapped within 0.1 ppm accuracy. For this technique a frequency modulation system, or open-loop principle, described in Section 2.1.3, giving a 2 – 3 Hz precision on a 40 MHz frequency, which corresponds to less than 10^-8.

With such equipment, fitted with a good clock, temperature-controlled oscillator and a tightly fixed probe, we can estimate the decay of a superconducting magnet, for 1- or 2-hour intervals to an accuracy of 10^-8 per hour with good confidence.

5.6 Automatic probe calibrator in Saclay

Another common use of NMR measurement is performed in CEA Saclay by M. Tkatchenko and C. Evesque, who have set up an automatic system to calibrate Hall sensors. Several probes are connected to a multiplexer. In this system, one point is at 30 gauss, with an EPR probe, to partially fill the gap towards low fields (< 400 gauss). This development has been achieved in a collaboration between CEA Saclay and Metrolab.

REFERENCES


BIBLIOGRAPHY

A. Abragam, Principles of Nuclear Magnetism, Clarendon, Oxford.
HALL GENERATORS

J. Kvitkovic
Institute of Electrical Engineering, Slovak Academy of Sciences, Bratislava, Slovak Republic

Abstract
The use of Hall generators is rapidly increasing in electrical engineering as well as in physics experiments. This chapter starts with a brief historical review of these generators followed by the Hall effect theory. Then follows a detailed analysis of Hall-generator parameters and a review of the various applications of these devices.

1. INTRODUCTION

The Hall effect has two important properties: the possibility to measure constant and varying magnetic fields, and the ability to perform multiplication. It is a very powerful tool for the determination of the semiconductor material parameters. The Hall effect has been used for many years in physics, but only after the discovery of the new semiconductor materials have the interesting and useful properties of the Hall generator been used in practical instrumentation and in industry. Recently developed cryogenic Hall generators can be used in superconducting research, magnet testing and sample investigation.

2. GALVANOMAGNETIC PHENOMENA

2.1 Historical review

The Hall effect was discovered by Edwin H. Hall in 1879 at the Johns Hopkins University in Baltimore, USA [1]. The discovery was not made accidentally as is often the case during research on other phenomena. Hall, as a graduate physics student, had been inspired by Maxwell’s book on magnetism. In Hall’s paper of 1880, he describes the measurements obtained on an iron foil where he found the coefficient (later Hall coefficient) of iron to be about ten times larger than that of gold and silver. In the next year he studied nickel and cobalt. Magnetoresistance (Gauss effect)—increase of resistance of a conductor in a magnetic field—was discovered by W. Thomson in 1856. The first report about the magnetometer using the Hall effect in germanium was made by G.L. Pearson in 1948 [2]. More extensive use of Hall generators was made after 1952 when the technology of InSb production was developed.

2.2 Transport phenomena

Let us imagine a semiconductor material with free electrons, no external fields (electric or magnetic) and no thermal gradients. The electrons collide occasionally with one another and with lattice atoms. They have different velocities and their velocity distribution is determined by the temperature. In the absence of an electrical field the mean velocity of the electrons is zero and no current flows through the semiconductor but when we apply an external field the charged particles are accelerated. The movement of the charged carriers causes charge transport and is therefore called transport phenomena. In cases where charge transport is possible, it is due to the conductivity of the material. In the next step if we apply the external magnetic field to the semiconductor with internal current density the electric field strength has not in general the same direction as current density. The angle between them is
called the *Hall angle*. In semiconductors and metals three types of transport effects can be observed as shown in Table 1.

**Table 1**

Transport effects in metals and semiconductors

<table>
<thead>
<tr>
<th>Effect</th>
<th>Magnetic field</th>
<th>Thermal gradients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Galvanomagnetic</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Thermoelectric</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Thermomagnetic</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

### 2.3 Magnetoresistance

Magnetoresistance is defined as a change in resistance of a sample produced by an applied magnetic field as illustrated in Fig. 1 (magnetoresistor dependence). Charge carriers drifting in a semiconductor material under the combined action of transverse electric and magnetic vectors are subject to the Lorentz force that deflects the charge carriers and produces the Hall field. If all the charge carriers are identical then the Hall field compensates the Lorentz force, the current lines are invariant with magnetic field, and in consequence the magnetoresistance $\Delta \rho/\rho_o$ is zero. If the charge carriers are not of the same charge type then the Lorentz force is not fully compensated by the Hall field. This produces a current transverse to the electric and magnetic vectors and a change in resistivity of a sample which can be measured along its longitudinal direction of current flow. This is *physical magnetoresistance* while the changes in magnetoresistance caused by changes in shape and dimensions of samples are called *geometrical magnetoresistance* [B1]. Magnetoresitors used this effect, but due to the very high temperature coefficient of the magnetoresistivity and the nonlinear output dependence on the magnetic field they were rarely used for magnetic field measurements.

![Fig. 1 Basic difference between Hall effect and magnetoresistance](image)

### 2.4 Hall effect

Electrons, considered as charged particles, drift in the direction of an electrical field $\mathbf{E}$ with a velocity $\mathbf{v}$ driven by the force $\mathbf{F}_e = -e\mathbf{E}$. In the presence of a transverse magnetic field $\mathbf{B}$ the drifting electrons are subjected to the Lorentz force $\mathbf{F}_m = -e(\mathbf{v} \times \mathbf{B})$ and some of electrons are deflected in a direction orthogonal to both the electric and magnetic field vectors producing a Hall field which compensates the Lorentz force. The combination of electrically and magnetically-induced forces $\mathbf{F} = -e(\mathbf{E} + \mathbf{\tau} \times \mathbf{B})$ acts on the electrons and produces the *Hall effect* [B2]. If the mobility of the electron is defined as:
\[ \mu = \sqrt{E} = (-ev) / F \]  
(1)

Ohm’s law as:

\[ J = \sigma E \]  
(2)

and the material is characterised by conductivity:

\[ \sigma = ne\mu \]  
(3)

then for an homogeneous, isotropic, and rectangular Hall generator infinitely long with point Hall and ohmic contacts not connected to the load, the Hall voltage is equal to:

\[ U_H = (R_H IB) / t \]  
(4)

where the Hall coefficient \( R_H \) for electrons is defined by:

\[ R_H = -1/(ne) = -\mu / \sigma . \]  
(5)

For Hall generators with finite dimensions it is necessary to include a geometrical correction factor \( G \):

\[ U_H = (R_H I B) G(a/b, s/a, B) / t \]  
(6)

where \( J \) is the current density, \( \sigma \) the conductivity, \( n \) the concentration of electrons, \( I \) the Hall generator control current, \( t \) the thickness of the active area, \( a, b \) the Hall generator dimensions.

2.5 Parasitic effects

The Hall generator current density is distributed inhomogeneously due to the presence of a magnetic field and local inhomogeneities in the semiconductor. Localised Joule heating produces large thermal gradients and these in turn can affect the properties and behaviour of the Hall generators. The Hall generator contacts (contacts between semiconductor and metal) can be regarded as thermocouple junctions which produce thermoelectric voltages. Parasitic effects from the point of view of the Hall generator are thermoelectric and thermomagnetic effects. These effects occur at the same time as a Hall effect and can increase measurement errors.

\textit{Seebeck effect.} This causes the generation of thermoelectric voltage. If a semiconductor plate with metallic electrodes is heated inhomogeneously in a longitudinal direction a voltage which is proportional to the temperature difference can be measured between the electrodes.

\textit{Peltier effect.} This is the reverse of the Seebeck effect—the generation of the thermal flow by the electrical current. Potentials generated by thermomagnetic effects (\textit{Eitnghausen, Nernst, Right-Leduc}) are usually negligible in comparison with transverse Hall voltages unless large thermal gradients are present [3]. It is important therefore to reduce such gradients to a minimum. All thermomagnetic effects are proportional to the magnetic induction. Due to this fact great attention must be paid in the measurement of high fields.

3. FABRICATION OF HALL GENERATORS

The Hall generator consists of a thin semiconductor plate of dimensions \( a \times b \times t \) equipped with four contacts (see Fig. 2). The control current is supplied by two current contacts, CC, while Hall contacts, HC, used for measurement of the Hall voltage are placed on the plate sides. The Hall generator is most sensitive to the component of magnetic field perpendicular to the semiconductor plate plane.
Bulk material Hall generators are prepared by cutting from an ingot [4], followed by abrasion, polishing and etching techniques. The thickness should be reduced to a minimum in order for the Hall voltage to be as large as possible [5]. Semiconductors, in contrast to metals, are extremely brittle and have to be glued to a ceramic substrate which provides mechanical support and protection for the fragile Hall generator.

Thin films prepared by vacuum deposition of intermetallic $A^{m}B^{y}$ compounds on isolating substrates are used for the construction of miniature Hall generators. The semiconductor is formed to the desired shape by photolithography and etching. The electrodes are soldered to copper wires or, for miniature Hall generators, wire bonding is used. Like other semiconductor devices Hall generators must be encapsulated in order to protect them from light, humidity, dust, chemical corrosion and other environmental influences. The package also provides for the electrical connections of the chip with the external circuits. The packages of Hall generators must be non-magnetic and the thermal expansion of the chip and package must also be taken into account, since an inadequate combination of thermal expansion coefficients may lead to an additional offset due to mechanical stress. For high stability measurements unpackaged Hall generators are also viable [6].

### 3.1 Hall generator materials

In the design and fabrication of Hall generators it is very important to choose the proper material and they are usually prepared from n-type semiconductors where the dominant charge carriers are electrons which have much higher mobilities than holes. The following materials can be used: Ge, Si, InSb, InAs, GaAs, etc [B3-B6]. Germanium, used in earlier years, is not in extensive use nowadays. InSb and InAs are the favoured materials for present day use due to their high mobility. Micron-thin InSb films can be grown from liquid or vapour-phase epitaxy on insulating substrates. Films less than 5 $\mu$m in thickness have electron mobilities smaller than those of bulk InSb so that the increase of sensitivity due to decrease of thickness is compensated. Pure InSb has a strong temperature dependence of its electrical parameters because of its small bandgap, Fig. 3, curve a). In order to decrease this temperature dependence donor impurities must be introduced into it, curve c). In comparison with InSb, InAs has a larger bandgap and consequently a reduced temperature dependence of parameters. However its electron mobility is much smaller and it is more difficult to grow in the homogeneous form. The basic material parameters of Hall generator semiconductors are presented in the Table 2.
Table 2
Material properties of semiconductors for Hall generators at 300 K, [B6]

<table>
<thead>
<tr>
<th>Material</th>
<th>Eg [eV]</th>
<th>n [cm⁻³]</th>
<th>(\mu_n) [cm²V⁻¹s⁻¹]</th>
<th>-Rₚ [cm³C⁻¹]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ge</td>
<td>0.67</td>
<td>2.4x10¹⁵</td>
<td>3.9x10⁵</td>
<td>9x10⁴</td>
</tr>
<tr>
<td>Si</td>
<td>1.12</td>
<td>2.5x10¹⁵</td>
<td>1.3x10⁴</td>
<td>2.5x10³</td>
</tr>
<tr>
<td>InSb</td>
<td>0.17</td>
<td>9x10¹⁶</td>
<td>7.0x10⁴</td>
<td>70</td>
</tr>
<tr>
<td>InAs</td>
<td>0.36</td>
<td>5x10¹⁶</td>
<td>2.2x10⁴</td>
<td>125</td>
</tr>
<tr>
<td>GaAs</td>
<td>1.42</td>
<td>3x10¹⁵</td>
<td>6.4x10⁴</td>
<td>2.1x10³</td>
</tr>
</tbody>
</table>

Fig. 3 Temperature dependence of Hall coefficient of n-InSb impurity concentrations:
a) -10⁻¹⁴, b) -10⁻¹⁶, c) -10⁻¹⁷, d) -thin film 10⁻¹⁶ cm⁻³ [B7]

Recent developments in integrated circuit technology has led to the fabrication of super lattices and to the possibility of manufacturing GaAs/AlGaAs and other heterostructures. Such modulation-doped semiconductor layers can be used for the formation of quantum wells with two-dimensional electron gas. The thickness of the device’s active region is less than 10 nm and electron mobility is extremely high 300 000 cm² / V.s. These layers can be used for the design of scanning Hall probe microscopes with high spatial resolution, 0.85 μm, and very high magnetic field resolution, 2.9 x 10⁻⁸ T Hz⁻¹/₂ at 77 K [B10].

3.2 Hall generator types
Hall generators are developed and produced in a large number of shapes and dimensions depending on their application. From the user point of view the Hall generators (HG) can be divided into several types according to:

- operating temperature
  - cryogenic HG
  - room temperature HG
- magnetic field range
  - low field HG - 02 T
  - high field HG - 240 T etc.
• package
  - transverse HG - measurement in gaps
  - axial HG - measurement in holes
  - unpackaged HG - high stability generator

• active area size
  - normal HG - general purpose
  - microsize HG - inhomogeneous magnetic field measurement
  - number of elements on chip
  - single HG - one element
  - multisystem HG – 215 elements in a line- special
  - tangential HG - measurement of the tangential component of B
  - near to surface HG - measurement of perpendicular component of B very close to the sample surface
  - corner HG - centre of the active area is at the package corner
  - 3 axis HG - measurement of all three perpendicular components of B
  - multisensors - e.g. HG and temperature sensor in one package

This classification was simplified in order to achieve clarity for the reader. The details can be found in [6-8].

4. PARAMETERS OF HALL GENERATOR

4.1 Magnetic field sensitivity

The Hall generator magnetic field sensitivity is one of the most important parameters. This parameter is the ratio of Hall voltage variation to the variation of external magnetic field and all other parameters (such as control current, temperature, pressure, etc.) are constant.

The absolute magnetic field sensitivity \( S_a \) is defined as:

\[
S_a = |\frac{\delta U}{\delta B}| \quad \text{[V/T]} \tag{7}
\]

or

\[
S_a = \frac{(R_H I_G)}{t} \quad \text{[V/T]} \tag{8}
\]

The relative sensitivity \( S_r \) is defined as the ratio of the absolute sensitivity to the Hall generator control current:

\[
S_r = S_a/I = \left(\frac{1}{I}\right)|\frac{\delta U}{\delta B}| \quad \text{[V/AT]} \tag{9}
\]

4.2 Offset voltage

Offset voltage is defined as the output signal from a Hall generator supplied with nominal control current in the absence of a magnetic field, \( B = 0 \).
The output voltage of the Hall generator, $U_{\text{out}}$, in the presence of a magnetic field is:

$$U_{\text{out}} = U_{H} + U_{o},$$

where $U_{H}$ is the Hall voltage and $U_{o}$ is the offset voltage. Because the Hall voltage and the offset voltage are combined in the output voltage, the offset cannot be distinguished from the output signal if $B \neq 0$. The offset voltage is, in most cases, small and in the range of $\mu$V, and is linearly dependent on $I_{H}$ up to $I_{\text{max}}$. It is temperature dependent and this dependence is usually expressed as a temperature coefficient of the offset voltage. The offset voltage can be caused by several reasons: the Hall electrodes not being placed at the same equipotential line (misalignment voltage), Fig. 4 [B7], or the electrical asymmetry resulting from material inhomogeneities and also from external influences such as light, mechanical stress and variations in temperature.

![Fig. 5 Hall generator offset-compensation circuits [B7]](image)

The offset voltage can be compensated by means of various external electric circuits, Fig. 5 [B7] but one must keep in mind that no current flow is desirable in the Hall circuit. With each compensation the Hall generator current stability is reduced. However, it appears that a perfect correction of offset voltage cannot be made for all arbitrary control currents, such correction must be made only to those currents usually used in practice.

In the case of AC fields or flying-mode measurements the inductive offset voltage due to small loops in the Hall circuit can be observed. This effect is usually expressed in equivalent area dimensions.

### 4.3 Temperature dependence of sensitivity

The galvanomagnetic and material parameters of semiconductors used for the fabrication of Hall generators are in general temperature dependent as also are the magnetic field sensitivity, offset voltage, input and output resistances due to their Hall coefficient temperature dependence.

The temperature coefficient of sensitivity $\gamma$ can be defined as:

$$\gamma = (1/U_{H})(\delta U_{H}/\delta T) \quad [\text{K}^{-1}]$$

(10)

where $\delta U_{H}$ is the difference in Hall voltage caused by temperature change and $\delta T$ is the value of the temperature change.

The temperature dependence of the offset voltage is defined as:

$$\varepsilon = \delta U_{o}/\delta T \quad [\mu \text{V/K}]$$

(11)

If Hall generators are to have a temperature independent response over a wide temperature range, then they must be made of materials with sufficiently large energy bandgaps so that
intrinsically generated carriers represent a negligible fraction of the total carrier density in this temperature region. In materials with small bandgaps the donor impurity density must be high enough to extend the interval where the Hall coefficient is constant, Fig. 3 curve c). However, this procedure leads to a decrease in sensitivity owing to the lower electron mobility produced by ionised impurity scattering [7]. An alternative procedure is to use external circuit elements in conjunction with Hall generators in order to compensate temperature dependence. Loading the output or input terminals of the Hall generator with a thermistor can produce such compensation.

4.4 Linearity error

When using Hall generators in magnetometers it would be very desirable for the relation between Hall voltage and magnetic induction to be linear. Since this is not always the case, it is useful to define the linearity error as in Fig. 6.

\[
NL = \left(\frac{\Delta U_H}{U_H}\right) \times 100 \quad \text{[\%]}, \quad (12)
\]

or

\[
NL = \left(\frac{\Delta U_H}{U_{H_{max}}}\right) \times 100 \quad \text{[\%]}, \quad (13)
\]

where \(NL\) is the linearity error, \(\Delta U_H\) is the maximum deviation between the real Hall voltage dependence and the ideal linear Hall voltage dependence on the magnetic field, \(U_H\) is the Hall voltage at the point of maximum deviation and \(U_{H_{max}}\) is the Hall voltage at \(B_{max}\).

The Hall generator exhibits nonlinearity if its sensitivity (absolute and relative) depends on the magnetic field:

\[
NL = \left(\frac{\Delta S}{S_0}\right) \times 100 \quad \text{[\%]}, \quad (14)
\]

where \(\Delta S\) is the deviation from constant sensitivity value \(S_0\) at \(B = 0\).

The linearity error consists of two components [10]:

\[
NL = NL_m + NL_g \quad (15)
\]

where \(NL\) is the total linearity error, \(NL_m\) the material non-linearity and \(NL_g\) the geometrical non-linearity.

The material and geometrical non-linearities exhibit the same quadratic magnetic field dependence, but with opposite signs. Moreover, the values of non-linearity coefficients are of the same order of magnitude. These facts can be used to design a Hall generator in which the non-linearity effects are compensated or reduced, [10-12]. The geometrical \(NL\) may be
compensated by loading the Hall output with the appropriate value of loading resistor. The value of the linearity error depends not only on the Hall generator properties and the magnetic field interval employed but also on the type of linearity error definition and approximation criteria for ideal Hall voltage dependence. The measurement errors caused by non-linearities can be reduced by Hall generator calibration and data correction.

4.5 Input and output resistances

Input resistance is the resistance of the current circuit while output resistance is that of the Hall circuit. These are measured at a specified temperature usually (300, 77 or 4.2 K) in the absence of an external magnetic field and with open Hall and input terminals.

The output resistance must be as small as possible since the lower is the resistance the lower is the Hall generator output noise. A Hall generator is usually connected to the measuring apparatus by long connecting leads and cables, and it is more advantageous to have low resistance at the end of the cable. Moreover, a low input resistance Hall generator can be supplied with a higher control current and therefore higher sensitivity can be obtained. Since the input resistance increases with magnetic field the heating increases with magnetic induction. If the cooling is sufficient the heat generated in the semiconductor can flow out towards both surfaces of the active area.

Electrical excitation is a very important parameter for Hall generator operation. It is usually expressed as a value of nominal control current at which the Hall generator usually works. The control current (0.1 - 100 mA depending on Hall generator type) can be increased up to the specified maximum value. Never do experiments with the control current above the maximum value since the result will be a damaged Hall generator!

4.6 Frequency dependence

The Hall effect is frequency independent up to approximately 1 GHz. It means that the amplitude and the phase of the Hall voltage does not change with frequency as the magnetic induction and/or control current are driven at high frequencies. The frequency dependence of a Hall generator is defined as the amplitude ratio of \( U_\mu(f)/U_\mu(0) \) to the frequency of the magnetic field. It is mostly influenced by the input and output circuit parameters: connection leads, cables, waveguides and eddy currents in the semiconductor plate. Eddy currents have a strong frequency dependence. In applications where a Hall generator is placed in an air gap of a magnetic circuit (e.g. multipliers), the frequency response depends on the material parameters of the magnetic circuit and on the width of the air gap.

The equivalent figure of merit is the response time. It is defined as the time needed by the output signal to reach a certain percentage (e.g. 98%) of its final value following a step change in the magnetic field or control current. The response time of Hall generators is in the range 10^{-14} to 10^{-8} s [B11].

4.7 Directivity

The magnetic field sensitivity of the Hall generator depends on the angle \( \alpha \) between the magnetic field vector and the active area plane \( U_\mu = ((R_dIB)/h)/\sin \alpha \). If the B vector is perpendicular to the active surface of the Hall generator, the maximum of the Hall voltage is observed. Any deviation from the angle \( \alpha = 90^\circ \) reduces the Hall voltage. If the B vector lies in the Hall generator plane in the ideal case the output is zero. If the control current flow is not absolutely parallel to active surface, the vector B parallel to that surface generates an output signal.

Planar Hall effect

Measurement errors can be frequently caused by the fact that the magnetic induction vector is not precisely perpendicular to the Hall generator plane. In this case one component
of the magnetic field lies in the Hall generator plane and generates an additional voltage which is added to the Hall voltage. If the angle between $B$ and the $x$ axis is $\phi$, Fig. 7, then we can obtain the transversal component of electric field:

$$E_y = - \frac{\rho_o \cdot a \cdot B^2 \cdot j \cdot \sin 2\phi}{2}$$

where $a$ is the magnetoresistance coefficient and $\rho_o$ the resistivity without magnetic field.

From this expression it is clear that the transversal electric field is equal to $B^2$ and $\sin 2\phi$. This means that the error Planar voltage is zero for angles $0, \pi/2, \pi$, etc. as shown in Fig. 8.

We can use this property when we know the orientation of the component lying in the Hall generator plane. We could rotate the Hall generator so that the direction of control current is parallel or perpendicular to the magnetic field component in the Hall generator plane [14,15]. It is important to note that this error increases with increasing magnetic field. As presented in Fig. 8 the amplitude of the Planar voltage depends on the semiconductor material parameters. It is seen that the material with higher conductivity has a lower amplitude of the Planar voltage. For precise measurement of the magnetic field with unknown orientation it is desirable to use the Hall generator prepared from material with higher conductivity.

![Fig. 8 Planar Hall voltage dependence on the angle between B and I at 1 T. a) n-type InSb, $\sigma = 750 \ \Omega^{-1} \cdot \text{cm}^{-1}$, b) intrinsic InSb, $\sigma = 200 \ \Omega^{-1} \cdot \text{cm}^{-1}$ [B4](](image)

![Fig. 9 Hall generator absolute sensitivity dependence on magnetic field](image)

The polarity of a magnetic field can also affect the sensitivity i.e. the sensitivities for $+B$ and $-B$ are different $U_{H}(B) \neq U_{H}(-B)$, Fig. 9. The differences are in the range of 0.1 - 5%.
5. HALL MAGNETOMETRY

5.1 Calibration and precision

Since the sensitivity of the Hall generator after fabrication is not known, it is necessary to measure the sensitivity at least at one value of the magnetic field or, for high precision measurements, to calibrate it over the whole operational interval of magnetic field. In addition, the Hall generator output voltage is not a linear function of B, therefore a calibration at only two points is not sufficient. From the metrology point of view the calibration consists of comparing the measuring instrument with a common standard of the unit [16, 17]. The magnetic field standard is usually a coil supplied by a known current, the magnetic induction being calculated from the current and coil dimensions. Requirements for the standard coil are:

- high stability of geometrical dimensions
- possibility to measure precisely its dimensions
- high homogeneity of magnetic field in a working space
- free access to a working space
- high accuracy of B data.

The standards are usually the one-layer or Helmholtz coils with a diameter of 250–350 mm wound on a frame made from material with a low, linear, expansion temperature coefficient (e.g. flint, 5.10⁻⁷ K⁻¹). Another possibility is to use a Cooper or superconducting magnet with measuring equipment capable of measuring the B with a high degree of accuracy. This instrument is based on the NMR method [17] but the flowing-water NMR method can also be used in order to overcome the limitations of the classical NMR method [18, 19].

Since the magnetic field sensitivity of the Hall generator is temperature dependent, for high-precision measurements it is necessary to calibrate the Hall generator at the same temperature at which it will be used in practice. Also the direction of the control current and the polarity of Hall leads must be kept the same, i.e. the Hall generator must work at the same conditions as during calibration. It is very important to reach high accuracy and precision during calibration. These parameters are affected by: homogeneity [20] and stability of the magnetic field, stability of Hall generator current source, stability of the working space temperature, accuracy of the voltmeter used for measurement of the Hall voltage, and also by the accuracy of the NMR apparatus [1921]. During calibration at weak magnetic fields the Earth’s magnetic field and its variations must be taken into consideration.

Precision of the Hall generator measurement is defined as the repeatability of measured data customarily expressed in terms of standard deviation. This means that it is necessary to measure the same data several times under the same conditions. The inherent precision of a Hall generator, under the condition that the magnetic field, temperature and control current are constant, depends upon its noise level (e.g. a Hall generator with a sensitivity \( S_t = 100 \, \text{mV/T} \) and a noise voltage \( U_n = \pm 0.1 \, \text{μV} \) has a precision of approximately \( \pm 1 \, \text{μT} \)).

5.2 Compensation of temperature changes

For high-precision measurements in areas where the temperature might change and where the temperature coefficient is not sufficiently small it is necessary to eliminate the influence of temperature on the Hall generator. This task can be solved in several ways such as:

- a) proper selection of the semiconducting material it is possible to reduce the temperature coefficient of sensitivity (described in section 4.3)
- b) using a thermostat. In order to minimise the influence of ambient temperature changes the Hall generator can be placed in a thermostat at constant temperature. For room temperature measurements this temperature should be around 40 °C. The Hall generator is embedded in a split aluminium or copper block to insure uniform temperature
distribution. Around the block is wound a heater coil and a temperature sensor is placed in the vicinity of the Hall generator and both are connected to the temperature controller. The complete assembly can be encased in epoxy resin. In this way the Hall generator can be kept at constant temperature to within 0.1 K or better. Great attention must be paid to any additional background magnetic field generated by the heater which must therefore be bifilar-wound. The temperature controller must be proportional, not bistable, in order to avoid transients from switch-on to switch-off of the heater. Unfortunately, this technique results in large overall dimensions of the measuring equipment.

c) computer correction with simultaneous measurement of magnetic field and temperature [24]. Knowing the actual temperature, Hall voltage and from calibration tables of the Hall generator at various temperatures, it is possible to calculate the corrected value of the magnetic field. Also, the magnetic field error of the temperature sensor can be eliminated. The measuring method consists of auto-zero and auto-calibration procedures to reduce the thermal drifts of the amplifier and A/D converter. In this way the error of the magnetic field measurement can be reduced from 1 to less than 0.1% in magnetic fields of 0.6 T and a temperature range of 20–60 °C.

5.3 Noise

Hall generator noise is a basic parameter that determines the lowest magnetic field $B_{\text{min}}$ detectable by the Hall generator as well as the stability of the output signal. The typical noise spectrum of a Hall generator is shown in Fig. 10 where several different types of noise can be seen [B11]:

- **Low frequency noise** (LF) is caused by the Hall generator temperature variations. Slow temperature fluctuations generate thermovoltages on the Hall contacts which, however, can be eliminated by the use of Hall generator AC control current. In the case of measurement of a DC magnetic field with a Hall generator supplied with DC control current the noise can be reduced by filtering techniques or by an integration voltmeter in which the statistical noise will be averaged to zero. For high-precision measurements it is necessary to keep the Hall generator temperature stable. The fluctuations of the offset voltage are due to the instabilities of the heat exchange between the generator and its surroundings.

- **1/f noise** (1/f) is due to the current flow through the Hall generator. In the generation of this type of noise the surface-to-volume ratio and surface conditions play a dominant role. It shows different behaviours in the presence or absence of a magnetic field.

- **Generation-recombination noise** (G-r) is caused by spontaneous fluctuations in generation and recombination of charge carriers [25]. The fluctuation in charge carrier density causes the noise voltage in the direction parallel and perpendicular to the current flow. Very low levels of this type of noise are due to the short lifetimes of carriers in InSb and InAs.

- **Thermal noise** (T) (Johnson, Nyquist) is generated by the random motion of charge carriers in the Hall generator semiconductive material. The mean square noise voltage $<u^2>$ i.e. the average value of the square of the noise voltage generated by the resistor $R_u$ is:

  $$ <u^2> = 4kTR_uA_f, $$

  where $k$ is the Boltzmann constant, $T$ is the absolute temperature, $R_u$ is the Hall generator output resistance and $A_f$ is the bandwidth of the detector.
The magnitude of the measured Hall voltage is also important though its non-linearity and temperature dependence are not significant for many applications and may be compensated by means of external circuits or by numerical corrections. However, accidental fluctuations of the offset voltage and noise cannot be compensated by means of external circuitry.

5.4 Stability and ageing

Hall generator output signal stability depends on noise, stability of offset voltage and stability of magnetic field sensitivity at constant temperature and magnetic field. The output voltage in a magnetic field is made up of the sum of the offset voltage and Hall voltage. Long-term variation of the output voltage may be due to a variation of residual voltage or variation of Hall coefficient. These variations may be caused by a change in geometry of a Hall generator because of corrosion of the surface or electrodes. For stability tests it is necessary to keep the external influences as constant as possible i.e. constant control current, temperature and magnetic field. For the offset voltage testing the Hall generator must be placed in a zero-gauss chamber, to cancel the influence of the earth magnetic field. Evaluation can be made from a large set of measurements (to minimise errors) by statistical analysis. H. Weiss [B4] reported on InAs Hall generator experiments in which he found the mean error of individual measurements to be $1.4 \times 10^{-5}$. The offset voltage varied over a period of 8 months by less than the equivalent Hall voltage generated by a magnetic field of $10^{-5}$ T. At the same time the Hall coefficient remained constant to $\pm 2.10^{-5}$. The ageing process is very important for cryogenic Hall generators because they undergo thermal cycles between 300 K and 4.2 K. InSb generators after 100 thermal cycles change their sensitivity less than $\pm 0.04\%$ and the offset voltage change is $< \pm 5 \mu V$ [6]. These changes can be caused by active area material changes (microcracs), changes in contacts, solder, glue, resin, package etc. A first indicator of some of the instabilities is a change of the offset voltage or change in input and output resistances.

6. HALL GENERATOR APPLICATIONS

Hall generators already have innumerable applications and these continually increase. The direct application, measurement of magnetic field, was described in the previous section. The remaining are indirect applications in which the measured quantity is transformed to a magnetic field which is then measured by the generator. These applications result from the independent variables of the output voltage that is dependent on the product of the control current and magnetic field. The magnetic field also has two additional variables—the
distance of the magnetic field source from the Hall generator and the angle with respect to the generator plane.

Hall generators can be used to measure current, power, position, number of revolutions and pressure as well as for multiplication of two input signals, for brushless DC motor control, as a contactless switch and applications in the automotive industry. Some of the most interesting applications are described in the following sections.

6.1 Current transducers

Since an electric current is associated with a magnetic field Hall generators are able to detect and measure it so avoiding the need to insert an ammeter into the circuit. This method is especially advantageous for very large values of DC currents since it is not necessary to put a shunt into the circuit. A simple measurement of conductor tangential field is highly linear as non-linear elements such as an iron core are not present.

![Current transducer diagram](image)

In accordance with Ampere’s law we have: \( \int H \, dl = I \) and \( I = 2\pi r U_h/\mu_0 S_h \), where \( I \) is the measured current, \( U_h \) the Hall voltage, \( S_h \) the Hall generator sensitivity, \( r \) its distance from the conductor centre to the Hall generator and \( \mu_0 \) is permeability of free space. This simple method has several limitations. The magnetic field and the signal from the Hall generator are very weak so that background magnetic fields and ferromagnetic objects could cause disturbances to the current measurement. These disadvantages can be overcome, and the sensitivity of measurement can be increased, by using a ferromagnetic core with a Hall generator in its gap Fig. 11. This is the basis for a clip-on ammeter. Using this method currents as high as 400 kA have been measured with an accuracy of 0.52% [B7].

6.2 Contactless signal generation

In this application the Hall generator is used as a receiver for magnetic signals over short distances in order to obtain information about the relative position of two objects—receiver and transmitter. The transmitter is usually a small permanent magnet and the measurement is reduced to detecting the presence or absence of a magnetic field or to detecting the sign of the magnetic field. For this purpose only a very small magnetic induction is required while temperature coefficients have little or no influence on the result. Another advantage is that a permanent magnet requires no energy and therefore no connecting leads are necessary. Furthermore the magnitude of the signal is not dependent on the relative velocity between transmitter and receiver. The output signal is usually processed digitally. Amongst many examples of this application it is important to mention its use in keyboards, end-switches, in the automobile industry and as revolution counters and positioners (Fig. 12). The latter consist of permanent magnets with alternating polarity placed on a disc of non-magnetic material fastened to the shaft to be measured. In front of the disc is placed a Hall generator, the resulting voltage depending sinusoidally on the angle of rotation of this shaft. This method can be used also for the digital control of angular motion (with many magnets) or simply for counting revolutions (with one permanent magnet).
6.3 Displacement transducers

*Linear displacement transducers* depend on measuring along a constant magnetic field gradient with a linear Hall generator. Such transducers can be used as proximity switches.

*Angular displacement transducers* can be used for the measurement of angles especially where it is important to convert an angle into a voltage without sliding contacts. A circular magnetic circuit is made from ferromagnetic material with four air gaps for Hall generators. A permanent magnet is located at the centre and rotates with the shaft to be measured. Pairs of opposite generators are connected in series in such a way that their output voltages are additive. Transformation of the angles to the analogue signal is purely ohmic up to high frequencies without any distortion. These transducers have the great advantage that they are independent of environmental factors such as dust, smoke and moderate temperature changes.

7. CONCLUSIONS

Future development of Hall generators will take advantage of their high stability, sensitivity and linearity, their low residual voltage, temperature coefficient (of both Hall and residual voltages), input and output resistance and power dissipation as well as their small dimensions. They offer many advantages compared to other sensors: simple measuring arrangement, low cost and the possibility of making continuous measurements. Easy operation and simple electronics make them ideal for many experiments.

ACKNOWLEDGEMENTS

I would like to thank Dr. M. Majoroš for reading the manuscript and for his many valuable comments. Very little is my own work so I wish to express my gratitude to all cited authors for their contributions to Hall generator research and especially to E.H. Hall for his discovery of the effect.

REFERENCES


BIBLIOGRAPHY


MECHANICAL EQUIPMENT

G. Moritz
GSI, Darmstadt, Germany

Abstract
Mechanical equipment is an integral part of each magnetic measurement. In this contribution the importance of mechanical tolerances is demonstrated with some examples and the principal rules in order to reduce the influence of mechanical errors are discussed. In the second part typical benches are presented for almost every measurement task. In the appendix a non-complete list of companies which produce or sell mechanical equipment is given.

1. INTRODUCTION
The successful operation of an accelerator requires:
- extremely tight magnetic field tolerances which have to be proved by measurements
- a proper alignment of the magnets
Thus we need mechanical equipment
- to check the mechanical tolerances of the magnets
- to make accurate magnetic measurements
- to set precise fiducial marks which are accessible after the installation of the magnets in the accelerator
In the field of magnetic measurements the mechanical equipment is mainly used for the positioning and moving of the sensor (coil, Hall probe etc.). Its specific design depends on the applications, i.e. on the field property we want to measure (transfer function, field boundary, field quality etc.), on the accuracy that is required, and on the shape and the symmetry of the magnetic field.

2. THE INFLUENCE OF MECHANICAL TOLERANCES ON THE ACCURACY OF MAGNETIC MEASUREMENTS
The accuracy required for magnetic measurements leads usually to very tight mechanical tolerances. I will demonstrate this with four examples.

2.1 Rotating harmonic coil
The influence of the mechanical errors on the results of magnetic measurements with a rotating harmonic coil is covered theoretically in several publications [1–3]. The main sources of measurement errors during coil operation are:
- error due to transversal displacement of the rotational axis
- error due to angular shifts
These effects depend on the angular position.
Transversal displacement is caused by the coil not rotating in a perfect circle due to sag and bow-effects and imperfect bearings. This error is not important for dipoles.

Angular shift means azimuthal vibrations induced by a lack of stiffness of the shaft and couplings or by the stepping motor. This effect influences the results for quadrupoles and dipoles as well.

These mechanical imperfections lead to errors in the measured harmonic coefficients. The relation is given in the following formulas

\[ \varepsilon b_n = \frac{n}{2} \sum_{k=1}^{\infty} \left( \delta_k (b_{n+1-k} + b_{n+1+k}) + \varepsilon_k (a_{n+1-k} - a_{n+1+k}) \right) \]

\[ \varepsilon a_n = \frac{n}{2} \sum_{k=1}^{\infty} \left( \varepsilon_k (a_{n+1-k} + a_{n+1+k}) + \varepsilon_k (b_{n+1-k} + b_{n+1+k}) \right) \]

for transversal displacement

\[ \text{Displ} = R_{\text{ref}} \cdot \sum_k \left( \delta_k \cdot \cos(k\theta) + \varepsilon_k \cdot \sin(k\theta) \right) \]

and by

\[ \varepsilon b_n = \frac{n}{2} \sum_{k=1}^{\infty} \left( \delta_k (a_{n-k} + a_{n+k}) + \varepsilon_k (b_{n-k} - b_{n+k}) \right) \]

\[ \varepsilon a_n = \frac{n}{2} \sum_{k=1}^{\infty} \left( -\delta_k (b_{n-k} + b_{n+k}) + \varepsilon_k (a_{n-k} - a_{n+k}) \right) \]

For angular shifts [2].

\[ \theta_{\text{mean}} = \theta + \sum_k \left( \delta_k \cos k\theta + \varepsilon_k \sin k\theta \right) \]

\( \delta, \varepsilon \) are the relative amplitudes of the distortion and \( b, a \) are the normal and skew field coefficients.

A field harmonic present in the magnet can introduce erroneous terms of higher and lower orders, and the error is basically given by the product of the amplitude of the harmonic coefficient and the error amplitude. Thus the main contribution is due to the main field component.

In order to reach a relative accuracy of the higher harmonic coefficients of the order of \( 10^{-4} \) we need for example, at a typical radius of 40 mm, a stability of the axis of some \( \mu \text{m} \) and a precision of the angular measurement better than 0.1 mrad.

It is almost hopeless to fulfil these conditions.

### 2.2 Mapping of the Electron Cooler solenoid

In an Electron Cooling device the ėwarmí ion beam is cooled down by a ėcoldí electron beam. By elastic Coulomb collisions between electrons and ions travelling together through the field of the cooler solenoid the velocity distribution of the ion beam is reduced, the beam gets colder. Good cooling efficiency requires a very high field homogeneity of this solenoidal field. Especially critical are the transverse field components, because these limit the reduction of the transverse velocity. It is almost impossible to build a solenoid with transverse components smaller than 0.1% of the main longitudinal component. Thus, so-called correction windings are usually installed in order to compensate for the transverse components.
The solenoid is 2–3 m long and the variation of the transverse components relative to the main component should be less than 0.01%. We at GSI used a 3D probe head with three Hall probes (Fig. 1). The very critical mechanical requirement is that the orientation of the probe head in front of a 4-m long cantilevered beam must be stable within about 0.01 mrad. Otherwise the main component would lead to a wrong measurement of the transverse component. That demands very high accuracy and stability of the stage that is used. Using the mirror in front of the probe head we check this regularly by autocollimation, i.e. by measuring the angle between the incident and reflected light.

![3D probe head with three orthogonal Hall probes](image)

**Fig. 1** 3D probe head with three orthogonal Hall probes

### 2.3 Mapping of an Insertion Device

The field of an Insertion Device (ID) (wiggler or undulator) should not disturb the closed orbit of a stored beam, that means that an electron should neither be deflected nor transversely displaced. This leads to the requirements that the first and second field integral have to be zero, i.e.

\[
\int B(s) \, ds = 0 \quad \text{and} \quad \int \int B(s\hat{y}) \, ds\hat{y} \, ds = 0
\]

Thus these field integrals have to be measured. The measurement errors typically have to be less than 100 G·cm and 100 G·cm² respectively.

One method that measures both the local and the integral fields is the mapping of the ID point by point in the longitudinal direction. The integral field is numerically calculated by summation over the point measurements:

\[
\int B \cdot dz = \sum B_i \cdot \Delta z_i
\]

The integral reproducibility error is given by

\[
\Delta \int B \cdot dz = \Delta B^{rms} \cdot \Delta z \cdot \sqrt{N}
\]

where \( N \) is the number of points, \( \Delta z \) the step width and \( \Delta B^{rms} \) the rms error of every point measurement.

Let us consider only the field error due to erroneous positioning:
I use as an example the Advanced Light Source (ALS) Insertion Device [4, 5]: It is 5.5 m long, $\Delta z = \pm 0.22$ cm, $N = \pm 2500$, maximum field 0.9 T, wavelength $\lambda = \pm 5.0$ cm. The field oscillates rapidly with the wavelength $\lambda$ between -0.9 T and +0.9 T.

The error of the local field measurement due to a position uncertainty $\delta z$ is given by the formula:

$$dB = B_0 \cdot \frac{2\pi}{\lambda} \cdot \sin\left(2\pi \cdot \frac{z}{\lambda}\right) \cdot \delta z$$

and

$$\Delta B_{rms} = B_0 \cdot \frac{2\pi}{\sqrt{2}} \cdot \frac{\delta z}{\lambda}$$

With the position reproducibility $\delta z$ of $1 \mu$m we get $\Delta B_{rms} = \pm 0.8$ G and the contribution to the integral error by the position accuracy only is $10$ G$\cdot$cm.

2.4 Measurement of the integral field strength $|Bdl|$ with the stretched wire method

When a stretched wire is moved transversely through a magnetic field by a distance $dx$, the flux through the wire loop changes and a voltage $V$ is induced that can be integrated:

$$\int Bdl = \frac{1}{dx} \ast \int Vdt$$

One sees immediately that the relative error of the position has to be of the same order as the desired accuracy of the integral field. A relative integral field accuracy of $10^{-4}$ and a step width of 10 mm requires a precision of the position measurement of $1 \mu$m.

3. STRATEGIES FOR REDUCING THE INFLUENCE OF MECHANICAL ERRORS

3.1 Use the best and most suitable equipment you can get

For example:

- good spring-loaded bearings for the minimisation of transverse vibrations
- good encoders
- interferometer for position measurement
- DC motors instead of stepping motors to reduce torsional vibrations

Do the components also operate properly in a high field and low temperature environment?

3.2 Use the appropriate material

- no magnetic material
- no conducting material
- high torsional stiffness (for example for the shaft and coupling of the rotating coil, $100\,\mu$Nm/mrad)
- high bending stiffness for a long cantilevered beam (good modulus/weight ratio)
3.3 Measure the correct sensor position

The position or orientation of the sensor (Hall probe/coil etc.) should be measured as close to the sensor as possible. That means for example:

- place the encoder near the rotating coils
- place the mirror for autocollimation measurements close to the Hall probe head
- place the interferometer reflector correctly for accurate position determination

3.4 Use intelligent design

The components may not be available on the market or are too expensive, the tolerances in machining are too tight.

- bucking coils compensate the main harmonic and thus reduce the influence of lateral and torsional vibrations on the measurement results
- damp the vibrations of the motor/gear by flexible coupling
- vertical orientation of the rotating coil/stretched wire avoids the influence of sag
- double search coils for gradient measurement are independent of the position accuracy

3.5 Use another method

A general strategy should be to double check the results by several methods!

- instead of mapping: stretched-wire method
- instead of stretched wire: flip- or search-coil method
- instead of mapping the three components of the field, measure the orientation of the field by a compass needle [6] (Fig. 2)
3.6 Measure the errors
If you cannot get the required precision of your equipment, then you have to measure the errors and correct for:
- systematic errors which can be measured off-line and corrected for later using calibration tables
- random errors which can be measured on-line
For example: Calibration of the 3D Mapper stage versus on-line measurement by autocollimation and/or interferometer.

3.7 Average the results
By appropriate averaging over different orientations mechanical errors will cancel out:
- offset of the inclinometer
- misalignment of a Hall probe can be checked by a rotation of 180°
- coil orientation (relative to gravity) can be found by end-to-end inversion of the coil

4. SPECIAL EQUIPMENT VERSUS UNIVERSAL EQUIPMENT
This is an often discussed question. Should we build for each measurement task an individual test stand for every magnet or should we have just one test stand, that can be used for each measurement task and for all magnets in the laboratory? Obviously both ways do not work.

The advantages of special equipment versus universal equipment are given in the following table:

<table>
<thead>
<tr>
<th>Special equipment</th>
<th>Universal equipment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well suited since it</td>
<td>Available in the laboratory or on the market thus saving</td>
</tr>
<tr>
<td>- saves measuring time</td>
<td>- money</td>
</tr>
<tr>
<td>- improves accuracy</td>
<td>- time</td>
</tr>
<tr>
<td>allows †better†comparison (reference magnet)</td>
<td>- man power for design and construction</td>
</tr>
</tbody>
</table>

Obviously we will build a well-suited specialised test stand for large series of magnets as well as for special magnets and magnets of extreme size.

In general each laboratory has some universal equipment. Whether you build a special test stand for a single magnet or not depends on your man power and the possible special demands the customer has.

5. MECHANICAL COMPONENTS
In Appendix 1 you will find a (incomplete) list of the most important mechanical components.

In the next two sections I will show you examples of mechanical equipment that have been built in the laboratories depending on the task that had to be done, i.e. on the properties of the magnet that had to be measured.

6. BENCHES FOR THE MEASUREMENT OF THE MECHANICAL PROPERTIES OF MAGNETS

During the production of normal-conducting magnets (especially in estrange magnets such as the LEP dipoles which are filled with concrete, and in laminated magnets) mechanical errors of the iron core are unavoidable. Thus one has to measure the

- gap height
- median plane position

as a function of the axial co-ordinate.

From these data one can find the average gap height, a possible gradient and the twist. One can then align the magnet correctly by bringing it up to the average mid-plane position and by removing the average roll and pitch.

I give two examples: One for the LEP dipoles [7] and one for the SLAC B-Factory High Energy Ring (HER) dipoles, both are C-type magnets.

The LEP bench is shown in Fig. 3. A carriage moves along the axial co-ordinate by rolling on the lower pole face of the gap and being pulled by a belt driven by a stepping motor. Five proximity detectors measure the position of the carriage in the gap and thus the gap height and the median plane. One four-quadrant photocell measures the position of the carriage relative to a laser beam and an inclinometer compares the tilt of the gap with the tilt of the reference surface on top of the core.

The bench used at SLAC is quite similar, but the proximity detectors are attached to an arm that is moved in the axial direction by a linear stepping motor. An inclinometer controls the orientation of the arm during the movement. A ěgarageï (an idealised gap) serves as a reference for mid-plane and gap height.

Fig. 3 Bench for measurement of the

Fig. 4 DESY HERA stretched-wire system
mechanical properties of the LEP dipoles

7. **BENCHES FOR THE MEASUREMENT OF THE MAGNETIC PROPERTIES OF MAGNETS**

7.1 **Standard accelerator magnets**

7.1.1 **Integral field strength**

We know several methods to measure the integral field strength directly, i.e. not by mapping the field point by point and then integrating the field numerically.

**Stretched-wire method**

Figure 4 shows schematically the bench at DESY for the HERA measurements [8]. A wire (usually tungsten or CuBe) is stretched between two points, fixed at one end and held by a tension spring at the other. The return wire closes the flux loop outside the yoke of the magnet. The wire can be moved by two high-precision x-y stages (resolution: 1µm), which operate in the master-slave mode. The induced voltage is integrated. The relative accuracy of the step width must be of the same order as that of the \[B_{l}\]. Thus the position is often measured by an interferometer. Figure 5 shows the Sincrotrone Trieste bench while measuring an Insertion Device [9]. The x-y stages are mounted on a common 3.2-m long granite base plate. Similar benches can be found at SLAC, ESRF, the former SSC [10], CEBAF [11] and FERMILAB [12].

![Fig. 5 Sincrotrone Trieste Elettra stretched-wire system](image1)

![Fig. 6 CERN CLIC Test Facility stretched-wire system](image2)
For the CTF (CLIC Test Facility) (Fig. 6) at CERN, a stretched-wire bench for quadrupoles of apertures as low as 10 mm has been developed [13]. In this case the magnet is moved, not the wire. The bench is vertical. The positions are controlled by new precise x-y capacitive sensors developed at CERN.

The stretched-wire bench can be used also for a *flip-coil* measurement. In this case the necessary positional accuracy of the stages is reduced. A rotation stage added to the x-y stages and driven by a step or servo motor flips the search coil. The principle is shown in Fig. 7.

If the stages are mounted on two separate base plates, you can omit one stage and replace the coil by a point probe at the end of a long probe arm so creating a very universal measuring device. Sincrotrone Trieste are presently changing the design of their bench in this way.

These methods work well, as long as the magnet is basically straight or at least the sagitta small compared to the aperture. For curved magnets flipping is not possible, so the *curved search coil* moves azimuthally out of the magnet, either on rollers or air cushions. Two possible schemes are shown in Fig. 8. At the top is the solution adopted at GSI for the SIS dipoles [14], the other solution was proposed for the ESRF dipoles. In both cases emphasis is put on the necessity to make a null measurement relative to a reference magnet. A different method was adopted at ANL for the curved C-type dipole magnets of the Advanced Photon Source (APS) [15]. A search coil is mounted on a flat board and moved in radial and vertical direction by several precision stages mounted on a common 4-m long base plate (accuracy: 0.01\text{"}mm) (Fig. 9).

GSI: SIS-Dipoles (2 coils)

![Fig. 7Schematic of a flip-coil device](image1)

![Fig. 8Azimuthally moving search-coil systems for curved dipoles](image2)
7.1.2 Field harmonics

*CERN LEP quadrupoles rotating coil*

The prototype of a standard bench for the rotating coil was developed at CERN by Louis Walckiers and co-workers for the LEP quadrupoles [16]. This type was later built by Danfysik and sold to different laboratories (ESRF, ANL, etc.). It is used to determine the field harmonics, the field axis and the field direction. A schematic side view is shown in Fig.10.

![Fig. 10 CERN LEP quadrupoles rotating coil](image)

The coil system consists of:

- radial coil and bucking coil (coil radius typically 100 mm)
- end coils/central field coil
- air bearings (lateral displacement less than 0.01†mm)
- DC motor
- absolute angular encoder (triggers the integrator)
The automatic magnet positioning and aligning system consists of:

- motors for vertical/horizontal translation and rotation around three axes. Air cushions are used for the horizontal translation.
- laser, position-sensitive light detector and electronic inclinometers (resolution: 0.01 mrad) for pre-alignment and for setting the fiducial marks

Overall specifications:

- relative accuracy of integrated main harmonic: $\pm 3 \times 10^{-4}$
- accuracy of a multipole component relative to the main component: $\pm 3 \times 10^{-4}$
- angular phase absolute accuracy: $\pm 0.2$ mrad
- lateral positioning accuracy of magnetic center with respect to the rotation axis: $\pm 0.03$ mm
- positioning accuracy of alignment targets with respect to coil axis: $\pm 0.03$ mm

The photograph of the bench shown in Fig. 11 was taken at ANL.

**Coil train for the CERN LHC dipoles**

For the measurement of the high magnetic field of the 15-m LHC dipoles two, long induction coil trains (12 coils connected mechanically together, rotating in ceramic pipes with an outer diameter 36 mm) are rotated from the outside of the magnets through shafts by a Twin Rotating Unit (TRU). From the specification of this TRU, I took the following scheme, that shows the principle layout with all the necessary components (Fig. 12).
1. Signal cable from CERN measuring coil
2. Connection flange
3. Rotating support for axle
4. Flexible cable connection
5. Reference surface
6. High precision, hollow shaft angle encoder
7. Level meter (gravity sensor)
8. Torque meter (Strain gauge)
9. Flexible coupling
10. Overload clutch (opens if torque > 0.1 Nm)
11. End switches and limiter of rotation
12. Drive unit

Fig. 12  Schematic view of the Twin Rotating Unit of the LHC dipoles bench

Main rotating parts:

1) Signal cable from CERN measuring coil
2) Flange connection to the shaft of the rotating coil train
3) Rotating support for axle (high quality ball bearings)
4) Flexible cable connection (three turns in both directions, flat-ribbon cable (50 twisted pairs))
5) Reference surface of axle (adjustable, to measure the orientation of the axle relative to gravity): ±0.05 mrad
6) High precision, hollow, shaft-angle encoder: accuracy ±0.05 mrad
7) Adjustable level meter on the top: measuring range: ±15 mrad, accuracy: ±0.05 mrad
8) Torque meter supervises the friction during rotation: max. 0.2 Nm
9) Flexible coupling for smooth rotation of the shaft and minimisation of vibrations
10) Overload clutch for protection (can open if torque > 0.1†Nm)
11) End switches and limiter of rotation
12) Drive unit (servo motor/reducer)

The torsion stiffness upstream of the encoder (in the coil direction) is specified as >100 Nm/mrad.

Mole for the low-field measurement of the CERN LHC magnets

Much work has been invested at BNL, SSC [17, 18] and at CERN [19] in the development of a rotating-coil device that can be moved through the small bore of a long superconducting magnet, the so-called émole. Let me present here the low-field mole (±500†Gauss) for LHC dipole measurements. Figures 13 and 14 are taken from the specification.
The mole (OD 46±mm) is pulled through the magnet by pulleys over two reels at either side of the magnet. The lateral position is determined by a PSD that monitors the laser beam. Internally it consists basically of the following components (all sealed in a stainless steel cylinder):

- level meter (limited range, accuracy: 0.05 mrad) and level motor
- coil motor (must not create an outside field >0.02 G)
- encoder (reproducibility of each trigger < 0.05 mrad, εzero of within ± 0.05 mrad)
- rotating harmonic coil (700 mm long)
- rotating PSD
- reference surfaces on the mole in order to check the εzeros of the PSD, level meter, and encoder

Figure 15 shows how the prototype mole looks in real life. Here the coil is driven by a long shaft, no motor is used. You see (from right to the left) the gravity sensors, the air brake, the ball roller, the encoder, the slip rings and the coil itself. Some moles are also equipped with Hall or NMR probes. Components for high-field moles have also been tested at CERN.
Rotating coils of extreme size

You can find both examples at LANL, one quadrupole with a bore as large as 1.25 m and as small as 10 mm. The corresponding coils are built on a large G-10 frame; For the small one a printed-circuit technique was used. They are presented in the proceedings of the CAS Magnetic measurement and alignment course 1992 in Montreux [20].

7.1.3 Field Mapping

Field maps are often required for:

- ray tracing
- the calculation of fringe field properties
- the determination of field overlap effects
- the calculation of special integrals (for example: \( \iiint B \cdot ds \cdot ds \))
- the construction of correction windings

One can design either special equipment, usually a little carriage carrying a field sensor that moves along a given curve (mostly the bending radius of the curved dipole magnet) or use co-ordinate measuring machines (CMM) with several degrees of freedom. The sensors measure either the main component or all field components. Some benches use several probes calibrated and oriented in situ.

Examples:

Standard CMM bench

Figure 16 shows a typical CMM bench with three orthogonal axes. All three tables run on linear ball bearings and are driven by stepping or servo motors via lead screws. This stage was used to map the fringe field region of the dipoles of the APS at ANL [21]. A similar bench exists at TRIUMF. With an additional auxiliary table, operated in the slave mode, they operate the bench as a flip-coil bench.
**Fig. 16**  Standard 3D mapping bench (ANL: APS dipoles)

**The Luge operated at the ALS in Berkeley**

The Luge was constructed as a high speed mapping device for the measurements of the Insertion Devices of the ALS [4, 5]. A custom-built stage moves axially through the gap of the 5-m long ID. It is supported on one side by a rail and on the other side by a pneumatic cylinder that provides the mechanism for axial translation (Fig. 17). The horizontal and vertical positions can be varied. The field components are measured by a Hall probe and point coils. As we have seen earlier in the error discussion, the accuracy of the axial position must be \(1\mu\text{m}\). Thus it was measured by a laser interferometer. The 6-m scan takes less than one minute.

**Fig. 17**  Luge for the mapping of the Berkeley ALS insertion devices

**The GSI Mapping device**

This device has six degrees of freedom, three translations and three rotations [22]. The probe head is mounted at the end of a long probe arm. In order to minimise oscillations of this arm, it was necessary to minimise any friction, so only air cushions and air-bearing slide shoes are used.
The complete system is shown in Figure 18. The three-dimensional measuring machine itself consists of a large granite table on a support sub-frame, two longitudinal slideways for the x-direction (horizontal), and one each for the y-direction (vertical) and the z-direction (orthogonal), respectively. The mechanical parts are driven on air-bearing shoes by motors using a continuous multiple steel band. The linear scan ranges are 2700 mm in the axial (x) and 1000 mm in the transverse (z) and vertical (y) direction. Resolution is 1µm and 20µrad. The rotation range around the x-axis is 360°, ±100° around the y-axis and +30° around the z-axis. The maximum speed is 27 mm/s. The whole machine can be moved either by a crane or air cushions.

![Figure 18 The GSI mapping device](image)

The probe arm consists of three long, carbon-fibre, epoxy cylinders that fit into each other and are bonded together to a maximum length of 4 m. Three Hall probes are mounted orthogonally to each other in a temperature-stabilised probe head. The orientation of the probes can be determined relative to the mirror in front by auto-collimation. The internal direction cosines of the probe are measured with a precision of 1 part in 10000.

The system was tested in three ways:

- Mechanically by electronic level meters: The maximum deviation of the flatness of the table was 8µm, the slideways were straight within 4µm. The three axes are orthogonal within 0.01 mrad.

- Optically by auto-collimation using two mirrors, one attached to the vertical sliding and the other in front of the probe: Moving with a speed of 20 mm/s along the x-axis, the maximum angular deviations were 0.01 and 0.08 mrad, respectively—the latter value includes the effect of the vibrations of the long probe arm.
- By magnetic measurements in a quadrupole: By averaging over 100 ms the results were improved by a factor of 4.

**CERN PS magnet mapping bench**

The combined-function magnet of the PS (dipole, quadrupole and sextupole!) is mapped in the following way [23] (Fig. 19): A trolley is driven pneumatically on 6-m long titanium rails along the longitudinal axis in steps of 20 mm defined by precise notches. The trolley carries a set of 15 Hall probes, aligned perpendicularly to the longitudinal axis. The lateral displacement is ±10 mm. It is used to calculate the gradient. Each time the trolley has moved axially the absolute lateral position is checked and corrected by a laser/PSD system to an accuracy of 0.01 mm.

![CERN PS magnet mapping bench](image)

**Fig. 19** CERN PS magnet mapping bench

**Measuring benches at LNS**

At the LNS (Laboratoire Nationale de Saturne) at Saclay, large benches with modular rails (curved and linear) up to 6-m long were developed. A carriage that is moved pneumatically along these rails can carry up to 100 Hall probes on a transverse 2-m long arm. Thus an area of 12 m\(^2\) can be measured in a short time. A special hall probe calibration bench allows 16 encased probes to be calibrated automatically in a reference magnet within 3 hours.

7.1.4 Field direction

The field direction is usually found by mechanical measurements (only for conventional magnets) as described above or by the rotating-coil or stretched-wire method already described.

7.1.5 Field axis
7.1.6 Material

Permeameter/Coercimeter

This topic is covered in the contribution of J. Billan to these proceedings [25].

Permanent magnet block measurement (LBL)

For the ALS Insertion Devices the magnetic moment of blocks of permanent magnets had to be measured [26]. The following objectives for a Helmholtz coil system were established:

- measure the three components of magnetic momentum to an accuracy of ± 0.1 %
- fast processing: 20 blocks per hour, more than 10000 altogether
- easy to use for an unskilled operator

This led to the decision to build an automated system. It is shown schematically in Fig. 20. The block holder in the centre of the Helmholtz coils attached to a long shaft is rotated 360° by a servomotor. The angle is measured by an encoder. A Fourier analysis of the voltage induced in the coils gives two components of the magnetic momentum. Then the block holder is flipped by 90° and rotated again by 360°. This measurement delivers the third momentum component.

![Diagram of Permanent-magnet block measurement (ALS Berkeley)](image)

Fig. 20 Permanent-magnet block measurement (ALS Berkeley)

7.2 Special magnets

7.2.1 Cyclotron magnet mapping
At PSI, Switzerland, and NAC, South Africa special benches for the mapping of cyclotrons have been built. These are described in the proceedings of the CAS êMagnetic measurement and alignmentí course 1992 in Montreux [20].

7.2.2 Detector solenoids

Again this topic is covered in the proceedings of the CAS school in Montreux [27] and in the contribution êDetector magnet measurementí of D. Newton to these proceedings [28].

7.2.3 Electron cooling device (including toroids)

The storage rings LEAR at CERN [29] and ESR at GSI were fitted with electron cooling devices. This required mapping of the whole device including the toroids, gun and collector solenoids. The solution was to bring the whole device into the horizontal plane. Then the whole field path was split in boxes that were measured from various directions (Fig. 21). The field components inside the boxes can be evaluated from measurements on the boundaries only [30].

![Fig. 21 Measuring boxes of the GSI ESR electron cooler.](image)

ACKNOWLEDGEMENTS

I thank all my colleagues in the laboratories in Europe and the US for helpful discussions and generously providing information and material.

* * *

REFERENCES


[28] D. Newton, "Detector magnet measurement", these proceedings.


## APPENDIX 1

### Mechanical/Optical Components

<table>
<thead>
<tr>
<th>Component</th>
<th>Specifications/Features</th>
<th>Companies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear-, Rotation</td>
<td>For example: Micro controle MT160 Resolution : 0.1µm Reproducibility: 0.1µm Accuracy: 1µm (for 100mm travel length) Max. speed: 40 mm/s</td>
<td>Micro Controle/Newport, France Klinger Scientific Corp., U.S.A. Spindle &amp; Hoyer, Germany Burleigh Instruments Inc., U.S.A. SKF, Germany</td>
</tr>
<tr>
<td>Table/Stages/ Optical Components</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear encoder</td>
<td>Best accuracy of the system: 0.1µm Temp. coeff.: 0/°K (substrate: glass ceramics) to 10 ppm/°K (glass)</td>
<td>Heidenhain, Germany</td>
</tr>
<tr>
<td>Rotary encoder (incremental)</td>
<td>Line counts up to 10000/revolution Interpolation by factor 4 and 5 Accuracy: 0.05 mrad</td>
<td>Heidenhain, Germany Gurley Precision Instruments, U.S.A. BEI, Encoder System Division, U.S.A. Litton Precision Products, Germany Baumer Electric AG, Switzerland CODECHAMP, France</td>
</tr>
<tr>
<td>Interferometer</td>
<td>Resolution: up to 1nm U.S.A.</td>
<td>Hewlett-Packard, U.S.A. Zygo Corp., U.S.A. Teletrac Inc., U.S.A. Spindle &amp; Hoyer, Germany</td>
</tr>
<tr>
<td>Capacitive sensor</td>
<td>Position accuracy (x-y): 1µm</td>
<td>FOGALE nanotech, France Capacitec, U.S.A.</td>
</tr>
<tr>
<td>Air-motor</td>
<td></td>
<td>Zo-Air Co., U.S.A. Physik Instrument, PI Ceramic GmbH, Germany Shinsei, Japan</td>
</tr>
<tr>
<td>Piezo-Electric motor</td>
<td></td>
<td>MicroPulse Systems, Inc, U.S.A.</td>
</tr>
<tr>
<td>Step, microstep motor</td>
<td>Microstep: up to 50000 steps/rev.</td>
<td>Parker Hannifin Corporation, Compumotor Division, U.S.A. Micro Controle/Newport, France Berger GmbH, Germany FENWICK, France</td>
</tr>
<tr>
<td>Servo-motor</td>
<td></td>
<td>Micro Controle/Newport, France Maxon motor Interelectric, Switzerland Maxon Precision Motors, Inc, U.S.A. Minimotor SA, Switzerland Portescap, Switzerland Portescap U.S., Inc., U.S.A.</td>
</tr>
<tr>
<td>Component</td>
<td>Specifications/Features</td>
<td>Companies</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>---------------------------------------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>Ball bearings (steel)</td>
<td>Centre-position stability: 1–2µm</td>
<td>FAG, Germany</td>
</tr>
<tr>
<td>Ball bearings (ceramic glass)</td>
<td>Centre-position stability: some µm</td>
<td>Wemh’ner&amp;Popp, Germany</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Koba Electronics, Japan</td>
</tr>
<tr>
<td>Slip rings</td>
<td></td>
<td>Schleifring, Germany</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Litton Precision Products, Germany</td>
</tr>
<tr>
<td>Inclinometer</td>
<td>Principle: capacitive position measurement of a pendulum</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Accuracy: 0.04 mrad</td>
<td>Wyler, Switzerland</td>
</tr>
<tr>
<td></td>
<td>Resolution: 0.002 mrad</td>
<td>Schaevitz, c/o Althen Messtechnik, Germany</td>
</tr>
<tr>
<td>Gravity sensor /Tilt sensor</td>
<td>Principle: resistance measurement of an electrolytic liquid in a vial</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Range: ± 1°, nearly-zero method</td>
<td>Spectron / G+G Technics AG, Switzerland</td>
</tr>
<tr>
<td></td>
<td>Resolution: 1 µrad</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Temp. coefficient &lt; 2µrad/°C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Temp. compensated</td>
<td></td>
</tr>
<tr>
<td>Laser for alignment</td>
<td>For mounting in a spindle</td>
<td>Hamar Laser Instruments, U.S.A.</td>
</tr>
<tr>
<td></td>
<td>0.01 mm beam centre position</td>
<td>OPTILAS, France</td>
</tr>
<tr>
<td></td>
<td>0.002 mm/hr/°/°C centering stability.</td>
<td>Gerhard Franck Optronic, Germany</td>
</tr>
<tr>
<td>Position sensitive detector/diode</td>
<td></td>
<td>United Detector Technology, U.S.A.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yamamatsu, Japan</td>
</tr>
<tr>
<td>Four-axis target</td>
<td>Two-axis centering: 0.01 mm</td>
<td>Hamar Laser Instruments</td>
</tr>
<tr>
<td></td>
<td>Squareness (pitch/yaw): 0.04mm/m</td>
<td></td>
</tr>
<tr>
<td>Auto-collimator</td>
<td>Resolution: 0.1µrad</td>
<td>Micro Controle/Newport, France</td>
</tr>
<tr>
<td></td>
<td>Range: ± 2µrad</td>
<td>Rank Taylor-Hobson, UK</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spindler &amp; Hoyer, Germany</td>
</tr>
<tr>
<td>Torquemeter /strain gauge</td>
<td></td>
<td>Kistler, Switzerland</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Applied Geomechanics Inc., U.S.A.</td>
</tr>
</tbody>
</table>
MAGNET MEASUREMENTS FOR SERIES PRODUCTION

P. Wanderer
Brookhaven National Laboratory, Upton, NY, USA

Abstract
This paper discusses magnetic measurements that are important in
qualifying magnets for use in accelerators. Experience gained in the
testing of superconducting magnets for the Tevatron, HERA, and
RHIC is summarised. The interactions with the accelerator physics
and magnet production groups are also reviewed.

1 INTRODUCTION
This paper brings together aspects of magnet measurements that are important for
certifying that magnets made as part of a production series are satisfactory for installation into
a particle accelerator. Series measurements check whether magnets are within accelerator
tolerances and provide fiducial information for the installation crew. For reasons of
economy, the number of measurements of series magnets is held to a minimum.

It was noted in a previous paper on this subject that, "The realisation of modern high
performance accelerator magnets requires a close collaboration between the measuring team,
the beam optics specialists, and the staff responsible for production" [1]. Certainly that has
been very much our experience in the Relativistic Heavy Ion Collider (RHIC) project at
Brookhaven National Laboratory. This note is intended to complement the discussion in Ref.
[1] by restating the most important aspects of series measurements and adding experience
from the production runs of the superconducting magnets for the Tevatron at Fermilab [2],
HERA at DESY [3], and RHIC [4].

This paper is organised by time period, with measurement activities divided into
planning, low rate production, high rate production, and postproduction periods. In each of
these four periods, there are characteristic discussions with the accelerator physics group and
with the magnet production group, as well as planning internal to the measurement group.

2 PLANNING FOR SERIES MEASUREMENTS

2.1 Interaction with the accelerator physics group

Generally, this discussion centres on magnetic field quality specifications developed
from magnets built for existing accelerators and from R&D magnets built for the new
machine. These values are then used for beam simulations. The measurement group must
reconcile the field quality specifications with measurement accuracy and with the budget.

Traditionally, magnet field tolerances are stated in terms of the mean and r.m.s.
variation of field quality variables (harmonics, integral field, field angle, etc.) which are
assumed to have Gaussian distributions. For RHIC, the accelerator physics group agreed to a
more general model, in which uncertainty was allowed in the mean itself [5]. This
uncertainty took into account possible systematic changes in the field quality variables which
could occur in the changeover from R&D to series production.
To minimise the cost of the correction system, the RHIC accelerator physics group requested that the field quality tolerances be those actually expected for the production run, rather than those which could be guaranteed. When specific field quality variables, or quench performance, have been worse than the “expected” values, their effects have been modelled in the accelerator and then corrected either on the production line or by appropriate placement in the lattice or both.

This is the time period to establish the links between the magnet test database (which is usually of modest size and not too hard to use) and the accelerator modelling database (more powerful and difficult to use). The links, of course, include people, as well as hardware and software.

2.2 Interaction with the magnet production group

2.2.1 Issues independent of measuring location

The smallest set of measurements needed to determine whether magnets meet the specifications is established at this time. The planned production measurements are a subset of the detailed measurements made on the R&D magnets. For example, R&D measurements made at many currents establish the excitation curve of field versus current in sufficient detail for good modelling of the accelerator. In principal, it would be sufficient to measure only the end points of the excitation curve to confirm that the production magnets meet the tolerances. Prudence, however, dictates that measurements be made at a few other currents so that at least the first round of questions which might be asked at a later time can be answered without remounting the magnet on the test stand. Another example is the effect of thermal cycles or quenching on harmonics. If these effects are significant, such studies will have to be incorporated into the testing of series magnets.

Discussions with the magnet production group establish an initial list of magnet components or features which are important for field quality but which may change during production. The list is shorter if thorough tests are made on components such as iron or superconductor. Mechanical and electrical measurements made during production, such as the weight of the yoke and the position of the magnet poles, also help insure magnet field quality. Ideally, if these measurements are within specifications, the magnet will be satisfactory. For unsatisfactory magnets, these data are used along with magnetic measurements to find the source of the problem.

Often the relative accuracy of measurements is much greater than the absolute accuracy. Accurate relative measurements are useful for the early identification of trends in magnet production.

All of the superconducting magnets installed in the Tevatron and HERA were cryogenically tested. However, because of the cost of this testing, subsequent accelerators have had to consider whether this is the best way to spend limited project resources. RHIC will be the first accelerator built with superconducting magnets which have not all been cryogenically tested. The initial “cold test fraction” must be established in this time period because of the long time needed to build cryogenic test facilities. For RHIC, the plan for the largest series of magnets (about 300 arc dipoles) was to cold test each of the initial 30 magnets but reduce the cold test fraction to 10% for the remainder.

Other topics needing discussion at this time include plans for magnetic measurements made near magnet production areas. Welding can raise havoc with sensitive electronics. Plans must accommodate the maximum rate. If magnetic measurements are part of the
production line (e.g., collared coils before yoking), the equipment may need an extra degree of reliability.

2.2.2 Magnetic measurements at a distance from the lab

HERA faced the greatest challenge here, with measurement stands supplied by HERA used at superconducting magnet production facilities in Holland, Germany, and Italy. For RHIC, BNL supplied equipment for measurements made at a factory which produced arc dipoles about 60 km from the lab.

The staff at the magnet factory may be unfamiliar with magnetic measuring equipment. It is worthwhile to spend some time at this stage explaining it so that there is greater confidence in the equipment and methods if a magnet fails a test later on. For RHIC, we began with a talk which gave an overview of the physics and the accelerator and stressed the importance of the magnets. This was followed by a talk which introduced magnet harmonics and stressed the relationship between the allowed and unallowed harmonics and various symmetries in the magnet components. Examples of the sensitivity of various harmonics to critical dimensions, such as the amount of skew quadrupole produced by small coil size differences, were given. (Eventually, the staff at the magnet factory asked for a table of all our available studies of harmonics versus magnet component sizes.) To build confidence in the measurements, the talk included examples of unwanted harmonics found initially with magnetic measurements and later confirmed with mechanical data.

Short instruction books (5-10 pages) for using the magnet measuring system were written by both HERA and BNL. The chief technician from the magnet factory assigned to magnetic measurements received hands-on instruction at the lab. No vendor had a professional staff member assigned to magnet production who was an expert in magnetic measurements. Instead, the labs provided software that evaluated the measurements and the magnet while the magnet was still on the test stand.

The magnet measuring stands were, of course, built to be highly reliable. In addition, numerous diagnostics which could be run by the factory’s technicians were provided. Extra wires were added to the HERA cables so that the computer could check whether the cables had been correctly hooked up. The measuring systems could be run remotely from the labs. Both labs had enough confidence in their systems that they did not supply complete backup systems. HERA provided some extra electronics modules. BNL provided a spare rotating coil and drive. Overall, the measuring systems worked well, with few repairs needed.

Under most circumstances it is advantageous to supply reference magnets for the vendor. HERA supplied a set of wires stretched on a cylinder. This “magnet” was used as necessary to establish that the measuring system hadn’t changed. BNL provided a dipole made of permanent magnets so that the calibration of the gravity sensors could be checked before and after each measurement.

A couple of miscellaneous items are worth mentioning. HERA found it valuable to have a lab representative who was knowledgeable about magnets and who also was fluent in the languages of the vendors. At BNL, the vendor’s measuring technicians called BNL staff directly so that problems could be resolved as quickly as possible. When HERA magnets failed the built-in check, the computer printed the unmistakable message, “This magnet not accepted by DESY.”

2.3 Measurement group activities

Series measurements require many of the same reference magnets that are necessary for making accurate measurements of any magnet. These include a large dipole for determination
of dipole field angles and (together with NMR) field strength. Measuring coils with bucking windings also need a quadrupole field for calibration. Arrays of inexpensive permanent magnets can be used to create high-order multipoles to check that the measuring system gives the correct polarities. It may be desirable to construct an array of precisely-placed conductors (without iron) to generate fields of high multipolarity. At BNL it was found more useful to make one of the spare sextupole cold masses a reference magnet. The sextupole was periodically remeasured to check for changes in the measuring system (such as a slow drift in the gravity sensor system) that might have gone otherwise unnoticed.

Frequent rechecking of the HERA and BNL gravity sensor systems proved to be necessary. HERA built a small dipole from permanent magnets and added a level so that the gravity sensor system (which was coupled to the measuring coil) could be checked before and after each magnet was measured. A similar system was built at BNL.

Series measurements usually require multiple copies of the same measuring coil design. After the coils are calibrated against the reference magnets in the usual way, it may be useful to make final adjustments so that one coil becomes the reference coil, and calibration constants of the others are (slightly) adjusted so that all coils give the same result when measuring the same magnet.

A technique used to enhance the sensitivity of measurements is to run a “test” magnet and a “reference” magnet in series, install nominally identical coils in each of them, and measure the difference signal. This is commonly done when testing resistive magnets but has not been used for superconducting magnets because of the expense of cryogenic testing. A recent example (Ref. [6]) is a system that used flat coils moved horizontally in the magnet gaps, and a fit to the x-dependence of the field strength difference, to measure the dipole, quadrupole, and sextupole terms.

The initial plan for the minimal set of production measurements is made in this period. The list of data for the database of magnetic measurements is developed from the set of production measurements. Some examples from RHIC arc dipole testing illustrate the issues. Integral harmonics were measured by moving a 1 m-long rotating coil stepwise through ten axial positions. The magnets were measured twice at the vendor’s, at room temperature (“warm”). Although it added about an hour to the measurement time, the second axial scan was particularly useful in resolving questions that arose later when measurements at one of the ten positions seemed out of line. The 10-position integral harmonic measurement was repeated at BNL for similar reasons. Also at BNL, the integral dipole field was measured with a single long, nonrotating coil by ramping the magnet. (The integral dipole field was also calculated by summing data from the 1 m rotating coil. Although not the definitive measurement of the integral field, it was quite useful for diagnostic work.)

After the magnets were mounted on the test stand, the integral harmonics were again measured “warm”. This measurement, together with the later “cold” measurements, served to check that the dipole field angle didn’t change as a result of the cooldown of the magnet. When cold, the magnets were quenched until four consecutive quenches had about the same current. Integral harmonic measurements were made at only three currents (injection, an intermediate field, and full field) because they were time consuming (ten positions using a 1 m rotating coil system, as for the measurements at the factory). With the axial position of the 1 m coil fixed, 30 measurements were made as the magnet was ramped stepwise from below injection to above operating current and back down. Finally, the integral dipole field was measured with the long, nonrotating coil at frequent intervals as the magnet was ramped continuously from injection to operating current. Another measurement was made on the down ramp.
An important part of determining the minimal set of measurements is to establish the magnet current cycles needed to have reproducible measurements. Resistive magnets are cycled between high and low currents a few times and then usually measured on the downward side of the hysteresis loop. Superconducting magnets need to be cycled at least once before measurements, but perhaps more often for nested coils such as correctors. Depending on the interstrand coupling resistance, it may be necessary to quench the magnets before cycling them. Dynamic effects associated with magnet ramping may require separate equipment for data-taking at high rates.

3 LOW RATE PRODUCTION MEASUREMENTS

3.1 Interaction with the accelerator physics group

The magnets made in this time period may well be the first ones which are designed to meet the tolerances in all respects. Also, this may be the first group with enough identical magnets to establish the random variations. Thus, comparison with magnet tolerances is much more detailed than before.

The initial magnet acceptance meetings are held at this time. A fair amount of work is involved in the acceptance process, so a division of labour is appropriate. For RHIC, all magnet test data were stored in a PC spreadsheet with just the approved runs moving to the database, so we were able to take advantage of the flexible plotting capabilities of the spreadsheet and provide all the plots for the acceptance meetings. The accelerator physics group's relational database had limited plotting capability, so it performed the task of checking all data for the magnets under review and flagging values beyond set tolerances. The list of magnets up for review was available by early Tuesday. An initial review by the accelerator physics group was held on Wednesdays. Sometimes, this meeting asked for extra plots to be made. The acceptance meetings were held on Fridays.

Reviews of these initial magnets lead to the first negotiations on acceptance. It is not possible to model accelerator performance for all possible magnet errors which exceed the tolerances, but specific errors can often be checked in a few days. It is easier to accommodate a small number of magnets which are beyond tolerances than a larger number, so informal review of measurement data on a daily basis (occasionally followed by consultation with the accelerator physics group) is of great value. Examples of this are given in Sec. 4.1.

3.2 Interaction with the magnet production group

A major focus is the comparison of the magnetic field measurements with the "expected" values. Differences between design and measured values need to be understood in terms of details of the production process. A critical question during this period and during high rate production is deciding when changes in magnet performance are trends instead of statistical fluctuations. At this early stage there are apt to be numerous small changes in the magnets, complicating the decision.

This time period is apt to be the magnet manufacturer's first experience with correlating magnet harmonics with construction details. For example, a shim change early in the production of RHIC arc quadrupole cold masses was incorrectly installed. The first several quads produced after the change had some large harmonics. We guessed that the large harmonics may have been produced by incorrectly installed shims and confirmed the guess with a calculation. Inspection of the cold masses revealed that this was the case. Because
only a few magnets were involved, it was possible to accommodate them in the lattice. Strictly speaking, we could have required the manufacturer to rebuild the magnets.

At this time it may be discovered that additional measurements or measurement techniques beyond the minimal planned set are needed because of unanticipated problems with the magnet production. This happened in three instances in RHIC. Problems with accuracy were found when a “CQS” cold mass was formed by welding together a corrector, a quadrupole, and a sextupole. This meant that magnetic measurements were needed to help improve the welding process and to confirm the correct alignment. Plans to develop room temperature measurements of the field centres with a colloidal cell and with a “survey antenna” were quickly implemented [7-9]. In the second instance, quench testing of corrector cold masses needed to be increased due to problems discovered during low rate production, so an additional vertical dewar was brought into service. In the third case, quench problems with the sextupoles during high rate production caused the cold test requirements to be increased. These examples are discussed in more detail in Sec. 4.3.1.

3.3 Magnet testing

Most or all of the detailed testing from the R&D period carries over into this period. In particular, all of the superconducting magnets are tested at 4.3 K. This testing is necessary to establish the correlation between “warm” and “cold” field quality measurements. Even for programs which will test all superconducting magnets cold, the correlation is important because it allows warm measurements to be used for earlier feedback to the production line. Studies of the effect of quenching or thermal cycles on field quality will also need to be continued.

Warm-cold correlation data for the normal sextupole of the RHIC arc dipoles is shown in Fig. 1. (The harmonics follow the definition given by A. Jain [10], with a reference radius of 2.5 cm. However, “U.S.” notation is used here, with b, denoting the normal quadrupole, and so on). The line is drawn at precisely 45°. At the top right corner of the plot is the mean warm-cold difference, followed by the uncertainty in the mean. Warm-cold correlation data for the integral transfer function $\int B \, dl / l$ are given for the three selected magnet currents in Fig. 2. Each line is the best fit to the data. At 5 kA, the maximum operating current, there are two sets of data due to a planned change in the total yoke weight made during the production run.

![Diagram](image)

Fig. 1 Warm-cold correlation of the normal sextupole for RHIC arc dipoles
4 HIGH RATE PRODUCTION

4.1 Interaction with the accelerator physics group

The challenge here is to determine whether magnets which have problems with one or more field quality specifications, but are otherwise acceptable, can be used in the accelerator. A key tool is sorting the magnets by position in the lattice so that any effects of the problems are minimised. A complicated sorting algorithm was employed for Tevatron dipoles [11]. At HERA, the arc dipoles from the Italian and the German manufacturers differed in effective length by 1 cm. Also, the time-dependent characteristics of the superconductor were not identical because there were two sources for the cable. The effects of the differences were minimised by installing the magnets in alternate sectors of the accelerator [12].

Fig. 2 Warm-cold correlation of the integral transfer function for RHIC arc dipoles
Several methods of sorting were used for RHIC. Dipoles with integral fields higher than the mean were installed next to those with integral fields lower than the mean to reduce the overall load on the dipole correctors [13]. A number of the sextupoles had quench currents that were just 30% higher than the operating current instead of the usual 100% expected from magnets operating at the limit of the conductor. These were installed at locations where the operating currents were expected to be lower than average. The worst magnets of each type were designated spares.

Sorting for RHIC insertion region magnets made use of the fact that two of the six regions had higher luminosity requirements than the other four. The insertion magnets with the best field quality were assigned to one of the two “golden” regions. This, however, required that test data be available on a significant fraction of the magnets for the acceptance meeting so that locations could be assigned. Because the final stages of insertion magnet production were location-specific, the acceptance meeting itself became (reluctantly) a part of the production schedule.

4.2 Interaction with the magnet production group

4.2.1 Interaction with magnet manufacturer

An important feature of this period was regular interaction with the manufacturer. Although the complete database of approved test results was available to the manufacturer, the information received the most scrutiny at the weekly magnet acceptance meeting. To assure that the dipole manufacturer was aware of the current issues affecting acceptance, we faxed the plots shown at the meeting (marked to indicate the specific magnets reviewed and any problem data) to the manufacturer. These were then reviewed by the manufacturer’s production staff. Once a month, an update of the same information was presented at a joint magnet production meeting of the staffs from BNL and the manufacturer. Also, proposed changes in magnet production were often informally reviewed by all the relevant groups at BNL. This review provided another opportunity to make the manufacturer aware of the connection between mechanical specifications and magnet field quality.

4.2.2 Changes during magnet production

In series production, magnetic measurements are often used to help understand or confirm changes in a series. Examples from RHIC production illustrate the types of changes that can occur. The changes can be divided into three types: planned changes, unplanned changes to a series, and unplanned changes to individual magnets.

Planned changes generally affect a series of magnets. After the first 30 RHIC arc dipoles had been tested, a coil cross section change was made to reduce the allowed harmonics. Also, the manufacturer was directed to place the heavier half of a yoke on the bottom to reduce the skew quadrupole at 5 kA. A change that was planned, but not at a fixed time, was switching to iron laminations punched from a new die, after the previous die broke. A one-of-a-kind change was to relax the limit on coil size differences for the last dipole, to minimise the number of coils made.

There were several unplanned changes to the series. In one instance, the radial thickness of a plastic spacer used in the arc dipoles decreased, causing the average conductor radius to increase. This was first noticed in the sharp drop in $B/I$, the transfer function, at magnet 89 (Fig. 3). The average value of the transfer function for magnets 89 through 100 is significantly lower than the average for the previous ten magnets. Because it took about a week for magnets to be completed and measured after the spacers were installed, and because the manufacturer of the spacers claimed initially that there was no problem, it took several
weeks for the problem to be fixed. Magnets 101 through 112 were made with spacers each of which had been measured. By this time the problem was identified. However, throwing out the defective spacers would have stopped the production line. It was at this point that the idea to install dipoles in pairs according to transfer function (one high, one low) was conceived and implemented [12].

![Graph](image)

**Fig. 3** Trend plot of the integral arc dipole transfer function at room temperature

It should be noted that, when any such problem is first detected, the magnet test group has to be able to defend its measurements. This is much easier if two independent methods can be used. In this case, the data from the nonrotating integral coil supported the data from the 1-m probe.

At a later time a change in the saturation of the skew quadrupole was traced to a calibration error in the scale used for weighing yokes. Determining the cause of this change was made more difficult by the fact that the spacers and yoke subsections were not used in the same order they were received or assembled. This was a necessary production economy.

Another example of an unplanned change to a series was a decrease in the quench performance of the RHIC sextupoles starting about serial number 250 (Fig. 4). (The first production sextupole is serial number 101.) The figure plots the number of quenches required to reach 95% of the conductor limit against serial number. The plot reflects the nominal 10%
cold test fraction used until that point. The magnet-to-magnet fluctuation in the number of quenches is large and it was difficult to demonstrate unequivocally to the manufacturer that there was a problem until the magnets required more than the contract limit of ten quenches to reach 95% of the conductor limit (serial number 283), or did not reach it at all. My colleague A. Ghosh switched quantities, to the minimum quench current, and smoothed the data through the use of an eight-magnet moving average (Fig. 5). His plot indicates a problem much earlier (soon after serial number 230) than the plot of Fig. 4. The quench problem was traced to a reduction in the amount of epoxy used in the winding of the magnet coils. Coils were not installed in magnets in the order in which they were made, so the deterioration in performance was not abrupt. Fortunately, it was possible to identify all of the suspect coils and magnets. With the number known, the accelerator physics group was able to find places where the operating current was expected to be relatively low so that all could be used in the accelerator.

![Fig. 4 Trend plot of the number of quenches required to reach 95% of the superconductor limit](image1)

![Fig. 5 Eight-magnet moving average of the minimum quench current](image2)

Some unplanned changes affect only individual magnets. The most important example for RHIC occurred when, due to the use of the wrong type of plastic spacer, the conductor was not restrained at the magnet pole over two 15-cm lengths separated by 61 cm. In these two regions, the cable nearest the pole was about 1 mm further from the midplane than it should have been, producing a large skew quadrupole in this region because of the top-bottom asymmetry of the coil. The magnet-to-magnet variation of the integral skew quadrupole was large enough (Fig. 6) that the effect of the locally-large skew quadrupole was not visible in a trend plot of the integral data. It was readily visible in plots of the axial variation of the skew quadrupole measured at 1 m intervals in the magnet straight-section (Fig. 7). The error was not seen in the normal quadrupole because the left-right symmetry is unaffected. (Note the difference in vertical scale for the two plots of Fig. 7.) The error also appeared in the expected higher order harmonics.

Based on our best guess as to the source of the problem, a computer model of the fault was created. Since the plastic spacers were only 15 cm long, it was important to measure the axial variation in steps of this size or smaller to confirm the model. This was done in two ways: Hall probes for the field strength and differential measurements made in 15-cm steps with the 1-m rotating coil. The measurements were in good agreement with calculations based on the model. Due to the cost of each magnet, the manufacturer elected to disassemble the completed magnet enough to inspect the faulty area. The inspection confirmed that the wrong spacers had been used. A similar error occurred a second time, 97 magnets later (Fig. 7), before sufficiently tight inspection procedures were put into place.
4.3 Magnet testing during high-rate production

The most important issues here were adapting the “10% cold test” rule to the individual types of magnets and optimising the technician effort. This plan (described in Sec. 2.2.1) was followed exactly for the arc dipoles. Each of the first thirty dipoles was cold-tested and all reached the current-carrying capacity of the superconductor. (One took more than a few quenches to do so. It was subjected to two additional thermal cycles and then reached the conductor limit.) No cold leaks or hipot failures were found. The correlation of warm and cold field quality measurements was better than the tolerances. Based on these good results, only 10% of the remaining 270 magnets were cold tested. (Thus, in all, about 20% of the dipoles were tested cold.)

For reasons of production convenience, usually every tenth magnet was cold tested. However, in several instances, magnets with unusual repairs were selected for cold testing. No problems were found in any of the remaining arc dipoles during cold test. (The magnets with the wrong spacers, discussed above, were not formally shipped to BNL.) The “10% cold test” rule was also followed for the arc quadrupole cold masses, which were made by the same manufacturer.

Evaluation of this strategy must wait until the cooldown of all of RHIC, scheduled for early 1999. Thus far, one sixth of RHIC has been successfully cooled down [4]. However, many of the magnets in this portion of RHIC had been individually cold tested.
The arc sextupole series had training problems due to low epoxy content in the coils, as was noted above. All magnets with “low epoxy” coils were cold tested. Once the problem was solved, we returned to 10% cold testing. However, about 10% of the individual coils were quench tested at BNL as soon as they were manufactured so that we could carefully monitor the quality of the coil production. (During the R&D program it was found that the quench performance of the magnets closely followed that of the individual coils.)

The four-layer arc corrector cold masses, which were produced at BNL, were made by a process which had been found to occasionally damage the conductor, which was a single wire. It was not possible to efficiently check for wire damage during production, so all of these magnets were cold tested. This testing, plus the added sextupole testing, was more than we had made provision for in our plans. Fortunately, we were able to adapt a short test dewar to cool down and quench test two magnets at a time. One in ten of the correctors was tested in the standard setup, where both field quality and quench data were collected.

The above units were welded into a single CQS cold mass and then put into a cryostat. The first thirty CQS units were tested cold with success. Since the individual magnetic elements were previously tested satisfactorily (with 10% to 100% cold tested, as noted above), it was felt that little would be gained by testing the CQS units as a whole, so testing was reduced below the 10% level for the rest of the production series.

Several things were done to optimise the magnet test technicians’ work during high rate production. Data-taking was automated as much as possible. The software permitted a run to be restarted from an intermediate current, and the data for a single current to be overwritten, to handle the occasional interruption of the control system. It was particularly helpful to let the computer control the bucking and the power supply.

The data-acquisition program generated plots of the harmonic coil output and the bucked data. It also displayed a single number (the r.m.s. variation of the speed of coil rotation) that was useful in characterising the uniformity of rotation, which is a critical element in the performance of the measuring system. The technicians were quite good at noticing changes in this information and knowing when to contact the physics staff.

A separate program was provided so that the test technicians could check the data for the most common errors before it was transferred to the network computer for final analysis and review.

At its peak, magnet testing was scheduled two shifts a day, five days a week. (Fortunately, we were able to avoid working on the midnight shift.) A physicist was always available for consultation. Most quench traces were reviewed by a physicist.

Technical milestones were celebrated with bagels and coffee several times a year.

4.4 Short production runs

In the RHIC insertion regions, just 24 of each type of magnet are needed, plus two spares. This is only about a tenth of the number needed for the arc regions, so a different approach was needed in order to minimise the effort needed for magnet development. We took advantage of the fact that the most stringent field quality requirements only applied to magnets in the two “golden” crossing regions, as discussed in Sec. 4.1.

We also made extensive use of variable-thickness shims to reduce the effects of production tolerances on field quality. Shims were used to improve the harmonics in the insertion quadrupoles [14]. They were also used to improve the alignment of the axes and field angles of the four-layer insertion correctors. This program increased the number of warm measurements of these types of magnets by a factor of two to three. These
measurements were part of the magnet production line, and hard work was necessary to avoid delays.

4.5 Resistive versus superconducting magnets

Both types of magnets need to be cycled to high current before magnetic measurements can take place. Often, the iron in superconducting magnets is taken well into saturation, but this is uncommon in resistive magnets.

Usually, the poles of resistive magnets can be used to align measuring equipment [15-17]. This provides a powerful consistency check for the measurements. Typically, fewer measurements are needed to characterise the field of a resistive magnet. Recent examples of series production of resistive magnets are those for the APS at Argonne National Laboratory [6], the Main Injector at Fermilab [15, 16], and, at SLAC, the A-line [17] and the B-factory [18].

There are several features of superconducting magnets that help reduce the measurement task. Generally, the centre of the iron yoke closely coincides with the field centre. If this is borne out by careful measurements during the R&D program, fiducial notches on the yoke perimeter can play the same role as pole pieces in resistive magnets in locating the measuring coil and providing a cross check of the integrity of the measurement and the magnet. The fiducial notches can also be used by the survey crew. Warm and cold measurements of the geometric multipoles (those due only to conductor location) are well correlated. Nonetheless, extra measurements will be needed to characterise the magnetisation of the superconductor and the dynamic behaviour of the harmonics.

5 AFTER COMPLETION OF MAGNET PRODUCTION

The experience with the Tevatron and HERA indicates that a modest capability for testing superconducting magnets needs to be maintained after production is complete. Post-production studies of the time and spatial dependence of the field quality have been necessary at both labs.

Accelerator modelling efforts require that the magnet measurement data be available as long as the accelerator runs. It is also possible that there will be some reanalysis of these measurements.

REFERENCES


DETECTOR MAGNET MEASUREMENTS

D. Newton
Dept. of Physics and Chemistry, University of Lancaster, UK

Abstract
The many differences between detector magnets and accelerator magnets are emphasized, especially the large magnetic volumes and stored energies. External measuring machines are described, as used for mapping the rather open magnets of the 1960s and 1970s. For the almost closed magnets of modern detectors an internal machine is described. The required precision is discussed. Methods for checking the field map are presented, with actual results from one detector.

1. DETECTOR MAGNETS AND ACCELERATOR MAGNETS
The magnetic measurements of detector magnets and accelerator magnets have a few similarities and many differences. The similarities are that the order of magnitude of the fields have up to now been similar, usually between 1 and 2 Tesla, so that the traditional techniques of search coils, Hall probes, NMR probes are suitable for both. Iron yokes have usually been involved, with consequent saturation effects at full field, and the well-known hysteresis effects.

The differences stem from the use of the magnets; in accelerator magnets particles of well-defined momentum pass all the way through, but in detector magnets particles of unknown momentum ranging over a factor of 1000 (e.g. from 0.1 GeV/c to 100 GeV/c, and either sign of charge) may have any trajectory, as shown in Fig. 1.

Fig. 1 Tracks of various momenta in a typical detector

The trajectory is detected at sampling points, or continuously in a bubble chamber, and the momentum calculated. If the field were uniform and the trajectory a helix, this would be simple. Over large volumes the attainable fields are not uniform, hence the need for a detailed field map. Then the measured trajectory, even though not a helix, allows the momentum to be
determined. The procedure for doing so is far from obvious, and much effort has been devoted to various software techniques[1].

Each detector magnet is unique; the degree of non-uniformity has no precedent. Twenty or thirty years ago this could bring some nasty surprises, but nowadays computer design codes for magnets are very good. Perhaps most important, there is only one opportunity to make the magnetic measurements. Between the engineers assembling and commissioning the magnet, and the physicists filling it with detectors, there will be a few days allowed for the magnetic measurements. Clearly a suitable measuring machine must be ready for this opportunity and must work without breakdown throughout a carefully prepared schedule.

2. STORED ENERGIES AND POLE-TIP FORCES

It is worth drawing attention here to some of the consequences of the very large volumes of magnetic field, in essentially air, of these detector magnets. This contrasts sharply with accelerator magnets. It also shows a spectacular, perhaps frightening, trend over the past few decades and towards the future.

The stored energy density is

\[
\frac{1}{2} \mathbf{B} \cdot \mathbf{H} = \frac{B^2}{2\mu\mu_0} 
\]

\[
= 0.4 \times 10^6 B^2 \quad \text{for air}
\]

Joule per m³, and negligible for the iron because the permeability \( \mu \) has a value of several thousand [2].

At a typical field of \( B = 1.6 \) T this gives a stored energy of one megajoule (MJ) per m³. In bubble chambers of the 1960s the magnetic volumes were under one m³. In the 1970s bubble chambers, and the increasingly popular counter experiments, had magnetic volumes of several (1 to 3) m³. In the 1980s detectors at fixed target machines became less popular than those at storage rings ISR, \( \bar{p}p \), and Petra, and some of these had magnetic volumes of order 10 m³. Now in the 1990s we have at LEP and HERA some huge solenoid detectors like Aleph, Delphi, H1 which have magnetic volumes of order 100 m³, and the giant L3 with magnetic volume of order 1000 m³. The stored energies in these magnetic fields is of order 100 MJ.

It should be noted that a stored energy density of one MJ per m³ in air but negligible in iron implies a pressure of one million Newtons per m² pushing the pole tips together (at the typical field of \( B = 1.6 \) T, and varying as \( B^2 \) for other fields). This is a pressure of ten atmospheres, or ten tons per square foot! It so happens for the magnet shapes used over the past 30 years that the fields and areas of pole tips have been such as to result in a force pushing the pole-tips together approximately equal to the weight of the magnet in each case. Thus the forces have grown from tens of tons, through hundreds of tons, and are now thousands of tons. For example in the H1 magnet the distance between the pole tips decreases by several millimeters when the magnet is turned on, an important fact when mounting a measuring machine for field mapping, or when mounting precision tracking devices for physics.

3. DETECTOR MAGNETS AND EXTERNAL MEASURING MACHINES
Detectors of the 1960s were chiefly bubble chambers or counter experiments, the latter consisting of scintillation counters on their own or triggering optical spark chambers, and both involving magnets for momentum determination.

A typical bubble chamber of the 1960s was the 150 cm x 50 cm x 50 cm hydrogen chamber used at CERN. The beam traversed the 150 cm length of the chamber, passing between the two coils and through holes in the surrounding iron yoke. The cameras viewed along the axis of the coils and thus along the magnetic field direction. Their depth of field was the full 50 cm depth of the liquid hydrogen. Thus the volume over which tracks could be detected, and over which the magnetic field map was required, was almost all of 150 cm x 50 cm x 50 cm. The field was approximately 1.3 T and was mapped with Hall probes, done of course before the liquid hydrogen chamber was introduced into the magnet.

In the 1960s and 1970s the most common type of detector magnet for counter experiments at the fixed target accelerators was like that shown schematically in Fig. 2. The pole-tips were rectangular and horizontal, typically 0.75 m apart, producing a field of between 1 and 2 T over a volume of 1 to 3 m³. A detailed description of such a magnet is given by Barber et al [3].

![Fig. 2 Front view of a typical detector magnet for counter experiments](image1)

![Fig. 3 Local distortion of the field produced by a half-cylinder hump on the lower pole-tip](image2)

The most common type of measuring machine had a set of small probes, search coils or Hall plates, on a long aluminium arm supported from a modified lathe bed. There would often be three mutually perpendicular probes to measure the three components of the field. With the coordinate system shown in Fig. 2 the long arm would be parallel to the z axis. The lathe bed would stand outside the magnet with the greatest extent of travel, the length of the lathe bed, parallel to the z axis. One stepping motor would allow a long scan in z, typically in steps of 2 cm. Readout was from the probes while stationary between steps. Other stepping motors on the x and y motions of the lathe bed would allow a full three dimensional scan, often taking several hours. An NMR measurement at the magnet centre would give absolute normalization.

Frequently the travel in z was insufficient to go right through the magnet. In such cases the measuring machine had to be taken round to the other side of the magnet, with resulting problems in relating the two scans to each other. A method for dealing with this was as follows. Before moving the machine an extra scan would be done from the first side. For this extra scan a small iron strip would be fixed onto the lower pole-tip. The shape of the iron strip was not important. It produced a local disturbance to the field shape just above the lower...
pole tip. Figure 3 shows the effect of a half-cylinder strip, the lines in the diagram being the equipotentials and field lines.

Clearly the scan over the strip would show a symmetric peak in the main component of the field and an anti-symmetric curve for the minor component, thus recording accurately the position of the strip in the coordinate system of the measuring machine. Then the machine would be moved to the other side of the magnet, and another scan made over the strip. Finally the strip would be removed and the proper magnetic measurements continued.

Several detector magnets had circular pole tips protruding through circular coils. The Omega Spectrometer at CERN is typical in this respect (but untypical for the 1970s in having superconducting coils). The natural coordinate system would seem to be the axis of symmetry as $z$ axis, radius $r$ from the $z$ axis, and azimuthal angle $\varphi$. One might expect a measuring machine to scan in these $z$ and $r$ and $\varphi$ coordinates. I cannot remember any such machine. The machines actually used were of the type already described, small probes on a long arm driven by a stepping motor system standing outside the magnet and giving scans on an $x,y,z$ coordinate system. The large size of the Omega Spectrometer meant that the complete field map was a combination of two field maps made with the measuring machine on opposite sides of the magnet, and the technique described above used for relating the two field maps.

4. **Solenoid Magnets and an Internal Measuring Machine**

The rising popularity of storage ring accelerators in the 1980s was accompanied by the development of the "4• detector". This is one in which the interaction point of the colliding beams is more or less at the centre of the detector. Secondary particles from the interaction are detected whatever their direction or momentum or charge, or at least very nearly so. Such detectors are almost completely closed; the only access when operating is through the small holes left for the accelerator beams. Solenoidal magnetic fields have become most popular, usually created by a cylindrical superconducting coil producing a field between 1 and 2 T. Figure 4 shows the typical arrangement with the axis of symmetry horizontal to coincide with the axis of the colliding accelerator beams. Surrounding the coil is the barrel part of the iron return yoke. At the ends are the end-caps which complete the iron return yoke and act as pole-tips. Each end-cap has a central hole for the accelerator beams.

![Fig. 4 Typical solenoid detector showing axis, coil, barrel and end-cap iron yoke](image-url)
4.1 Internal measuring machine

Clearly for such a detector the magnetic field map must be made by a machine inside the almost closed volume. The small holes in the end-caps are sufficient for feeding in mechanical shafts from external stepping motors, and bringing out the readout signals, but the main movement of the arms carrying the probes must be done inside.

Towards the end of the 1980s three such detectors planned (in Europe) were Aleph and Delphi for LEP and H1 for HERA. It happened that the dimensions of the magnetic region were similar. CERN arranged for the construction at MPI, Munich, of the measuring machine shown in Fig 5. This machine was used for magnetic measurements of these three detectors. It was also used for mapping the central part of the L3 magnetic volume and, with some modification, for Opal at CERN and for Zeus at HERA.

Fig. 5 The measuring machine used for mapping the Aleph, Delphi and H1 magnets

In Fig. 5 the end-caps of the iron return yoke are indicated as 1 and 3, and the barrel part of the yoke as 2. The superconducting coil is shown as 4. An external stepping motor was used to drive a shaft (not shown) through an end-cap hole into one of the two bearings which support the main shaft (6) of the measuring machine. These bearings were mounted from the coil housing, not from the end-caps. This was important because of the end-cap movements under the enormous forces described in section 2. The measuring arms (7 and 8) extended each side of the main shaft. They were fixed into a collar which could be stepped along the shaft. The whole shaft and collar, and thus the measuring arms, could be stepped through azimuthal angle ϕ from zero to 360 degrees.
The measuring arms carried groups of six Hall plates to measure all field components. Such groups of six were distributed at intervals of 20 cm (at small radius) or 10 cm (at large radius) along the arms. This was the case out to a radius of 2 m for mapping the Delphi magnet, using 120 Hall plates, and a radius of 1.775 m using 96 Hall plates for mapping the H1 magnet. The measuring arms also carried seven NMR probes. Using this machine the extent of the field map was a cylinder of approximately 4 m diameter and 6 m length.

4.2 Static field sensors

A completely different approach has been adopted for mapping the major part of the 1200 m$^3$ of the L3 magnet. Approximately one thousand static magnetoresistive plates are fixed in the detector to measure the (approx. 0.5 T) main component to the required precision of 0.4 % [4]. The minor components were not measured, but calculated from parameters resulting from a fit to the main component in the manner to be described in subsection 5.3.1.
5. **USE OF THE FIELD MAP**

One needs to keep in mind the ultimate use of the field map. This will determine the required precision of the magnetic measurements, the degree of detail which should be provided to the user, and the ways in which the field map can be tested.

5.1 **Required precision**

As explained in section 1 the whole objective is to measure the momentum of each charged particle track in the detector. This clearly depends on the degree of bending which can be established for a track, which in turn depends on the resolution of the track detectors (bubble chamber, optical spark chambers, multi-wire-proportional chambers, drift chambers, etc.). If the momentum is so high that over the observed track length the sagitta is as small as the resolution of the detector then even the sign of charge may be uncertain! For this reason magnetic detectors are often said to measure $1/p$, with approximately Gaussian errors, rather than the momentum $p$ directly. A typical expression for the uncertainty $\sigma_p$ on a measurement of $p$ is

$$\frac{\sigma_p}{p} = 0.01 \times p \quad (3)$$

where $p$ is in GeV/c and of course the value of the constant varies from one detector to another. Clearly a relative precision of 1% at $p = 1$ GeV/c becomes 10% at $p = 10$ GeV/c. In the other direction it looks as if the relative precision becomes 1 part per thousand at $p = 0.1$ GeV/c, but the expression becomes invalid when multiple scattering becomes important. That depends on the amount of material in the detector.

These considerations limit the accuracy of momentum determination even if the magnetic field map is known with enormous precision. So there is no point in such enormous precision. Local errors of 1 part per thousand in the main field component, or 1% in the minor components will not affect the momentum determination, and are unlikely to be of any interest to the physicist user of the field map. Even a global systematic error of 1 part per thousand is usually acceptable. The field mapping is usually planned to determine the main component to 0.1% and minor components to 1%, well within the capabilities of temperature controlled Hall probes, or well calibrated search coils.

5.2 **Coarse structure and fine structure**

When the field mapping is complete it often shows local peculiarities, wiggles in one or more components near to holes in the iron or current lead connections. If these are at the level of 1 or 2 parts per thousand the physicist user does not need to know about them for momentum determination. They may be called "fine structure" effects of the field. With their neglect the field shape needed by the physicist may be called the "coarse structure" of the field. It has to be coded into the software so that a single subroutine call giving the space point will return with the field components. The physicist wants this subroutine to be fast!

The coarse structure of the field may have some symmetry, for example up-down symmetry or left-right symmetry in a rectangular pole-tip magnet as shown in Fig 2, or axial symmetry in a solenoidal magnet as shown in Fig 4. That is valuable for reducing the size of the field map which has to be stored in the fast subroutine. The reduction becomes really worthwhile when the symmetry is complete in one coordinate, like independence of azimuthal angle $\varphi$ in a solenoidal magnet. Then the field map is essentially two dimensional in $z$ and $r$ with field components $B_Z$ and $B_r$ only.

For the field map of the H1 magnet we found [5] several effects at the level of 1 or 2 parts per thousand. Although of interest to the magnet engineers, they were considered
unimportant for physics and so were classed as fine structure effects. Apart from these, the coarse structure which was coded into the fast subroutine had two important symmetries, axial symmetry (independence of azimuthal angle $\varphi$) and forward-backward symmetry with respect to the median plane.

5.3 Checking the field map

There are two stages of the checking process, one carried out by those responsible for the magnetic measurements before releasing the subroutine for physics use, and then another carried out by the physicists of the collaboration during the subsequent years.

5.3.1 Consistency with Maxwell's equations.

The Maxwell equations constrain the possible variations of the field components as a function of position in space. This fact can be used to check that the field map which it is proposed to release for physics use does not contain any inconsistent values. The technique is to find a magnetic scalar potential, given analytically as a function of position, which satisfies Laplace's equation and can be fitted to all of the field measurements with acceptable accuracy. If any part of the field map is seriously wrong, as could happen if a probe became skewed at some stage, then the wrong values stand out "like a sore thumb" in the fit. This technique was described in some detail in 1992 [5] and need not be repeated here.

5.3.2 Invariant mass histograms.

The classic test of correctly measured momenta is to study the decays of the strange particle

$$K^0 \rightarrow \pi^+ \pi^-$$

in which the momenta of the $\pi^+$ and $\pi^-$ tracks are measured. Some detectors, with Cerenkov counters for example, can identify the positively charged track as that of a $\pi^+$ and the negatively charged track as that of a $\pi^-$. Then the relativistic invariant mass is given by

$$m^2 = (E_1 + E_2)^2 - (p_1 + p_2)^2$$

$$\frac{1}{2}m^2 = m_{\pi}^2 + E_1E_2 - p_1p_2 \cos \theta$$

and can be calculated for the $\pi^+\pi^-$ pair, where $p_1$ and $p_2$ are their momenta making an angle $\theta$ at their joint origin, and $E_1$ and $E_2$ are their energies given by

$$E_1^2 = p_1^2 + m_{\pi}^2$$

$$E_2^2 = p_2^2 + m_{\pi}^2$$

Those $\pi^+\pi^-$ pairs which came from a kaon decay should have an invariant mass at the kaon mass, known to be [6]

$$m_K = 497.67 \pm 0.03 \text{ MeV} / c^2$$

Those $\pi^+\pi^-$ pairs having some origin other than the two-body decay of a particle may have any value for their invariant mass. Even positive and negative pairs not actually pions, but so called for calculating energies from measured momenta, will not have a unique invariant mass. Thus a histogram of invariant mass yields a clear kaon peak; it is only the level of the background under the peak which depends on the contamination from non-pions, and non-kaon decays.
The test of correct momentum determination is whether the position of the peak in the invariant mass histogram comes at the known kaon mass. Figure 6 shows the histogram resulting from studies in the H1 detector [7]. A fitted Gaussian shape to these data had a central value of

\[ m_{\pi\pi} = 497.9 \pm 0.2 \pm 0.2 \text{ MeV}/c^2 \]  

where the uncertainties are statistical and systematic. A discrepancy of less than 1 part per thousand from the known kaon mass is perhaps surprising, as the coarse structure of the field map was used, and is known to contain local deviations from the measured field by 2 or 3 parts per thousand. The explanation is probably that when averaged over thousands of tracks, in all parts of the detector, any local deviations from the true field become greatly diluted.

![Figure 6: The mass spectrum of pion pairs to check the momentum determination](image)

6. **THE FIELD MAP WITH HINDSIGHT**

However carefully one plans the magnetic measurements there often appears in later years (when it is too late because the magnet is full of detectors) some reason why extra measurements would have been useful. Our experience in H1 can be used to illustrate this.

Figure 4 shows schematically a typical solenoid detector like Aleph, Delphi or H1. The superconducting coil is, in principle, coaxial with the beams. The central region indicated with a dashed line surrounds the interaction point and can be taken as the region of the tracking detectors and thus the region over which the detailed magnetic survey was made. For physics analysis this has proved entirely satisfactory for all three experiments.

Figure 4 also shows schematically a few field lines rather parallel to the axis through the region of physics interest, but curving into the iron of the end-caps especially in the beam holes of the end-caps. For the H1 experiment magnetic measurements were not made within these holes of the end-caps.

One reason for regretting this is that physicists have added to the original detector several scintillation counters with phototubes in these holes. In order to optimize the magnetic shielding of the phototubes they wished to know the magnitude and direction of the field as a function of position. A second reason is that it was found that the orbit of the 12
GeV electron beam in HERA, at injection, was dependent on whether the H1 magnet was on or off. As can be seen from Fig. 4 there is a radial component of field inside the end-cap holes which increases with distance off axis. In principle there is no radial component on axis, indeed this can be taken as the definition of the magnetic axis of each end-cap hole. It may not be identical with the geometrical axis of each hole. In fact the HERA experts found that by moving the 12 GeV electron beam by several mm parallel to the geometric axis they could reduce or even eliminate the effect of turning on or off the H1 magnet.

With the benefit of hindsight one can thus say that measurements in the end-cap holes would have been useful for the machine operators, and for the hardware physicists, even though they have no relevance to the primary objective of momentum determination for physics analysis.

7. PROSPECTS FOR FUTURE DETECTOR MAGNETS

One reason for describing, in sections 2, 3 and 4, the development of detector magnets over the past three decades, was to attempt to see the next decade in perspective. The technical proposals for the CMS and Atlas detectors for LHC give the magnetic stored energies as thousands of MJ. This is not out of line with the trend already described, roughly an order of magnitude per decade. It is rather frightening nonetheless!

How much of our past experience with detector magnets will be applicable to the future? Perhaps magnetic measurements will not even be needed? Have computer codes for magnet design reached such a state of precision that they make measurements unnecessary? It seems to me prudent to make some form of measurement, even if it is only in limited regions of space, because that will determine the normalization for the calculated field.

* * *

REFERENCES

QUALITY CONTROL OF OBSERVATIONAL DATA IN METROLOGY

J. C. Iliffe
Department of Photogrammetry and Surveying, University College London, UK.

Abstract
This paper considers the treatment of errors in metrology. After a classification of the sources of error into systematic, random, and gross, the treatment of these is divided between management systems and statistical techniques. In the treatment of systematic errors, the emphasis is shown to be on minimising their impact by setting up appropriate documented procedures for calibration. Random errors are seen to be an inevitable occurrence, and the emphasis is on quantifying their effect on the quality of the final results, through the use of redundant observations and rigorous least squares techniques. The final classification, the gross error, is potentially the most serious in terms of the effect that this can have on the final solution, and hence the importance is stressed of not simply making check observations, but of quantifying the probability of gross errors remaining undetected.

1 INTRODUCTION

1.1 Error types

All measurements are subject to errors, which may come from a variety of different sources. The intention of this paper is to review the techniques that are available for minimising these errors, and for monitoring and quantifying them. Such techniques may take the form of management systems, or they may involve rigorous statistical methods.

In general, the types of error that can occur are classified as follows:

i) Systematic,

ii) Random, and

iii) Gross.

Whilst some sources of error can be thought of as belonging to more than one of these classes, this classification provides the most useful basis for examining errors, and it is the one that will be used in this paper to discuss the treatment of them. Before turning to each class in turn, however, it is necessary to discuss the fundamental terminology of quality assurance and quality control.

1.2 Quality assurance and quality control

Quality assurance can be defined as a set of procedures that "provide assurance that activities have been carried out to the required standard, and provides evidence that this has taken place" [1]. Hence, this is an area that is concerned with management systems rather than statistical techniques per se, and is therefore to be viewed as an over-arching framework that encompasses documentation and training.

Quality control, on the other hand, can be defined as including "those actions that provide a means to measure and regulate the characteristics of an item or service to established requirements" [1]. It is therefore concerned with physical actions, rather than the management environment, and would be expected to encompass mathematical techniques and statistical testing. Clearly in metrology there is a certain amount of overlap between the concepts of quality assurance and quality control. However, with these basic definitions in mind, we shall
proceed to examine each category of error in turn to establish the most appropriate methods for identifying them, quantifying them, and assessing their effect on the final result.

2 SYSTEMATIC ERRORS

A systematic error is one that is caused by a consistent bias in the observations, and is therefore something that will be present however many times an observation is repeated. The most common sources of this type of error are a mis-calibration of the instrument used and environmental factors.

Since there are unlikely to be any internal methods of detecting the presence of systematic errors (that is, by the self consistency of the results) this type of error is usually dealt with by adopting a particular set of procedures. In this sense, the treatment of systematic errors can be considered to fall within the scope of management (in the sense of arranging for certain procedures to be adopted and providing evidence that they have been carried out) and is therefore in many ways related more to quality assurance than to quality control.

The procedures that might be adopted will vary according to the type of measurement being made and the sources of systematic error that are likely to occur. Whilst this paper cannot present a complete set of procedures for every measurement that is likely to be made in metrology, one example will be given: that of electromagnetic distance measurement (EDM).

The technique of EDM essentially works by measuring a distance between an instrument and a reflector using a modulated electromagnetic wave of known frequency. In its simplest conception, the distance is determined by finding the number of full and part wavelengths along the measured line. The accuracy of the observation is dependent on the validity of certain basic assumptions:

i) That the electronic centre of the measurement coincides with the physical centre of the instrument over the point to be measured. Any difference between these will lead to a constant error (index error) in each observation.

ii) That the length of the modulated wave is known. Variations in this quantity may be caused either by errors in the frequency generator of the instrument, or by variations in the speed of the wave in the atmosphere. Whilst both these error sources will result in an error proportional to the length of line being measured (scale error), the first source will be constant and the second will vary with environmental parameters.

The key to the detection of these errors is calibration. A typical method of calibration [2] is to establish a baseline of several stable pillars in a line (Fig. 1).

![Fig. 1 A multi-pillar baseline](image)

By measuring either all the possible combinations of distance between the pillars (AB, AC, AD, etc) or a majority of them, it is possible to determine the index error of the instrument by the internal consistency of the observations. It is also possible to determine the scale error, but only if the distances between the pillars have previously been observed by an instrument of greater accuracy than the one being tested. An alternative to the latter is to measure directly the fundamental frequency of the instrument against a known standard.

As has already been pointed out, this paper cannot attempt to constitute a complete manual for the calibration of even one type or class of instrument. The concepts that can be drawn from this brief example, however, are those related to the management systems that govern this type of calibration. The key points that are of interest are those of traceability and documentation.
Traceability refers to the ability to trace a hierarchy of standards from the instrument under consideration, via instruments of higher standards, and ultimately to national and international standards. It is important to note that this applies not only to instruments that have been used to calibrate the test baseline, but also to the ancillary equipment such as that used to measure the ambient temperature and pressure of the atmosphere, and to equipment used to measure the fundamental frequency of the instrument.

Reference [2] gives examples of good practice for the documentation that should be produced to accompany these checks.

Thus it is seen that the management system, incorporating provable and documented checks, is the key to minimising the impact of systematic errors.

3 RANDOM ERRORS

The principal characteristic of random errors is, as the name implies, that they will not be the same for any repeat observations. All measurements will be subject to this type of error, due primarily to the limitations of the instrument being used. What is important is to be able to quantify these, and to assess their impact on the final result.

The treatment of random errors is a branch of statistics which can, and does, fill entire textbooks. This paper will summarise some of the key techniques that can be used. Fuller treatments can be found in, for example Refs. [3–5].

3.1 Error distribution

Repeated observations of a single quantity will result in a Gaussian error distribution. In the absence of systematic errors, the mean of this distribution will be centred on the true population value for an infinite number of observations (Fig. 2).

![Gaussian error distribution](image)

±1 σ = 68.2%

Fig. 2 Gaussian error distribution

Each individual observation, $\hat{\theta}$, differs from the best estimate of the measured parameter, $\bar{\theta}$, by an amount known as the residual, $\nu$. Thus:

$$\nu = \bar{\theta} - \hat{\theta}$$  \hspace{1cm} (1)

The variance of the observation is then defined as:

$$\sigma^2 = \frac{1}{n} \sum \nu^2$$  \hspace{1cm} (2)

where the quantity $\sigma$ is most commonly referred to as the standard error of the observation, and $n$ is the number of observations.
The use of the Gaussian error distribution leads to the well-known result that 68.2% of the observations will lie within ±1σ of the population value, and 95% will lie within ±1.96σ. Turning this around, we are more interested in the possible population values of the observed parameter that can be inferred from a single observation. The use of the Gaussian distribution means that confidence limits can be assigned for the population value based on the observation and its standard error.

In practice, then, the more usual situation is that instead of dealing with a very large number of observations, from which the standard error can be inferred, a small number of observations is made of any one parameter; the standard error is inferred from past experience or the specifications of the instrument.

This in turn leads to several questions, such as: is this estimate of the standard error of the observations a reasonable one? And what, in any case, is the effect of an error in the observation of one parameter on the quality of the parameters ultimately being derived? The most appropriate framework in which to address and answer these questions is through a least squares treatment of the observations.

3.2 Least squares analysis

Least squares analysis is the key to the questions posed in the previous section. In addition, it has the advantage of being a convenient mathematical framework for the treatment of complex situations, and also of giving an assessment of the quality of the results.

To give a brief summary of the procedures used, the first step is to establish a functional relationship between a set of observations, s (for example observed distances, angles, offsets), and a set of unobserved parameters, x (for example coordinates of points, bias parameters). This is written in the form:

\[ F(x; s) = 0 \]  \hspace{1cm} (3)

For example, the functional relationship between an observation of distance and the coordinates of the points it is measured between would, in two dimensions, have the form:

\[ F(\Delta x, \Delta y; s) = s^2 - \Delta x^2 - \Delta y^2 \]  \hspace{1cm} (4)

An expression of this form can then be linearised around a set of provisional values of the observed and unobserved parameters (\( \hat{s} \) and \( \hat{x} \) respectively):

\[ F(\hat{x}; \hat{s}) = F(\hat{x}; s) + \frac{\partial F}{\partial x} dx + \frac{\partial F}{\partial s} ds \]  \hspace{1cm} (5)

where the terminology (\( \hat{x}; \hat{s} \)) represents the final best estimates of these parameters, and:

\[ dx = \hat{x} - \hat{x} \]  \hspace{1cm} (6)

\[ ds = \hat{s} - \hat{s} = \hat{s} - \hat{s} + v = 1 + v \]  \hspace{1cm} (7)

The problem that then has to be solved is to determine the solution of a set of linear equations of the form:

\[ \frac{\partial F}{\partial x} dx + \frac{\partial F}{\partial s} ds + 1 + \frac{\partial F}{\partial s} v = 0 \]  \hspace{1cm} (8)

In simple cases this can be expressed as a matrix equation:

\[ A.x = 1 + v \]  \hspace{1cm} (9)

This is then solved according to the condition that the weighted sum of the squares of the residuals is a minimum, which is achieved by introducing a weight matrix, W, containing the variances and covariances of the observed parameters:
\[ v^T W v \rightarrow \text{minimum} \]  \hspace{1cm} (10)

It can be shown [3, 4] that the solution to this is:

\[ x = (A^T W A)^{-1} A^T W b \]  \hspace{1cm} (11)

or

\[ x = N^{-1} b \]  \hspace{1cm} (12)

3.3 Quality control of program

Equation (12) gives the least squares solution for a set of unobserved parameters as a function of a set of observations and their associated standard errors. A brief comment should be made at this stage about the potential for one of the most serious of all possible systematic errors: that of using a computer program that is itself incorrect [6].

As with other systematic errors, the solution to this problem lies in the management systems that are adopted. In this case, a new program must be checked in at least one of several different ways:

i) By comparing the solution with that obtained by other programs that share the same functionality. In the case where a new program is in fact just a transfer to a new platform, it is likely that the two programs will have all elements in common, and this is straightforward. In other cases, it is less likely that another program will be able to give the results of the new one.

ii) For simple cases, a comparison may be made with a manual solution of the problem, and the validity of the program confirmed by extension.

iii) Comparisons may be made with standard data sets at other institutions.

iv) Internal mathematical checks may be made on some stages of the least squares process. For example [3], it should be the case that:

\[ A^T W v = 0 \]  \hspace{1cm} (13)

In all cases, these checks should be adequately documented at every stage.

3.4 Observational errors

Having obtained a solution through equation (12), it is necessary to check several of the assumptions that have been made. The first of these is the question of whether the observational standard errors that were used to form the weight matrix were a reasonable reflection of the quality of the observations. This is important for two reasons. Firstly, if the wrong relative weight is given to different observations then this may affect the solution that is obtained. Secondly, an unbiased estimate of the quality of the observations is essential in determining the quality of the final result (that is, the standard errors of the unobserved parameters).

If the estimated values of the standard errors of the observations were correct, then it would be expected that the residuals, \( v \), would be on average about the same magnitude as the standard errors, provided that sufficient account is taken of the amount of redundancy in the situation. That is,

\[ \left( \frac{v}{\sigma} \right)^2 = 1 \]  \hspace{1cm} (14)

for the situation where the number of observations greatly exceeds the number of unobserved parameters. In the situation where there is no redundancy, the best estimates of the observed parameters will exactly equal the observed values and the residuals will equal zero.

More formally, the parameter \( \sigma_0 \) is defined as the standard error of unit weight, and is given by:
\[ \sigma_0^2 = \frac{v^TWv}{m - n} \] (15)

where \( m \) and \( n \) are the numbers of observed and unobserved parameters respectively. Hence a value of \( \sigma_0 \) less than 1 will indicate that on average the residuals are smaller than the standard errors had predicted, and these were therefore pessimistic. The opposite conclusion could be drawn from a value of \( \sigma_0 \) greater than 1.

In general, therefore, better overall \textit{a posteriori} estimates may be made of the standard errors of the observations:

\[ \sigma_{\text{a posteriori}} = \sigma_0 \sigma_{\text{a priori}} \] (16)

This will not affect the final result, as all observations are being treated equally, but it will improve the estimate of its quality. In fact, it is advisable to examine the individual residuals to see whether any particular group of observations had poorly estimated standard errors.

Of course, it is unlikely that \( \sigma_0 \) will be exactly equal to unity; how significant the difference is will depend on the number of observations. The significance of the difference can be tested using the Fisher F-test [4].

3.5 Dispersion matrices

In considering the quality of the observations and the unobserved parameters, an important concept that should be introduced at this stage is that of dispersion matrices.

Dispersion matrices are square matrices equal in size to the vector of parameters whose errors they are characterising. Thus, if there are \( m \) observations, then the dispersion matrix of the observations will be an \( m \times m \) matrix, which contains the covariances between all combinations of observations. The diagonal elements of the matrix will simply contain the variances of the observations. In fact \( D_0 \), the dispersion matrix of the observations, is related to the previously encountered weight matrix by the expression:

\[ D_0 = W^{-1} \] (17)

The underlying meaning of the dispersion matrix for the observations, is the relationship to the true population value of the observed parameter, \( \bar{s} \). This is expressed as:

\[ \bar{s} = \bar{s}^o + \bar{v} \] (18)

Thus, although the population residual, \( \bar{v} \), can never be known, the dispersion matrix for the observations contains the expected values of the cross products of these quantities for all observations.

In a similar fashion, the final best estimates of the unobserved parameters are related to the true population values by:

\[ \hat{x} = \hat{x}^o + \hat{\delta x} \] (19)

The dispersion matrix for the unobserved parameters, \( D_x \), then contains the expected values of the cross-products of the quantities \( \delta x \), which is the same as the variances and covariances of the unobserved parameters. It can be shown [3] that this is given by:

\[ D_x = \sigma_0^2 N^{-1} \] (20)

Other dispersion matrices are then defined in a similar fashion. These include: the dispersion matrix for the observed parameters, \( D_s \), which describes the quality of the best estimates of the observed parameters; and \( D_v \), which indicates the typical size (as opposed to quality) of the residuals. It is important to realise that in a situation where there is no redundancy, the residuals will be zero.
3.6 Precision of the result

The dispersion matrix of the unobserved parameters can therefore be used to extract the standard errors of final results. This is then a measure of the precision with which a parameter has been determined. Note the use of the term 'precision' in this context, as opposed to 'accuracy', which would imply a complete absence of systematic errors.

The precision of the final result could be presented in a form such as, for example:

\[ x = 123.41 \pm 0.03 \]  \hspace{1cm} (21)

where the \( \pm \) would usually be understood to represent \( \pm 1\sigma \), and hence gives 68.2% confidence limits on the solution. Alternatively, 95% confidence limits could be expressed by using \( \pm 1.96\sigma \).

For a situation in which the unobserved parameters are the two-dimensional co-ordinates of points, these standard errors could be expressed graphically in the form of error bars, as in Fig. 3.

If we consider the error component in each direction, however, it may be apparent that this does not give the best representation of the situation. This is illustrated in Fig. 4, which demonstrates that the direction of the maximum error is not necessarily aligned with either of the conventional axes.

Fig. 3 Standard error bars on a point in two dimensions

Fig. 4 Error ellipse for a point in two dimensions

By extracting from the dispersion matrix for unobserved parameters the variances of the x and y coordinates, \( \sigma_x^2 \) and \( \sigma_y^2 \), and the covariance \( \sigma_{xy} \), the orientation of the error ellipse may be found [3] from the expression:

\[ \tan(2\alpha_m) = \frac{2\sigma_{xy}}{\sigma_x^2 - \sigma_y^2} \]  \hspace{1cm} (22)

and the maximum dimension may be found from:

\[ 2\sigma_{\text{MAX}}^2 = \sigma_x^2 + \sigma_y^2 + \frac{2\sigma_{xy}}{\sin(2\alpha_m)} \]  \hspace{1cm} (23)

It is useful to note here that, provided the a priori estimates of the standard errors are accepted as reasonable, none of the assessments of quality outlined above requires any actual observations to be made. Thus, the quality of the final result can be predicted by a simulation based on a proposed observational scheme, to check beforehand whether this is likely to be acceptable.
4 GROSS ERRORS

The foregoing discussion on random errors is based upon the assumption that all errors are drawn from a normal distribution centred on zero and of determinate size, which would generally be the case when errors are due simply to the limitations of the measuring device. This is not always the case, however; some errors are due to causes that are not of a predictable nature, such as a disturbance during the act of measurement or a misinterpretation by the operator. The only predictable aspect of these is simply the fact that they do occur from time to time.

Since an error of this type is only likely to be distinguishable from a random error by its size, it is termed a gross error. Clearly, the undetected presence of an error of this type is likely to lead to a substantial bias in the solution, and to invalidate any conclusions that are drawn about the accuracy and precision of the result. The task, therefore, is to detect the presence of these errors, and to eliminate the corresponding observations from the data set.

4.1 Outlier detection

It is recalled that for a Gaussian distribution, 95% of residuals will lie between the limits ±1.96σ, where σ is the standard error. Therefore, an observation that gives rise to a residual outside these limits can be said to have a 5% probability of belonging to the normal population. That is, if the observation were rejected, then 5% represents the probability of rejecting a valid observation.

Similarly, if the residual lies outside the limits ±3σ, then the observation has a 0.2% chance of belonging to the normal distribution.

The flaw in this argument, however, lies in the question of which standard error is being referred to. Clearly, for a situation in which there is no redundancy (i.e. the minimum number of observations necessary to find a solution has been made) then the residuals will always be zero. Hence the concept of σ simply referring to the standard error of the observation is clearly not sufficient. More formally, then, two alternative hypotheses are proposed for the ith observation:

\[ H_0: \hat{s}_i = \bar{s}_i + \varepsilon_i \]  
(24)

and

\[ H_1: \hat{s}_i = \bar{s}_i + \varepsilon_i + \Delta_i \]  
(25)

H_0 is proposing that the observed value, \( \hat{s}_i \), is made up of the true population value, \( \bar{s}_i \), plus a random error, \( \varepsilon_i \). The alternative, H_1, is that it also includes a gross error, \( \Delta_i \).

The following test statistic is then introduced:

\[ w_i = \frac{\hat{d}_i}{\sigma_d} \]  
(26)

where

\[ \hat{d}_i = \hat{s}_i - \bar{s}_i \]  
(27)

Thus, \( \hat{d}_i \) is the difference between the observed value of the parameter and the value \( \bar{s}_i \) that is computed for this parameter using all available observations except \( \hat{s}_i \). The quantity \( \sigma_d \) is the corresponding standard error of this.

If H_0 is correct, then \( w_i \) will be distributed normally with a mean of zero and a standard error of unity. Otherwise, it will have mean

\[ \delta_i = \frac{\Delta_i}{\sigma_d} \]  
(28)

A test for gross errors can therefore be proposed, in which the value of \( w_i \) is computed for each observation. It may then be suggested that any observation that gives rise to a value of
$w_i$ greater than a key ratio, $a$ (for example 1.96) should be rejected. It should be recognised that in rejecting values on this (or indeed any) criterion, some valid observations may occasionally be rejected. Similarly, there will be situations where a gross error may be less than the cut-off value, or may indeed be greater than the cut-off value but combines with a random error of opposite sign to bring it back within acceptable limits. Thus, although the gross error is biasing the solution, it is not detected.

Figure 5 illustrates a situation in which observations with values of $w$ greater than $a$ (=1.96, say) are rejected; an alternative error distribution centred on a gross error is also shown. In the situation illustrated, this observation would be rejected 90% of the time; in 10% of cases the random errors would be sufficient to bring the observation within the acceptable limits. For this situation, the value of $b$ is 1.28.

![Fig. 5 Marginally detectable gross errors](image)

The gross error illustrated is therefore termed a marginally detectable error, since it can be defined as the limiting value of a gross error that has a 90% chance of being detected under the stated conditions of the test: any errors larger than this would have a greater chance of being detected, and any smaller ones would have a worse chance.

The size of the marginally detectable error is therefore governed by the value $\delta^u_i$, which for the stated test conditions is given as:

$$\delta^u_i = a + b = 3.24$$

and hence

$$\Delta^u_i = \delta^u_i \cdot \sigma^*_d = 3.24 \cdot \sigma^*_d$$

is the actual size of the marginally detectable error. Conveniently, for purposes of computation, [4] shows that Eq (26) can be simplified under certain circumstances to:

$$w_i = \frac{v_i}{\sigma_v}$$

and the standard error $\sigma^*_d$ can be found from:

$$\sigma^*_d = \frac{\sigma^2_i}{\sigma_v}$$

in which it should be noted that $\sigma_i$ is the standard error of the observation and $\sigma_v$ is the standard error of the residual, which is found from the dispersion matrix for residuals, $D_v$. 
4.2 Reliability of observations

The likelihood of being able to detect and reject a gross error is termed the reliability of an observation. The underlying basis for this is as shown in section 4.1, but it is convenient to be able to express the reliability of an observation in terms of a simple statistic.

This is done by introducing the statistic $\tau$. The full definition of this is rather complicated, but [4] shows that for situations in which all observations are uncorrelated the statistic can be defined for a particular observation as:

$$
\tau_i = \frac{\sigma_{d_i}}{\sigma_i} = \frac{\sigma_i}{\sigma_v}
$$

Incorporating this with equations (30) and (32) gives the marginally detectable error as:

$$
\Delta_i^u = \delta_i^u \sigma_i \tau_i
$$

Since $\delta_i^u$ and $\sigma_i$ are fixed for any given observation and set of test criteria, the size of gross error that may occur without being detected is principally proportional to the statistic $\tau$. Hence, a low value of $\tau$ is desirable.

If we consider the way in which Eq. (33) operates, it can be seen that a situation in which there is no redundancy at all will lead to residuals which are always equal to zero: hence the standard error of the residual, $\sigma_v$, will also be equal to zero as there is no variation. With decreasing redundancy, therefore, the value of $\tau_i$ in equation (33) will tend to infinity, and so therefore will the marginally detectable gross error.

Conversely, with a great deal of redundancy, the standard error of the residual will approach the standard error of the observation, which implies that the value of $\tau_i$ will approach 1 from above. Therefore, the values of $\tau$ range from a minimum of 1, representing highly reliable observations, to infinity, representing completely unreliable observations with no redundancy.

Thus, we have finally arrived at the point where we can quantify the contribution that "check observations" are actually making in an observational scheme. This is illustrated in the examples given in the following section.

4.3 Examples

To illustrate the use of the $\tau$ statistic in determining reliability, two rather extreme examples will be given. Consider first the situation illustrated in Fig. 6. In the diagram, the eight points A to H are of known coordinates. The point P is unknown, and is to be determined by distance observations made from each of the known points. Each observation of distance is made with a standard error of 1 cm.

![Fig. 6 An observational scheme with high redundancy](image)

![Fig. 7 An observational scheme with minimal redundancy](image)
As a result of the least squares computation of the position of P, the coordinates are found and the standard errors of the positions along each of the axes are found to be:

\[ \sigma_x = 0.005 \text{ m} \]

\[ \sigma_y = 0.005 \text{ m} \]

(35)

It would be possible to compute the value of any of the observed parameters (the distances from the fixed points to P) from the coordinates so obtained, and it is clear from (35) that this would be done to a higher precision than the original observations themselves. This is the key to the isolation of gross errors.

Table 1 shows the values of \( \tau \) that have been computed for each of the eight observations.

**Table 1**

Values of \( \tau \) for observations shown in Fig. 6

<table>
<thead>
<tr>
<th>Observation made from:</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau_i )</td>
<td>1.31</td>
<td>1.42</td>
<td>1.31</td>
<td>1.28</td>
<td>1.28</td>
<td>1.34</td>
<td>1.42</td>
<td>1.34</td>
</tr>
</tbody>
</table>

Taking the observation from A as a typical example, the meaning of the figures is clear from equation (34), which can for this observation be re-written as:

\[ \Delta_A^u = \delta_A^u \sigma_A \tau_A = 3.24 \times 0.01 \times 1.31 = 0.042 \text{ m} \]

(36)

That is, the largest gross error that has a worse than 90% probability of being detected is 4.2 cm.

For comparison, consider next the situation illustrated in Fig. 7. In this scheme, the three points A, B, and D are fixed, and the point P determined by distance observations from the fixed points, each with a standard error of 1 cm. After the computation of position, the standard errors of the coordinates of the point P are given as:

\[ \sigma_x = 0.007 \text{ m} \]

\[ \sigma_y = 0.010 \text{ m} \]

(37)

In this situation, it is seen that the y coordinate of P is found to the same level of precision as the original observations, whilst the x coordinate is found to a slightly better precision. Table 2 then shows the \( \tau \) values that are computed for the three observations.

**Table 2**

Values of \( \tau \) for observations shown in Fig. 7

<table>
<thead>
<tr>
<th>Observation made from:</th>
<th>A</th>
<th>B</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau_i )</td>
<td>2.00</td>
<td>202.0</td>
<td>2.02</td>
</tr>
</tbody>
</table>

In this situation, applying equation (34) to the observation of distance made from point B:

\[ \Delta_B^u = \delta_B^u \sigma_B \tau_B = 3.24 \times 0.01 \times 202.00 = 6.545 \text{ m} \]

(38)

That is, a gross error in the observation from B, equal to as much as 6.545 m, could occur with a 90% chance of being detected. To state the obvious, there is an unacceptable 10%
probability of this error remaining undetected, with disastrous effects on the coordinates that are adopted for the point P.

Clearly, the examples given here are somewhat extreme, and it might be said that common sense would have arrived at a conclusion on the validity of the observational scheme shown in Fig. 7. These examples have been chosen for clarity, however, and there will be situations that are less extreme and employ more complex measuring systems, in which a formal method of evaluating the ability to check for gross errors is invaluable.

5 CONCLUSIONS

The aim of this paper has been to examine the treatment of errors in metrology. It has been shown that the methods for dealing with errors vary from the use of appropriate management systems to statistical analysis, and that the choice of technique depends on the type of error that is being considered.

In fact, of course, there is no "choice" as such, as all types of error should be expected to occur in any observational scheme, and therefore all the techniques and systems discussed here would normally be brought to bear in parallel.

Quality means fitness for purpose, and in the context of metrology the final meaning of this is that positions or coordinates are to be determined within a specified tolerance and at a given level of probability. For a set of observations to be thought of as fully "quality controlled", this means that all possible sources of error have been considered, and steps taken to minimise and quantify them. This paper has demonstrated that, except in the most undemanding and obvious of circumstances, the answers to these problems lie in:

i) Documented calibration procedures to reduce systematic errors.

ii) Least squares analysis of observations, to quantify the effect of random errors on the final solution.

iii) An analysis of the reliability of observations, to quantify the probability of gross errors or blunders remaining undetected, and to ensure that this is within acceptable limits.

* * *

REFERENCES


SETTING REFERENCE TARGETS

* SLAC-Pub 7447, April 1997

* Work supported by the Department of Energy, Contract DE-AC03-76SF00515
Abstract
Reference targets are used to represent virtual quantities like the magnetic axis of a magnet or the definition of a coordinate system. To explain the function of reference targets in the sequence of the alignment process, this paper will first briefly discuss the geometry of the trajectory design space and of the surveying space, then continue with an overview of a typical alignment process. This is followed by a discussion on magnet fiducialization. While the magnetic measurement methods to determine the magnetic centreline are only listed (they will be discussed in detail in a subsequent talk), emphasis is given to the optical/mechanical methods and to the task of transferring the centreline position to reference targets.

1. INTRODUCTION
Particles travel undisturbed on their design trajectory only when the magnetic axes of all beam steering components form a smooth continuous line in space. In addition, the axes of diagnostic instruments need to coincide with the particles’ trajectory for good performance. Establishing these positioning conditions is the goal of the alignment process. The setting of reference targets is a key task in this process. These targets represent physically what otherwise cannot be accessed or referenced. In the task sequence of the alignment process, reference targets are used to represent the surveying coordinate systems (i.e. the surface net and the tunnel net), the geometric or electrical axes of diagnostic instruments, and the magnetic axes of the beam steering components.

The following definition conventions are used throughout the text: reference targets in the context of coordinate systems are referred to as monuments, and in the context of beam steering components or diagnostic components they are called fiducials. Surveying refers to measuring the position of an object, while alignment describes the action of adjusting an object’s position.

2. REFERENCE TARGETS
Reference targets are used to represent virtual entities like the magnetic axis of a magnet or a coordinate system, which otherwise cannot be physically accessed. The design of the targets is a function of the survey and alignment instrumentation used with these targets, the placement location, the physical size of their host component if applicable, and the required measurement accuracy. In general, reference targets can be categorised by whether they are fixed, installed targets or removable and by the number of degrees of freedom they reference.

2.1 Fixed or removable fiducials
While fixed targets have an advantage in terms of accuracy, removable targets or fixtures can have a significant cost advantage. In beam lines where hundreds of structurally identical magnets are used, a limited number of fixtures can substitute for hundreds of fixed fiducials.
2.1.1 Fixtures

Because of favourable magnetic properties, magnets are often constructed from stacked laminations. The same die is used to punch these laminations for all magnets, even for a large quantity. The dimensional features of the die can be machined to high accuracy, and are usually verified before and during use. Consequently, the mechanical dimensions of the laminations are very precise and their mechanical axis should coincide closely with the magnetic axis. Hence, it is possible to forego the fiducialization of individual magnets and instead reference the magnetic axis using clamp-on fixtures. These fixtures are seated in reference grooves and carry the actual targeting (see Fig. 1, 2).

![Fig. 1 Alignment fixture on quadrupole with reflector and inclinometer](image1)

![Fig. 2 Fixture referencing in lamination groove](image2)

2.1.2 Fixed fiducials

Usually, the fiducials are installed directly on the top or side of a magnet. While three fiducials, or two fiducials plus a roll reference surface are sufficient to reference all degrees of freedom, additional fiducials provide redundancy for error checking. If it is necessary to separate the fiducial from temperature caused expansion or contraction of the magnet body due to tight positioning requirements, the fiducials can be mounted on fiducial plates (see Fig. 3). These plates, made of invar, reference to the split planes of the magnet. Since the split planes do not move due to temperature, and since invar has a negligible expansion coefficient, the fiducial position can be considered invariant with temperature.
2.2 Fiducial and monument designs

2.2.1 One-dimensional reference targets

Traditionally, coordinate systems were represented by 2+1 D monumentation. Separate reference marks were used to represent the horizontal (2-D) and vertical (1-D) coordinate systems. Typical 1-D designs are rivets grouted into the floor (see Fig. 4) or levelling bolts set into walls (see Fig. 5).

2.2.2 Two-dimensional reference targets

The most basic 2-D reference mark can be scribe-lines on the floor (see Fig. 6, 7) or stick–on targets. A variation is the SLAC SLC floor-marks; to protect the stick-on targets from traffic wear, they are mounted inside small metal cans which are grouted into the floor (see Fig. 8). Other 2-D references are the standard Leica and Kern forced centering mounts. The Kern mount is commonly used on pillar monuments representing surface network points (see Fig. 9).
2.2.3 Three-dimensional reference targets

3-D reference marks provide both horizontal and vertical reference.

**Tooling Ball Reference**  The most inexpensive kind of 3-D references are tooling ball holes drilled into the bodies of components or tooling ball bushings tack-welded onto components (see Fig. 10). Common tooling balls inserted into the hole or bushing (see Fig. 11) provide an excellent reference for levelling mini-rods or optical tooling blades. For optical pointing, the tooling balls can be exchanged for targets with identical dimensions but with a cross hair or other target pattern at the exact centre of the virtual sphere (see Fig. 12). For distance measurements, reflector tooling balls are available, which have glass or air cubes mounted into the sphere such that the optical centre of the reflector coincides with the centre of the sphere (see Fig. 13). Commonly, tooling balls with either a 0.5-inch head diameter or a 0.875-inch head diameter are used.

**1.5-inch Sphere System**  The equivalent of the tooling ball bushing for the 1.5-inch sphere system is usually referred to as a nest. A three-point mount provides a kinematic centering for the sphere. Also used are mounts with conical surfaces (see Fig. 14). As with tooling ball targets, the 1.5-inch sphere can be host to a variety of target patterns and applications (see Fig. 15). A sphere-mounted reflector with front mirror surfaces (air cube) is the commonly used laser tracker reflector.
3.5-inch Sphere Target  The 3.5-inch standard is the evolutionary ancestor to both of the other two target systems. These spheres are a traditional item in optical tooling based alignment. As with tooling ball targets and 1.5-inch sphere targets, the 3.5-inch sphere can be host to a variety of targets (see Fig. 16, 17). However, because of its bulkiness, the 3.5-inch standard is rarely used anymore. A flavour of this size standard is the CERN socket (see Fig. 18), where a sleeve clamps a 3.5-inch sphere with a 30 mm bore onto a conical surface. This design provides the flexibility to adjust the sphere such that the bore is parallel to gravity. Using a special mounting adapter, targets and instruments can be mounted with reference to gravity on sloped surfaces (see Fig. 19).
Fig. 19 CERN socket with E2 mounted in sloped tunnel

3. TRAJECTORY DESIGN SPACE

Three translations and three rotations define the position of any object in space with respect to a Cartesian coordinate system. These values are a function of the design beam trajectory, the position of magnetic fields along the trajectory, and of the placement of the trajectory on our earth.

3.1 Design trajectory

Codes like MAD [1], NOMAD, or TRANSPORT [2] are used to trace the path of particles through idealised magnets. The particles’ path is described by a sequence of beam elements placed sequentially along a reference orbit. The reference orbit consists of a series of straight-line segments and circular arcs. A beam following coordinate system describes the
orientation of the beam at any point along its path through the accelerator (see Fig. 20). This system remains tangent to the orbit with its positive z-axis pointing downstream. The system is rotated such that the positive x-axis generally points out from the bending arc and lies in the plane of the curve. The positive y-axis is oriented to complete the right-handed coordinate system for the local beam. To transform the local coordinates into the global system, three shifts and three rotations are applied. Care has to be taken with the sequence of the rotations and their signs (see Table 1).

![Fig. 20 Beam following coordinate system](image)

**Table 1**

<table>
<thead>
<tr>
<th></th>
<th>MAD</th>
<th>TRANSPORT</th>
<th>Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yaw, around y-axis $\Theta$</td>
<td>Positive, when z-axis turns towards x-axis</td>
<td>Same as MAD</td>
<td>Same as MAD</td>
</tr>
<tr>
<td>Pitch, around x-axis $\Phi$</td>
<td>Positive, when z-axis turns towards y-axis</td>
<td>Same as MAD</td>
<td>Opposite</td>
</tr>
<tr>
<td>Roll, around z-axis $\Psi$</td>
<td>Positive, when x-axis turns towards y-axis</td>
<td>Same as MAD</td>
<td>Same as MAD</td>
</tr>
</tbody>
</table>

The output of the above mentioned codes gives three coordinates and three rotations for the beginning and end of each straight section and arc in the beam following system. It is now the magnet engineer’s job to design a component, which replicates its virtual cousin’s parameters.

### 3.2 Magnet reference target parameters

#### 3.2.1 Magnet coordinate system

The parameters of the reference targets of a magnet are defined in the magnet’s local coordinate system. The $u$, $v$, $w$ coordinate axes form a right-handed Cartesian system. The positive orientation of the axes is defined in the same way as for the $x$, $y$, $z$ beam-following local coordinate system. The datum of the $u$, $v$, $w$ system relative to a magnet is not universally defined. While the $w$ coordinate axis is usually coincident to the magnetic axis and the $u$-$w$ plane parallel to the zero roll plane, there are differences in the definition of the origin. Often, the origin is placed where the beam enters the magnetic field. This is a virtual point and requires entering empirical data into the fiducialization process. At SLAC the
midpoint of the magnet is the origin (see Fig. 21). Since the midpoint of the magnetic length usually coincides with the midpoint of the steel length, the origin can be physically determined without knowing the difference between magnetic and steel length.

3.2.2 Roll

The split planes of magnets are often accepted as the zero roll planes. However, if necessary, zero roll planes can be determined to better accuracy with magnetic measurement methods. A dipole’s roll can be measured by running a horizontal wire in a vertical plane. No voltage is induced if the magnetic field is oriented perpendicular to this plane. Hence, by adjusting the magnet’s tilt around the w-axis until no voltage is induced, the zero roll position can be determined. In the case of a quadrupole, no voltage is induced if a wire moves radially relative to one of the poles. If, e.g., a quadrupole’s pole design position is at a 45º angle, this method allows determination of the true 45º-field plane. From there it is straightforward to determine the zero-roll position.

In some instances, a geometrical correction to the design program’s roll value for dipoles needs to be applied. The w-axis of dipoles does not coincide with the x-axis of the beam-following coordinate system. Hence, if the origin of a dipole’s coordinate system has been fixed where the beam enters the magnetic field, the roll value from the trajectory design program does not reflect the correct dipole roll angle (see Fig. 22). A correction, which is a function of half of the dipole’s bending angle, needs to be applied. However, if the magnet’s coordinate system originates at the midpoint, then the design program’s beam-following roll at the midpoint is the correct physical magnet roll.

Larger machines need to take into account the convergence of the gravity vectors (see Fig. 23). While in circular machines magnets at diametrically opposing locations are at the same elevation, their ideal roll values need to be corrected. E.g. for PEPII the correction amounts to about 0.15 mrad, SSC magnets would have needed a 1.5 mrad correction. Also,
other geodetic corrections related to the beam line’s geometry may apply (see following Section).

3.2.3 Sagitta

In the dipole case the magnet is usually not centered at the midpoint, but shifted by a fraction of the sagitta height (see Fig. 21 above). This shift minimises the necessary aperture to accommodate a straight beam pipe, and moves the beam into the homogenous field region of the magnet. The amount of the shift varies. It is typically set to half of the sagitta, but two thirds are also not uncommon.

3.3 Global reference points

Global reference points are used to determine the overall shape of an accelerator. With the exception of small machines, these points represent the global coordinate system in the form of pillar monuments on the surface and floor marks in the tunnel.

3.3.1 Shape of the Earth and survey reference frames

The goal is to define a computational reference frame, i.e. a mathematical model, of the space in which the surveyor takes his measurements and performs his data analysis. Transformation algorithms and parameters between the surveying space and the machine layout coordinate system must be defined.

Ancient civilisations realised that the earth is round, and geodesy was born when the Greek Eratosthenes (born 276 BC) first attempted to determine its size [4]. The earth is actually of a more complex shape, the modelling of which is not easy. Three surfaces are of importance to the geodesist studying the shape of the earth:

*Terrain Surface*  The terrain surface is irregular, departing by up to 8000 m above and 10000 m below the mean sea level.

*Geoid*  The geoid is the reference surface described by gravity; it is the equipotential surface at mean sea level that is everywhere normal to the gravity vector. Although it is a more regular figure than the earth’s surface, it is still irregular due to local mass anomalies that cause departures of up to 150 m from the reference ellipsoid. As a result, the geoid is nonsymmetric and its mathematical description nonparametric, rendering it unsuitable as a reference surface for calculations. It is, however, the surface on which most survey measurements are made as the majority of survey instruments are set up with respect to gravity.
Ellipsoid  The spheroid or ellipsoid is the regular figure that most closely approximates the shape of the earth, and is therefore widely used in astronomy and geodesy to model the earth (Fig. 24). Being a regular mathematical figure, it is the surface on which calculations can be made.

In performing geodetic calculations, account must be taken of the discrepancy between the ellipsoid and the geoid. The deflection (or deviation) of the vertical is the angle of divergence between the gravity vector (normal to the geoid) and the ellipsoid normal (Fig. 25). Several different ellipsoids have been defined and chosen that minimise geoidal discrepancies on a global scale, but for a survey-engineering project, it is sufficient to define a best-fit local spheroid that minimises discrepancies only in the local area. Whatever ellipsoid is chosen, all survey measurements must be reduced to the ellipsoid before computations can proceed. This reduction of observations to the computational surface is an integral part of position determination [5]; the equations can be found in most of the geodetic literature, e.g., in Leick [6].

For large projects, the magnitude of these corrections is significant. A geoidal height determination was part of the LEP surface net measurement plan. It was found that the vertical deflections approached close to 15 arc seconds, which resulted in a separation between the local reference ellipsoid and the geoid of up to 200 mm (see Fig. 26). To nevertheless obtain a true plane in space, corrections varying between −40 mm and +100 mm had to be applied to the levelled elevations [7].

Fig. 24 Ellipsoid and geoid

Fig. 25 Deflection of the vertical

Fig. 26 LEP Geoid undulations (from [7])
3.3.2 Surveying Coordinate System

Computations with spheroidal (geographical) coordinates latitude $\phi$, longitude $\lambda$, and height $h$ are complex. They are also not very intuitive: when using spheroidal heights, it can appear that water is flowing uphill. Especially in survey engineering projects, coordinate differences should directly and easily translate into distances independent of their latitude on the reference spheroid. Therefore, it is desirable to project the spheroidal coordinates into a local Cartesian coordinate system or, going one step further, to project the original observations into the local planar system to arrive directly at planar rectangular coordinates.

A transformation is required to project points from a spheroidal surface to points on a plane surface. Depending on the projection, certain properties of relationship (distance, angle, etc.) between the original points are maintained, while others are distorted. It is simply not possible to project a spherical surface onto a plane without creating distortions [8] (see Fig. 27), but since these distortions can be mathematically

![Fig. 27 Mapping distortion](image-url)
modelled, it is possible to correct derived relationships, such as distances, angles, or elevations. This situation can be vividly shown in the example of the projection of levelled elevations onto a planar coordinate system (see Fig. 28). Table 2 shows the projection errors as a function of the distance from the coordinate system’s origin. Notice that the deviation between plane and sphere is already 0.03 mm at 20 m.

![Curvature correction](image)

**Table 2**

<table>
<thead>
<tr>
<th>Distance [m]</th>
<th>Sphere $H_s$ [m]</th>
<th>Spheroid $H_s$ [m]</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.00003</td>
<td>0.00003</td>
</tr>
<tr>
<td>50</td>
<td>0.00020</td>
<td>0.00016</td>
</tr>
<tr>
<td>100</td>
<td>0.00078</td>
<td>0.00063</td>
</tr>
<tr>
<td>1000</td>
<td>0.07846</td>
<td>0.06257</td>
</tr>
<tr>
<td>10000</td>
<td>7.84620</td>
<td>6.25749</td>
</tr>
<tr>
<td>25000</td>
<td>49.03878</td>
<td>39.10929</td>
</tr>
</tbody>
</table>

3.3.3 Survey networks

Monuments physically represent the surveying coordinate system. The coordinates of these monuments are determined using conventional trilateration or triangulation methods or, for larger size projects, satellite methods like the Global Positioning System [9].

*Surface network* In order to achieve the absolute tolerance and the circumference requirements, a surface network (see Fig. 29) with pillar-type monuments (see Fig. 30) must usually be established. Traditional triangulation and trilateration methods or GPS surveys can be applied to measure the coordinates of the monuments and of tripods over the transfer shafts or sightholes. Differential levelling of redundant loops is the standard method to determine the vertical coordinates. Proper reduction of measured distances also requires accurate elevation difference data.

![LEP surface network](image)

![Monument pillar](image)
Using state-of-the-art equipment in a small trilateration network with good intervisibility of monuments can yield standard deviations for the horizontal coordinates in the range of $2\,\text{mm} + 1\,\text{ppm}$. In medium size applications it has been shown that GPS, combined with terrestrial observations and careful control of the antenna eccentricities (GPS, too, has its fiducialization problems), can yield positional accuracies of about $2\,\text{mm}$ [10]. In large projects, like the SSC or LEP, positional accuracies of about $10\,\text{mm}$ can be achieved using two-frequency GPS receivers. Trigonometric and differential levelling are the only accurate methods to determine elevations; both methods yield the same accuracies—approximately $1\,\text{mm}$ for networks smaller than $2\,\text{km}$, and $20\,\text{mm}$ for a SSC size network.

![Fig. 31 Tunnel network](image)

**Tunnel** Tunnel networks are usually long and narrow (see Fig. 31), and incorporate points beneath the shafts as connections to the surface net. The monument system can be 2-D (horizontal only) or 3-D: common designs are the SLAC 2-D marks, the DESY-HERA 3-D reference cups or the standard 1.5-inch floor cups and magnet mounts (see Fig. 11 in Section 2.3.2). Some kind of tripod or column-like monopod is used for the instrument set-up. The SLAC set-up (see Fig. 32) is designed to accommodate slopes of up to $15^\circ$; the HERA design is more optimised towards efficiency, virtually eliminating the task of centering instruments and targets over monuments (see Fig. 33) [11]. The elevation of the instrument above the 3-D reference cup is known very accurately, which facilitates 3-D mapping with theodolites. Both set-up types are forced centered over known points. Laser tracker and the more recent generation of Total Stations allow the efficient use of the free-stationing method (see Fig. 34).

![Fig. 32 SLAC instrument set-up](image) ![Fig. 33 HERA instrument set-up](image) ![Fig. 34 Free-stationed laser tracker](image)

4. **FIDUCIALIZATION**

Fiducialization is a fancy name for relating the effective electromagnetic axes of components to some kind of mark which can be seen or touched by instruments. The beam, influenced only by the electromagnetic field of a component, knows nothing about fiducials.
It is therefore of at least the same importance to accurately relate the magnetic axis to the fiducial marks as to correctly position the marks to their nominal coordinates. The term fiducialization is commonly also used to describe the task of relating the axis of mechanical components to their marks. These marks are then aligned to the nominal positions calculated with the information described in the previous chapters.

Fiducialization is a two step process: firstly, the axis of the component is determined, and secondly, the axis position is related to the fiducials.

4.1 Determination of centreline

Magnets in accelerator beam lines have, for the most part, been made with ferromagnetic poles, and traditionally these pole surfaces have been used as the reference for external alignment fiducials. This practice assumes that the magnetic field is well defined by the poles. However, this fails in the presence of saturation and in the case of superconducting magnets, which have no tangible poles. There are other well-known difficulties: the poles of an iron magnet are never perfectly flat or parallel. Where is then the magnetic mid-plane [12]? The equivalent problem for quadrupoles or sextupoles is that there is no unique inscribed circle that is tangent to more than three of these poles; this makes it quite difficult to describe where the centreline really is.

While without a doubt the electro-magnetic determination of the centreline is the most accurate method, quite often, budgetary, time, or other constraints make it necessary to rely on mechanical means.

4.1.1 Mechanical representation of axis

*Design dimensions* Magnets are often constructed from stacked laminations, what guarantees very repeatable dimensions. This permits the design of features into the shape of the laminations, which have a known and accurate relationship to the mechanical axis. Most commonly, these features have the shape of grooves or edges (see Fig. 2 above). As already described in Section 2.1.1, these features can be used to reference fixtures with targets, which then in turn will represent the centreline.

*Bore target and mandrel* If magnets could be machined perfectly, a bore target (see Fig. 35) in the shape of a short cylinder with the nominal bore diameter would touch all four poles of a quadrupole. Since these required tolerances can hardly be achieved with affordable fabrication methods, the target is likely to touch only three of the poles. This method is therefore limited to lower accuracy applications, unless the typical method is modified. This improvement is accomplished by moving the target in the bore such that the target sequentially references different sets of poles. These four target positions, in case of a quadrupole, describe a circle. The centre of this circle references the mechanical axis.

Fig. 35 Bore target
A mandrel (see Fig. 36) is a combination of two bore targets on a common mechanical axis. Inserted into a quadrupole, it references and represents the mechanical axis. But, as before with the bore target, fabrication limitations will cause the mandrel only to touch three poles at two longitudinal locations. Moving the mandrel in the same manner as the bore target will alleviate the problem.

Scanning of poles  The bore target and mandrel will always pick-up the highest point of a pole. Therefore, stacking imperfections or remaining burrs will bias the measurement. This can be avoided by scanning the cross section of each pole in combination with a shape least squares fit.
(see Fig. 37) [13]. In most cases circle shape fits are sufficient, otherwise hyperbolic or parabolic shapes can be fitted. A subsequent circle fit through the centres or focal points of the previous pole shape fits will reference the axis location at the cross section longitude. If this process is repeated at more than two cross sections, a line fit will further reduce determination errors. The scanning operation is best accomplished on a Coordinate Measurement Machine (CMM); however, if the magnet’s dimensions or weight exceed the available measurement or loading capacity, theodolite measurement system or laser tracker measurements will suffice.

4.1.2 Electro-magnetic axis determination

The inherent problems of approximating the magnetic axis of a component with its mechanically determined centreline can be avoided by direct magnetic measurements. Several methods are available using various properties of the magnetic field. Properties commonly used are the magnetic field’s symmetries and its lack of radial field in the centre [14]. The following is only intended as a listing of applicable methods [15], technical details are given elsewhere.

Rotating coil A multipole magnet field will induce a voltage, which is cancelled by the symmetry of the coil. The output voltage of the coil is proportional to the magnitude of the dipole field, which is only a function of distance from the magnetic centre.

Vibrating wire A light wire is tensioned by a weight at one end such that its lowest resonant frequency is around 50 Hz. Shaken in either the vertical or horizontal plane, the voltage induced in a loop containing the wire is observed, and minimised by moving its ends. At minimum first harmonic output, the wire will be along a family of lines, all of whom intersect at the nodal point of the magnet [16].

Taut wire This method uses the property of zero field at the centre. After the magnet is energised, a current is passed through the wire. If the wire is not at the magnetic centre, a deflection will be observed. The wire position is then corrected, and the process repeated until no deflection of the wire is observed when the wire current is turned on [17].

Ferrofluidic cell Ferrofluidic cells are a means to make the virtual magnetic axis visible. The physical mechanism of this method is based on the scattering of polarised light on aligned colloidal particles in multipole fields [18]. A target with the fluid is placed in the bore. White plane-polarised light is shone through the solution from one end of the magnet. From the other end, an observer looks at the target through a plane-polarising analyser. The analyser is aligned with the polarise of the incoming light such that complete cancellation of light should occur when the magnet is turned off. With magnetic field, complete cancellation does not occur except along two perpendicular axes, which cross at the magnetic centre of the quadrupole [19].

4.1.3 Electrical axis

Some diagnostic instruments need to be accurately positioned. E.g., the relationship between a beam position monitor (BPM) and its related quadrupole/ sextupole often needs to be known to high precision. This requires fiducializing the BPM accurately. To determine the
electrical centre of a BPM, a fast pulse is sent down a wire stretched through the BPM, and its position is sensed with the same electronics used to measure the beam position [20].

4.2 Relating axis to fiducials

The above-described methods will visualise the mechanical or magnetic axis by optical targets or by a wire. While optical instruments can acquire targets directly, a wire-represented axis requires an additional transfer to marks, which can be acquired by survey instrumentation.

4.2.1 Referencing a wire

Jigs are commonly used to reference a wire to physical marks. The wire can be related to a jig using capacitive sensors, laser scanning [21], or microscopes [22]. These measurements combined with the jig’s dimensions will reference the wire position to the jig’s fiducials.

4.2.2 Transfer of axis to component fiducials

*Fixturing*  An example for this method is the Danfysik measurement stand (see Fig. 38, 39), which is a commercial version of a measurement bench developed for LEP at CERN [23]. As a result of the measurement process, two CERN socket-type fiducials are mechanically adjusted to be precisely in the same vertical plane as the magnetic axes. The measurement bench automatically centres a quadrupole onto its rotating coil, which represents the magnet’s magnetic axis. A fixture positions a laser beam above the magnet in exactly the vertical plane of the coil at a given vertical offset. The laser acts on a quadrant detector, which is mounted inside a 3.5-inch sphere, which in turn is placed sequentially into the two CERN socket base plates. Each base plate is now manually adjusted such that the quadrant detector readings become zero. The z-coordinate of the base plates is provided by mounting references.

![Fig. 38 Danfysik rotating coil bench](image)

![Fig. 39 Side view of bench. Rotating coil assembly consisting of measurement cylinder (1), air bearings (2), DC motor (3), and angular encoder (4). Magnet positioning system consisting of magnet platform, horizontal movement gears and motors (5), vertical movement gears and motors (6). Alignment system consisting of laser (7), photo detector with Taylor Hobson ball (8), and calibration supports (9).](image)
**Optical tooling**  Optical tooling is the standard method at SLAC to reference a bore target or rotating coil to the magnet’s fiducials (see Fig. 40 – 42) [24]. The magnet is first levelled using observations to the split planes or to clean laminations. A transit is then “bucked-in” parallel to the magnet by measuring offsets to the iron or laminations, representing the magnet’s z-coordinate axis. A second transit is set-up parallel to the line-of-sight of first transit such that the range of its telescope micrometer includes the axis of the magnet. The offset amount is determined by simultaneously reading a common scale-bar with bothtransits. The second transit’s crosshairs are brought into coincidence with the bore or rotating coil target mark by adjusting the parallel plate micrometer. Subsequently, the micrometer is read, and the reading is added to the transit offset value. As a result, the horizontal offset between the master transit and the magnet’s axis is known. This value is transferred onto the fiducials by touching these fiducials with scales held perpendicular to the line-of-sight of the master transit and then reading the individual fiducials’ offsets with the master transit. Adding these offsets to the previously determined axis-transit offset will reference the fiducials to the axis. To determine the z-coordinate of the fiducials, a third transit is set-up perpendicular to the line-of-sight of the master transit by collimation of its trunion mirror. Scale readings perpendicular to the line-of-sight of the third transit will yield z-offset of the fiducials. To reference these z-offsets to the magnet iron, readings to the front and back face of the magnet are also taken. The vertical component is determined by standard levelling techniques. Combining all these values produces fiducial coordinates in reference to the axis.

**Theodolite Measurement Systems (TMS)**  Measuring horizontal and vertical directions from at least two stations to the same target, after the relative orientation of the stations has been determined, is sufficient information to calculate 3-D coordinates of that target (see Fig. 43). The model volume can be extended by adding more theodolites or by moving one theodolite to a new station. However, a minimum of three common points needs to be measured from each station. Following these rules, it is straightforward to measure the fiducial and the axis targets coordinates. TMSs are significantly more efficient and tend to produce more reliable coordinates than optical tooling.

![Fig. 40](image1.png)  Optical tooling from tooling bars

![Fig. 41](image2.png)  Optical instruments and bore target
Laser trackers (see Fig. 44) [25] and total stations measure not only horizontal and vertical directions, but also distances with respect to the same origin. This fact allows one to calculate 3-D coordinates using measurements from only one station. Polarly determined coordinates with laser trackers are in the same accuracy domain as TMS results. Total station (TC2002, TDM5000) based measurements are somewhat less accurate. However, the polar technique can provide an up to 300% productivity advantage over TMS procedures.

Coordinate Measurement Machine (CMM) measurements CMMs can address targets either optically or mechanically. They are available in a wide range of measurement volumes and accuracy capabilities. Because CMMs measure coordinate differences directly, they represent the most efficient approach and can also reach much higher accuracies (see Fig. 45).
5. SUPERCONDUCTING MAGNET FIDUCIALIZATION MONITORING

Superconducting magnets don’t have poles like a traditional warm magnet (see Fig. 46). The field is generated by a coil, which is maintained by collars and located at the centre of an iron yoke. Since there is very little iron, the axis of the field is determined by the position and shape of the coil, which itself is not geometrically stable [26]. It is therefore not possible to fiducialize the magnet in reference to the mechanical axis of the yoke. The coil and yoke, referred to as the cold mass, are insulated from the ambient temperature by heat shields. This assembly is inserted into a steel cylinder, which also provides the structural support. The situation is compounded by the fact that the cold mass is submitted to high temperature differentials during cool down. The cold mass supports need to provide stable support while at the same time permitting a longitudinal expansion/contraction in the order of several centimetre. This not only makes the cryostat the cold mass inaccessible, but the above design criteria make the cold mass to cryostat relationship mechanically unstable. Consequently, “we have to align an object that we can neither see nor touch directly” [27].

Because of these boundary conditions, precise fiducialization of superconducting magnets requires a determination of their magnetic centreline with one of the above mentioned methods at superconducting temperatures. The execution of these measurements is more involved than the axis determination of a warm magnet. Firstly, a superconducting magnet can only be cooled down if the vacuum chamber is under vacuum. Since it would be too difficult to install the magnetic measurement equipment inside a vacuum vessel, a secondary pipe is usually inserted into the vacuum chamber and sealed vacuum tight to the chamber. Secondly, since the cold mass is inaccessible, fiducials can only be attached to the cryostat, which does not have a stable relationship to the cold mass. Therefore, the fiducialization must be supported by a determination of the possible variations in the cold mass to cryostat relationship.
5.1 LHC magnets

Some magnets of the LHC string test set-up have been equipped with a monitoring system. Inside a cold mass support-foot a short silica rod is attached to the cold mass (see Fig. 47). The position of the rod is monitored in all degrees of freedom by capacitive sensors with respect to the cryostat. The sensors are capable of resolving μm size motion. First results indicate that the feet move as expected in the longitudinal direction, are less stable than expected in the vertical direction, and do not move in the perpendicular direction [28].

5.2 SSC

A test plan was developed to establish cold/warm relationships for SSC dipole and quadrupole magnets by making direct autocollimation and autoreflection measurements through windows to the cold mass (see Fig. 48, 49). The mechanical design was composed of a clear optical window in the vacuum vessel, an optically coated window in the 80K shield, a through hole in the 20K shield and autocollimation mirrors and targets attached to the cold mass. Four locations at two axial stations on the magnet were instrumented this way. The resulting heat leaks were deemed acceptable for R&D magnets, but would not have been accepted for production magnets. First measurements on one magnet showed the cold mass arching and moving by about 0.1 mm in the horizontal plane, a negative vertical translation of about 0.7 mm and inconclusive roll changes [29].
6. CASE STUDIES

The following are short descriptions of actual “setting reference targets” projects, which are exemplary for many future applications.

6.1 Warm magnets

6.1.1 FFTB magnets

The Final Focus Test Beam (FFTB) is a transport line designed to test both concepts and advanced technologies for application to future linear colliders. In order to achieve the desired spot sizes at the focal point ($\sigma_x = 1 \, \mu m$, $\sigma_y = 60 \, nm$) among others, demanding tolerances on alignment ($\sigma_x = 100 \, \mu m$, $\sigma_y = 60 \, \mu m$) must be met. To retain as much as possible of this tight error budget for the alignment process, a very accurate magnet fiducialization bench based on the vibrating wire technique was developed (see Fig. 50). Magnetic centre location was accomplished by placing the wire at six different small angles ($< 1$ mrad) with respect to the magnet axis, zeroing the first harmonic electromotive force signal each time. The spatial location of the wire was first detected with a modified tooling microscope. Then it was transferred from the microscope to fiducials by means of a coordinate measurement machine. The resultant six lines were used to find the point at which the average distance from the point to each of the lines was minimised. The fitting would produce a centre unconstrained by any mechanical reference to the magnet (see Fig. 51). The RMS deviation of any line from the point of closest approach was less than 4 $\mu m$ relative to the external fiducials of the magnet.
6.1.2 PEP II quadrupoles

The High Energy Ring (HER) of the PEP II project re-uses about 280 of the original PEP quadrupoles. The original PEP survey concept provided fixtures to reference the mechanical axis to fiducials. While changing the fixture design, the principle concept was retained. However, when verifying fixture functionality on test magnets, problems were encountered. An investigation traced the reference problems to dimension variations of the magnets. The quadrupoles are assembled from four quadrants that are individually stacked from laminations. Although dowel pins referenced the assembly, forces were introduced in the assembly process, which twisted the magnets sufficiently to invalidate the fixture to axis relationship. Subsequently, all 280 quadrupoles were mechanically fiducialized. A mandrel (see Fig. 52) was inserted into the bore and placed into four positions (see Fig. 53) such that all pole combinations were used. The same process was repeated with the mandrel shifted longitudinally by a small amount to eliminate burr or stacking biases. In each position the two end points of the mandrel were measured with a laser tracker (see Fig. 54). The laser tracker also swept the fiducials and floor control points from stations. Variations of up to half a millimetre were identified.

![Fig. 52 Mandrel inserted in bore](image1)

![Fig. 53 Mandrel referencing upper two poles](image2)

![Fig. 54 Fiducialization with laser tracker](image3)

6.2 Superconducting magnets

6.2.1 RHIC

A magneto-optical procedure was designed, to directly locate the null-field axis with respect to outside fiducial reference points of the magnet assemblies. The procedure has been used to determine the null axis, both when the magnet assembly was at room temperature, and
after cooling to 4° K. All of the arc CQS assemblies are measured at room temperature. A subset of these assemblies were also measured at liquid helium temperature, to obtain statistics on the null axis’ position shift caused by differential thermal contraction of magnet parts between these temperatures [30].

The magnetic measurement procedure is based on ferrofluidic cell measurements. The cell is placed in the magnet bore on a rail support and illuminated with collimated light (see Fig. 55), which passes through a polarising filter. The cell (see Fig. 56) is observed with an alignment telescope (see Fig. 57) through a polarising filter. The telescope is set-up such that its line-of-sight is close to the geometric axis. Two 3.5 inch target spheres on either side of the magnet are adjusted to the boresight. These spheres can later be observed by a TMS, which facilitates the spatial reference to external fiducials. Prior to energising the magnet, the polarises are crossed to give a sharp extinction of the light beam. When the magnet is energised, a characteristic cross shadow pattern is seen, being centrally symmetric about the magnetic axis (see Fig. 58). Vertical and horizontal components of the axis’ displacements from the previously established boresight are measured directly using the plane plate micrometers built into the alignment telescope. The procedure was automated by attaching a CCD camera to the telescope and replacing the micrometer reading with image processing [31].
6.2.2 CEBAF cavities

The CEBAF accelerators are powered by superconducting cavities. A cavity pair is mounted in a cryounit. Four cryounits are housed in one cryomodule. The internal alignment is specified to 0.5 mm in the horizontal and vertical plane and 0.5 mrad roll. After the cavities leave the chemical preparation, they are assembled on an alignment stand into pairs inside a class 100 clean room (see Fig. 59). Subsequently, four pairs are mounted in a helium vessel within an insulated dewar flask. To maintain the relative pair alignment, the cavity pair with its alignment fixture is slid on bearings riding on precision rods through the open vessel. Subsequently, the cavities are connected to the helium vessel, which after removal of the alignment fixture retains the pair alignment. The beam pipe is installed which will become the fiducial. Afterwards, the magnetic and thermal shielding is added, and then the unit is inserted into the vacuum vessel. The vacuum vessel is supported through a pair of Thompson rods and bearings. Nitronic rods are installed which support and axially restrain the helium vessel inside the vacuum vessel. Adjustments of the rods assist in keeping the mounted pair the cryounit.

Alignment has now been transferred outside the vessel to the two beam flanges and is maintained by the fixturing of the Thompson rail and bearings. The integration of four cryounits into one cryomodule is carried out on a precision assembly bench. From a line parallel to the axis of the module, offsets are measured to the cryounits’ flanges. Instead of reading scales, special offset bars of defined length with crosshair targets at the end are used (see Fig. 60). When all cross hairs line up on the reference line, the cryounits are in good relative alignment. Typical alignment is reported to be better than 0.25 mm and 1.25 mrad [32].

![Fig. 59 Assembling two cavities into one cryounit](image_url)
![Fig. 60 Aligning the cryounits in a cryomodule](image_url)

6.2.3 HERA magnets

The HERA superconducting magnets have been equipped with two active fiducials each, i.e. so that not only targets but also instruments can be referenced to their axes (see Figs. 61, 62). The fabrication dimensions are checked on an optical test stand. On either side of the magnet two columns are set-up. Each column carries two optical tooling base plates with a horizontal spacing equivalent to the distance between the nominal magnet axis and the fiducial. Two alignment telescopes are set-up on one side (see Fig. 63) looking at Taylor Hobson spheres on the opposite column. Then the magnet’s fiducials are aligned to the line defined by the one telescope and its target, and at the same time, the magnet’s roll is adjusted. The second telescope target combination now represents the nominal axis of the magnet. Sliding a self-centering target through the bore, horizontal and vertical deviations can be
measured (see Fig. 64). A stretched wire system is afterwards used to confirm the magnet axis’ position in a cooled down condition. There, the wire replaces the boresight equipment. By measuring the spatial relationship of the wire and the fiducial line-of-sight, the magnetic axis is referenced to the fiducials.

7. CONCLUSION

It should have become apparent that alignment targets are an intrinsic part of any accelerator. Even small machines cannot be installed and maintained without some kind of reference to the design coordinate system and the path of the particles. The choice of the targeting is directly correlated to positioning methodology and instrumentation, and therefore lastly to alignment cost. May our magnets always be aligned!
ACKNOWLEDGEMENTS

I wish to acknowledge the contributions of many friends and colleagues, in particular: Horst Friedsam –APS, Frank Karl – BNL, Franz Löffler – DESY, and Will Oren – CEBAF.

REFERENCES


[24] Basically, transits are used to set-up a rectangular coordinate system. This is facilitated by auto-collimating a second transit to a mirror which is mounted on the trunion axis of the master transit truly perpendicular to its optical axis. Scales, referencing a feature and held perpendicular to a line-of-sight, are read with the assistance of the plane plate micrometer. A transit in its geometry is very similar to a theodolite. However, an optical tooling transit usually does not have angle measurement capabilities. Its telescope is tuned to the predominantly short line-of-sights. Accordingly, it provides excellent line-of-sight stability and has a built in plane plate micrometer.


IMPLEMENTATION AND MAINTENANCE OF THE ALIGNMENT OF ACCELERATORS (PRESENTED AS İSINKING AND AGEINGİ)

Michel Mayoud
CERN, Geneva, Switzerland

Abstract
The precise alignment of the components of particle accelerators, and of their experimental equipment, requires special techniques derived from geodesy. The increasing size of these machines and the demand for tighter and tighter tolerances have led to the development of special instruments and methods. The present needs, in terms of relative accuracy along the beam lines, are around 0.1 mm (rms)—and they are fully satisfied by this geodetic metrology. But nothing is absolutely stable and perfectly rigid, neither the ground nor the structures: various geomechanical forces and progressive changes of mechanical properties in some material affect the positioning of the components (hence the subtitle), and the alignment is to be maintained regularly. This paper describes the basic concepts and techniques used for that purpose.

1. INTRODUCTION

Nothing is absolutely stable and perfectly rigid, everything can be moved and/or strained and distorted by forces. All human constructions are affected, even our planet itself is subjected to various parasitic movements, random constraints or cyclic deformations. Accelerators, like other objects inserted in the upper layers of the earth’s crust, are moved up and down, expanded or shrunk and distorted while getting older.

At the last stage of their installation, the most critical components are positioned (aligned) to within 0.1 mm (rms), and then the geometry is progressively altered and must be restored. Even not knowing why it is moving, one cannot keep ignoring by how much: according to their related effects on the circulating beams, all significant movements have to be detected and corrected. This metrology calls for special techniques, coming from geodesy.

Geodesy is an old applied science, devoted to the measurement of the earth and to "positioning" problems in general, and whose means are also used for the measurement of some large to huge objects—such as dams, bridges, industrial or scientific equipment, etc.—and such applications are often referred to as "micro-geodesy". As far as high precision is concerned, one can also speak of "geodetic metrology". The basic concepts and some of the special techniques used for the initial alignment of accelerators and for their maintenance surveys are presented in this paper.

2. CATALOGUE OF FORCES ACTING AGAINST MAN-MADE CONSTRUCTIONS AND MACHINERY

In addition to the general tectonics and its related catastrophes (earthquakes, volcanos, etc.), subsidence (isinking) and upheavals may be caused by geomechanical forces capable of moving and deforming constructions:

- local tectonics, oregeny (active faults and cracks);
• level of water tables, seas and lakes, position and size of glaciers - i.e. environmental load of the substratum;
• water content and pressure in the ground, locally or over the whole area (circumstantial or seasonal), in connection with the physical behaviour of materials with respect to the addition or the removal of water — i.e. hydrophily and swelling of clays, structural deformations according to pressure changes in some layers;
• compression/decompression effects due to changes in the local load and pressures (uplifts and convergence);
• specific weakness of the substratum caused by superposing underground constructions (galleries, caverns, etc.);
• thermal expansion of the ground around a heated construction (thermal convergence);
• progressive change of mechanical properties (due to ageing) in construction materials.

3. BASICS ON POSITIONING, DIMENSIONAL AND PHYSICAL GEODESY

Astronomy and geodesy have been linked since the very beginning of rational observations of the universe. Cadastral surveying appeared very soon, when establishing boundaries and taxes on land property. Astro-geodesy, as a set of precise positioning techniques, developed with the needs of navigators and cartographers. Then dimensional geodesy was a major science during the dispute about the exact shape of the earth (1735-44). Soon after, the development of specific mathematics and the improvement of measurement techniques induced questions on the physics of the earth. Physical geodesy opened a new field of investigation about gravity anomalies and their effects on measurements and data processing. Then came the era of spatial geodesy, using satellites (or even quasars) as links across or over continents. The whole geodetic science is closely connected to geophysics and other earth sciences.

What does a geodesist's toolbox contain? For global geodesy, the panoply is:
• Angular measurements: theodolites (best accuracy = 0.1 mgon / 0.3 arc second);
• Distance measurements: Electronic Distance-Meters (a few $10^{-6}$ accuracy, maximum $10^{-7}$ when using two carriers and high frequency modulation);
• Levelling: optical automatic levels plus Invar rods (accuracy <1mm per km$^{1/2}$);
• Accurate positioning: differential GPS, i.e. Global Positioning System with phase measurements on the two carriers while tracking the satellites (nearly $10^{-7}$);
• Intercontinental links: Very Long Base Interferometry (correlating signals from quasars), Satellite Laser Ranging, refined GPS;
• Mathematical reference surface and volume and systems: IUGG (International Union of Geodesy and Geophysics) or WGS (World Geodetic System) ellipsoid(s), local datum(s) and local tri-dimensional systems;
• Physical reference surfaces and volumes and data: geoid(s), mass models, gravity field and spherical harmonics, free air models for obtaining orthometric levelling data;
• Related mathematics and appropriate software.
When dealing with geodetic metrology, most of the above is still necessary (except intercontinental links) but some very accurate instruments must be added:

- Invar wires and other length measurement devices (a few 0.01 mm accuracy);
- Special instruments for mono- or bi-axial off-set measurements with respect to a stretched wire or a laser beam (a few 0.01 mm accuracy);
- Alignment telescopes and targets, rules and micrometers;
- Inclinometers ($10^{-6}$ rad) or horizontal pendulums (up to $10^{-9}$ rad);
- Hydrostatic levelling systems (up to a few µm resolution);
- Interferometric calibration baseline.

4. SURVEY AND ALIGNMENT TOLERANCES FOR ACCELERATORS

The specifications of accuracy, for alignment, are related to beam optics—i.e. to the magnetic elements of the lattice. Transverse errors in positioning are seen as imperfections of the guiding field: the particles no longer meet the theoretical magnetic field or gradient, and this creates a local perturbation of the motion—which is specially critical in focusing elements. Depending on the magnitude, location and distribution of these alignment errors, the resultant orbit may undergo deviations and oscillations of varying amplitude—with possible resonant effects in circular accelerators. Accordingly, tilt errors induce vertical distortions of the trajectory, and related tolerances must be also specified for this critical parameter.

The relative positioning of quadrupoles is therefore of major importance along particle beam lines, and this is the reason why the main criterion for precision is a relative and local one, leading to the best "smoothness" along the trajectory. To be correctly defined, this statistical criterion involves the consideration of a local trend curve, fitted to the actual data. The rigorous estimate is therefore the remaining dispersion of positions around this trend curve. Nevertheless, it can be roughly expressed as the standard deviation (rms) of the discrepancies on the radial or vertical position of each focusing magnet with respect to the adjacent ones, comparing actual sagittas to the theoretical (expected) values.

In former accelerators, with a rather large aperture, this criterion was much less critical and technically easier to manage. But with the progress of measuring techniques and the economical gains in reducing the aperture, it is commonly set to 0.1 mm—or even less for future linear colliders for which a few micrometers are sought in dynamic alignment systems.

But the absolute accuracy is not without importance: long-range errors in curvature (or straightness) may also induce oscillations of the beam and degrade the performances of the machine. For "small" accelerators—even with 200 m or 300-m diameter—it becomes confused with "relative" errors, affecting the local smoothness. One cannot neglect this correlation: there is no simplified metrology for small machines. Errors in geodetic control points will also induce deformations in the adjustment of the metrological reference network used for initial alignment. As a consequence, the best absolute geometry is also desirable for a good mastering of the whole alignment process.

When considering the technical means able to satisfy such requirements, it appears that the vertical control (height ordinates) is perfectly and easily ensured by appropriate levelling techniques. The radial control (plane co-ordinates) is much more difficult to obtain for such tight tolerances, and the complication is drastically increased by the huge size of some accelerators.
5. NETWORK STRUCTURES FOR RADIAL (HORIZONTAL) CONTROL

First of all, for each new accelerator project, the exact location is ensured by means of a surface geodetic network, which provides control points at appropriate places, via access galleries or pits to the underground infrastructures. The accuracy of this framework has, of course, a direct influence on the control of the absolute geometry of the machine to be built.

The design of the metrological control network has evolved with the size of projects and with the measuring tools used at different epochs. Various accurate instruments, commercially available, can be used for measuring angles (with theodolites) and distances (with electronic distance-meters). Measuring short or medium distances with the required accuracy can still be a problem, and the observation of misalignments around the accelerator is not a straightforward process.

Some special devices were developed at CERN for this metrology:

- the Distinvar, for length measurements from 0.4 to 55 m, using calibrated Invar wires, with an in-situ accuracy of $\sigma \approx 0.03$ mm;
- the wire offset device, for measuring the offset of a point with respect to a straight line provided by a stretched wire (Nylon, Kevlar or carbon fibre, up to 120 m), with an in situ accuracy of $\sigma \approx 0.03$ to 0.10 mm according to the wire length and observation conditions;
- the laser offset device where the reference line is a laser beam and having about the same accuracy, depending on environmental conditions.

When comparing the network structures designed in various HEP laboratories, they can be classified into three main categories:

1. Regular polygons with central point (or central kernel of points), well adapted to small or medium accelerators and ensuring a stiff control of the absolute geometry, for example the Proton Synchrotron (PS 200-m diameter) shown in Fig. 1;
2. Ring-shaped networks with a chain of large braced quadrilaterals, still ensuring a good control of the absolute geometry but without central point, as for the Intersecting Storage Rings (ISR 300-m diameter) shown in Fig. 2;
3. Ring-shaped networks with either a narrow chain of quadrilaterals or a simple polygonal contour, maintained by a few control points at access areas, as in the Super Proton Synchrotron (SPS 7-km circumference) or in the Large Electron-positron Collider (LEP 27-km circumference) shown in Figs. 3 and 4. In such cases, the span between control points is 1100 m (SPS) or 3330 m (LEP), and it becomes more and more difficult to master the flexibility of such arcs.

The PS network was measured with accurate distances (Invar) and angles, and the positioning of magnets was first ensured by the same means, producing non-homogeneous radial-error ellipses (from 0.2 to 0.6 mm rms). Using offset measurements all around the ring, instead of polar measurements, brought a significant improvement (0.2 ñ 0.3 mm rms) on magnets.

The ISR network was a pure trilateration network, well structured and accurately measured with the Distinvar, which produced very good results, despite the low redundancy of its design. The radial error, along a diameter, was only 0.4 mm rms.

However it is worthwhile concentrating on some critical aspects of large control networks. The size of projects may quickly reach a point where geodetic parameters must be rigorously included in the positioning data: the earth is an imperfect ellipsoidal volume, and a Cartesian geometry does not fit directly with spherical co-ordinate systems, altered by gravity
anomalies. In addition, these flexible quasi-linear networks may have a stochastic "behaviour" which raises many problems at different stages of their initial measurement, and later on with their successive use for the metrology of the actual object to be aligned and maintained, i.e. the accelerator itself.

Fig. 1 Proton Synchrotron network
6. VERTICAL CONTROL

As mentioned, vertical and horizontal (radial) control is often processed differently and separately - except in 3-D triangulation/trilateration blocks, mainly used for the metrology of large experiments.

Vertical control is much simpler: it makes use of the well known technique of geometrical levelling, using optical levels and measuring height differences between successive points - thus forming traverses, loops when coming back to a starting point and finally a network when connecting loops. For reliability and quality, multiple readings plus forward and backward measurements provide redundancy, whilst loops (whenever possible) ensure a local check of the data. Care against systematic errors consists of making regular
calibrations, ensuring symmetry (for cancelling refraction and earth curvature effects), looking at temperature effects on the instruments and on the measured structures, checking stability, etc. *Levelling is a rather simple process, complicated by many good reasons to do it wrongly!*

In modern instruments, there is no more 'human' optical reading: a CCD array and an 'encoded' staff (rod) do it. Along beam lines, in tunnels, the conditions of observation are very good and the rms accuracy of 'high precision levelling' can reach 0.3 mm per km $^{1/2}$ - with a rms error not greater than 0.04 mm in the height difference between adjacent quadrupoles (every 40 m in LEP). The closing error after the 27 km loop has often been (luckily) less than 1 mm!

For rather unstable sites (like ESRF), or for sensitive quadrupoles in low $\beta$ sections (where misalignments and movements have amplified effects on the orbits), it can be interesting to set up a real-time levelling system for measuring the changes in vertical positions - combined with motorised jacks for correcting movements or resetting given positions. In the environmental conditions of accelerators, and for high precision requirements, the best technology is made of hydrostatic levelling systems (HLS) equipped with capacitive sensors (Fig. 5).

![HLS vessel on a low-\(\beta\) magnet](image)

The principle of a HLS is very simple, but one has to avoid making either a thermometer or a barometer: only one input of air in the loop, and the smallest 'column' of liquid. When temperature conditions are too much different between measurement areas, and when (in addition) the circuit of water has to snake up and down, the solution is to circulate the water through a tank sufficiently large to quickly obtain a homogeneous temperature before taking height measurements.

Other technologies are also possible for measuring the level of liquid:

- contact measurement, moving down a needle and measuring the displacement up to the 'touching' signal;
- 'parallactic' measurement, i.e. oblique beam of light (LED + optics) reflected on the surface and observed via a PSD;
- ultrasonic measurement (less accurate).

7. METHODOLOGICAL ASPECTS OF LARGE NETWORKS

The high-precision metrology of large networks demands not only that a maximum of accuracy and care be taken in the measurements but also that all relevant geodetic concepts are taken into account. This is mainly the case over large (or long) accelerators, of which tri-
dimensional geometry is necessarily defined in a local geodetic system. For instance, at CERN, some major changes were introduced into the design of the SPS (2.2-km diameter) and LEP (8.6†km diameter) control networks for this reason.

First, in both cases, the computation of the theoretical XYZ co-ordinates of the machine has involved finer and finer consideration of the geometry of the earth. For the SPS, a spherical approximation was sufficient to express the effects of the earth's curvature in computing the Z ordinates, correcting the vertical "descent" of geodetic points along the shafts or properly tilting the magnets, in order to obtain a real plane in space. With the LEP project, which partly lies under the Jura Mountains, a further step has been to determine the vertical deflections generated by gravity disturbances, and then to express the separation between a reference equipotential surface and a reference (local) ellipsoid. This knowledge provides the necessary corrective factors to convert measured altitude into ellipsoidal heights in 3-D computations, to correct the co-ordinates of bottom points from the effects of vertical deflections or to reduce the gyro (azimuth) measurements for the difference between local horizon (physical plane) and geodetic horizon (local projection plane tangent to the reference ellipsoid).

One other change in the methodology is that repetitive measurements of the SPS or LEP control networks could no longer be thought of and managed as "absolute" surveys. For such long and flexible ring-shaped figures, the variations of the co-ordinates arising from different sets of comparable measurements may have no physical meaning. As mentioned, the trajectory of a beam within an accelerator is mainly sensitive to short-range errors. Long-range errors have less effect but are not negligible. In other terms, the figure must be smooth and this smoothing concept is fundamentally involved in a particular refinement process used for the first installation of a large machine and for any new partial or global survey when a re-alignment of components is to be carried out.

Finally it is worth mentioning that certitude in any accuracy problem cannot be acquired without a thorough knowledge of the stochastic behaviour of the measured networks. Although this statement sounds self-evident, it is in reality dependent on the method of estimating the actual errors and deformations which a network may undergo as a result of the random and systematic errors in the measurements, and also on the various constraints inherent to the chosen computational models.

8. STOCHASTIC ANALYSIS AND COMPARATIVE SURVEYS

Co-ordinates of geodetic networks are calculated by a least-squares adjustment of observations. The mathematical model (1, 2 or 3D) defines either a "free" network or a "constrained" one. The variances of adjusted parameters are derived from the Variance-Covariance matrix: $Vx = s . N-1$ and 1D error bars, 2D error ellipses or 3D error ellipsoids are derived from sub-matrices of $Vx$ and their eigen vectors.

These estimates are not exhaustive, and the a posteriori variance of (groups of) observations may be altered: it depends on the redundancy and relative "strength" of each group, according to the network structure. As a complement, Monte-Carlo simulations allow an artificial generation of random and systematic errors, with controlled constraints. The effect of random (Gaussian) errors can be therefore correctly assessed, and the a posteriori variances can be re-scaled. The deformations induced by systematic errors can be also isolated and identified. The whole gives true images of the distortions really suffered by a complex network. It provides "warning lights" to watch when actual measurement are made and processed.
For small accelerators, successive surveys must be processed in "free" network adjustment and then superimposed (for comparisons) in a congruent transform: 2D or 3D Helmert transform with respect to theoretical co-ordinates. For large ones, the apparent flexibility of the arcs makes that only local comparisons are valid - after removal of non-significant differences.

Along quasi-linear traverses or networks, correlation is locally good but becomes very poor between remote points. The solution of the least-square adjustment produces therefore a rather ill-conditioned system. When repeating such measurements, the stochastic process of the random cumulating of normal (Gaussian) errors gives different profiles, with no physical meaning of the differences. Within the envelope of Gaussian errors, all "wrong" lines have the same likelihood to be true—if no systematic errors alter the data, adding parasitic distortions.

The key problem is therefore to remove—analytically—these "apparent" differences of stochastic nature - thus making the true ones appearing really, as signals of significant movements. This must be made by trend curve analysis, completed with appropriate geometrical comparisons and relevant statistical tests (checking the signal/noise ratio). Applied to accelerators with the correction of the misalignments in order to reduce the dispersion, this process is referred to as ismoothing.

9. RADIAL/VERTICAL SMOOTHING IN LARGE ACCELERATORS

When installing the machine components, the first determination of the control network gives the displacement vectors between their actual rough position and their theoretical one. As explained above, magnets are finally positioned around an unknown mean trend curve (one among an infinity) contained within the envelope of maximum errors. The polynomial degree of the curve depends on the redundancy, the overlap of measurements, and the bridge distance between control points.

The final relative errors are a quadratic combination of those of the network itself and those of the positioning, i.e. installation errors. Their statistical nature is essentially Gaussian: the aligned elements are randomly and normally distributed around this mean trend curve (Fig. 6).

As the major requirement for the geometry of an accelerator is that the relative errors must be very small (σ ~ 0.1 mm), a compulsory step is to check the installation by measuring and—if needed—improving the smoothness of the initial alignment. Then, during the exploitation, when making successive maintenance surveys of these long and flexible figures, absolute comparisons would be a nonsense and the differences between trend curves, corresponding to each survey, must be analytically eliminated. The state of the alignment is expressed by the statistical dispersion (rms) observed around the mean trend curve—after removal of the biases (large bumps and hollows) and using preferably the same algorithm in these successive comparisons. To restore the alignment to the required state, outliers have to
be corrected and the global scattering has to be reduced up to the desired degree of smoothness.

The smoothing process consists of a set of radial or vertical measurements. For the radial case, it must be said that measurements cannot have the same span (and redundancy) as those of the metrological network, and the figure obtained—in co-ordinates—is less good, and more flexible. In any case, and for successive measurements as well, the problem is the difference between these distorted curves and the theoretical geometry. Each "image" of the ideal line has the same likelihood of being "true", within the envelope of errors, but is nevertheless different. This difference has (globally) no physical meaning, but local discrepancies or distortions may be the signal of a move, either for a single element or for a group, depending on the deformations of the supporting structure (floor and tunnel) due to geomechanical forces, and/or to some constraints along the machine (vacuum, dissipated energy, etc.).

Different concepts and methods have been tested for smoothing. The first one, used for checking the initial alignment, was a non-parametric method, which gave satisfactory results for this purpose but was not well adapted to the detection of movements. It even made us blind to some deformations! Parametric methods were then used with care.

As a matter of fact, polynomials fitted over such long lines and curves may induce correlations and constraints, which alter the image of the trend curve. Fourier analysis has also some drawbacks when choosing arbitrarily a given harmonic as the "best fit" and leaving cyclic errors around. Spline functions are rather heavy to handle and only piece-wise functions (attached arcs of polynomials, kept at a low degree) proved to be a realistic solution for a while.

Finally, a very satisfying method has been found. It consists in doing successive least-square fits of low-degree polynomials in a sliding window, shifted by steps all along the set of data. This concept allows one to retain the continuity of the trend curve—geometrically and statistically as well. It can be compared to a carpenter's plane used for smoothing an irregular plank: depending on the size of the tool and on the adjustment of its blade, one can obtain different qualities of planing with more or less waves on the wood.

In our algorithmic concept, abnormal offsets are located, and then removed if significant. This significance depends always on a signal/noise ratio at a given confidence level, and also to the choice of a given operational level. The adjustment of the blade is therefore a threshold, which defines a bandwidth beyond which data (i.e. actual offsets) are considered for a displacement back to the trend curve. The size of the sliding window can be fixed according to the desired degree of smoothing. Several iterations can be made, changing the parameters over the whole curve or in a given area when local discrepancies are observed at a higher degree of periodicity and have to be corrected more precisely. Such a method is very reliable, and flexible. Candidates for correcting movements are well identified, offset discrepancies are well quantified, and the whole work can be optimised according to the degree of "perfection" required.

CONCLUSIONS

It has been shown that nothing is stable and rigid, hence that everything is somehow floating—but not dramatically sinking—on the upper layers of the earth's crust, accelerators included. Nobody can prevent the evil but it must be now clear that everybody can cure it. Regular tests of geodetic metrology are necessary for the diagnosis, smoothing realignments make the potion for recovery, and particle orbits constitute the ultimate check on the good health of the geometry.
BIBLIOGRAPHY


J. Gervaise, Applied Geodesy for CERN Accelerators (Part I), M. Mayoud, Computing and Analysis Methods (Part II), Seminar on High Precision Geodetic Measurements, Facolta di Ingeneria di Bologna, 16-17 October 1984, Bologna, Italy

CERN Accelerator School on Applied Geodesy for Particle Accelerators, 14-18 April 1986, CERN, 87-01

First International Workshop on Accelerator Alignment, 31 July - 2 August 1989, SLAC, Stanford, USA

Second International Workshop on Accelerator Alignment, 10-12 September 1990, DESY, Hamburg, Germany

W. Coosemans, MÊtrologie pour un Grand Collisionneur LinÈaire, XVII Colloque de l'AFT, June 1991, CERN, Geneva, Switzerland

J. Gervaise, M. Mayoud, Geodesy for Particle Accelerators, Advances of Accelerator Physics and Technologies - Chapter 16, Editor : H. Schopper, Advanced Series on Directions in High Energy Physics - Vol. 12, World Scientific

Third International Workshop on Accelerator Alignment, 28 September - 1 October 1993, CERN, Geneva, Switzerland
ALIGNMENT OF EXPERIMENTS

M.J. Price
CERN, Geneva, Switzerland.

Abstract
Current methods of aligning detectors inside the planned LHC experiments are reviewed. Examples of how these methods will be used are discussed.

1. INTRODUCTION

The past hundred years have seen tremendous advances in particle accelerator technology. The earliest terrestrial accelerator was arguably the cathode ray tube which enabled J.J. Thomson to determine the ratio of the charge and mass of the electron. The Large Hadron Collider (LHC), scheduled for completion in 2005, takes the collision energy to unprecedented values. Protons circulate in opposite directions in the 28 km circumference ring and, on collision, will produce a centre-of-mass energy of 22 TeV.

Particle energy has risen dramatically since the invention of circulating beam accelerators; fixed target equivalent beam energies have risen about 8 orders of magnitude in the last 50 years. The impetus behind this ever-increasing progress is clear. According to Heisenberg’s uncertainty principle, to “see” details of the order of $x$ — and hence probe deeper into the structure of matter — a momentum of the order of $h/x$ is necessary, $h$ being Plank’s constant. The increase in maximum energy and complexity of particle accelerators, driven as it was by the needs of high energy physics, has been accompanied by a no-less impressive increase in the sophistication of particle physics experiments. The higher the energy of the colliding particles, the higher the energy of the debris of produced particles. A particle with high energy is more penetrating, i.e. more material has to be put in its way to stop it. Experimenters aim to contain as many of the emerging particles as possible, in order to thoroughly study their properties. This simple fact means that the experiments tend to become larger with increasing primary particle energy.

The properties of these out-going particles must be studied in the finest detail in order to understand what happened during the interaction. Ideally, the nature and momentum of each of the emerging particles should be measured. In practice, this is seldom possible and experiments are inevitably the result of compromise. The momentum of charged particles can be measured by observing their curvature in a strong magnetic field. The nature and energy of electrons, positrons, $\gamma$ rays and $\pi$ mesons can be inferred by their behaviour when interacting with the so-called electromagnetic calorimeters. (As the name implies, “calorimeters” measure particle energy.) In a similar way, the characteristics of the hadrons (protons, neutrons) can be measured by their behaviour in “hadron calorimeters”. $\mu$ mesons (muons) do not interact readily with matter and, after all the other particles coming from the interaction have been absorbed, it is inevitably the muons that remain. For this reason, modern experiments are massive with outer layers of detectors specifically designed to observe these highly penetrating particles. Because of these different requirements, modern particle physics experiments contain a large number of different detectors.

In order to measure of the physical characteristics of particles in these experiments the geometrical elements of the experiment have to be known with respect to each other with a high degree of precision. In particular we need to know

(i) the position of individual elements within a detector;
(ii) the position of the various detectors with respect to each other;
(iii) the position of the experiment with respect to the beam.
The methods by which the first two of these are determined is the subject of this paper. This bears the common name of “alignment”. In fact, this is something of a misnomer, implying as it does some active mechanism to bring misaligned elements back into place. One of the LEP experiments, L3, actually employs this method. There are no plans to do this in the LHC experiments. Instead, detector elements will be monitored and the necessary corrections made in the reconstruction programs.

In this paper I shall mainly concentrate on tools and systems to be used for the alignment of the two larger LHC experiments. Because of the large number and momenta of the particles produced at these experiments, they present greater challenges than previous generations of collider experiment. After describing the physics motivation for the LHC experiments, I discuss the positioning accuracies required. In the following section I describe some of the monitoring tools. In the final section I describe monitoring of the inner detectors of the two larger LHC experiments.

2. PHYSICS AND DETECTORS

2.1 The physics case for LHC

Particle physicists have been extremely successful in building up a model of the fundamental particles. The 50's and 60's saw an uncomfortable proliferation in the number of “fundamental” particles, however, theoretical and experimental work soon revealed an underlying order. Key experiments in the '70s determined the existence of the quarks. The fundamental theory of the quarks, “Quantum Chromodynamics” was inspired by the highly successful quantum electrodynamics. These theories, together with the Glashow-Weinberg-Salam theory of weak interactions, provide the theoretical underpinning of the so-called Standard Model. In this theory there are three families of particles. Each family consists of a massive and massless lepton and two quarks. Each family is accompanied by a family of anti-particles. In addition to the three families of particle there exist the “binding” particles, the photon, the intermediate vector bosons (Z°, W⁺/W⁻) and the gluons. In addition, the Weinberg-Glashow-Salam theory calls for at least one “Higgs” boson to explain the masses of the fundamental particles. Once produced, this particle will decay rapidly giving rise to charged by-products that can have momenta of 100s of GeV/c.

The discovery and subsequent investigation of the elusive Higgs particle is the main motivation for building the LHC, given this particle’s important theoretical significance. However, there are other fundamental problems to be studied. The CMS (standing for Compact Muon Solenoid) experiment and ATLAS (A Toroidal LHC ApparatuS) are the two larger, general purpose, experiments planned to run at the start-up of the LHC. As well as looking for the Higg’s particle, these experiments will also investigate possible top quark decays and search for supersymmetric particles. In addition, there are two smaller experiments (everything is relative since these, in themselves, are as large as LEP experiments). Alice will investigate heavy ion collisions and LHCB will explore CP violation in beauty particles.

2.2 Detector fundamentals

2.2.1 Detector varieties

One of the principle problems in designing the LHC detectors is the vast quantity of data that has to be handled. During collisions many secondary particles are produced. The vast majority of interactions are uninteresting, however, the detectors have to be capable of recording all the information in order to retain the interesting events. Events will be characterised by a large amount of extraneous data out of which high momenta particles will have to be sifted.

We have seen that, very broadly speaking, detectors can be divided into two classes: calorimeters and tracking detectors. Although calorimeters are fundamental to the
performance of an experiment, they do not need to be so precisely positioned as the tracking detectors. For this reason, we shall only consider the alignment of the tracking detectors. We recall that one of the principle tasks of these detectors is to measure the curvature of rapidly moving charged particles as they through the strong magnetic field. A necessary additional quality is to have sufficient resolution to be able to distinguish between the very many simultaneously passing particles.

In ATLAS and CMS there are tracking detectors both close to the interaction region and far away. The innermost trackers have the dual role of measuring the momentum of all types of emerging charged particles and of finding the topology of the interaction. To perform this second task they must have sufficient accuracy to point back to the interaction region to describe the geometry of the interaction. There are very high demands on the precision of these detectors. Towards the outside of the experiment there are tracking detectors that measure the momentum of the remaining muons. Although these devices do not have to describe the interaction geometry, there are still extremely stringent demands on their positional accuracy.

2.2.2 Intra- and inter-detector alignment

The reader will by now be aware that particle physics experiments consist of a large number of sub-units (detectors) and that each of these detectors itself contains many detecting elements. In turn, these detecting elements may either consist of a single element or be further subdivided. Let us take, as a concrete example, the barrel silicon tracker of the CMS experiment. The purpose of this device is to measure the momentum and vertex geometry of those particles emerging from the interaction region and making a large angle (greater than about 45°) with the beam direction. To do this effectively the detecting strips are parallel to the beam axis. The device consists of seven “wheels” that are distributed along the beam axis. Each wheel, in turn, is made up of seven layers of detecting elements. The detecting elements are wafers of silicon, each 12.5 cm long with an active width of 5.12 cm. Going into further detail, each detecting element has 1024 sensitive strips spaced with a pitch of 50 μm. The CMS collaboration has considered that the detecting elements within one of these wheels can be positioned with sufficient precision to render continuous monitoring unnecessary. Part of the challenge to the designers of this equipment is to ensure that these detectors, once assembled, do not move. During assembly the individual detector elements are surveyed so that their relative positions are accurately known. This knowledge is entered in a data base that will subsequently be used during the measurement phase. It should be noted that the ATLAS collaboration has adopted the opposite approach. In their inner detector there is a dense network of monitored distances to follow any possible distortion of the structure.

The evaluation of the relative positions of the detectors themselves poses another type of problem. It is very difficult to imagine that the relative positions of the detectors inside a large experiment remain unchanged. For example, although care is taken to use materials that have low coefficients of thermal expansion, there are inevitable temperature differences. At the very low relative displacements that we are considering humidity effects are also important. The magnetic field strengths in the LHC experiment are enormous. Centimetres of movement are expected in certain regions when the fields are changed. The displacements within the tracker will certainly be much less than this, however, there could be a measurable influence. During the rest of this paper we shall mainly concentrate on methods to measure the relative positions of the detectors. It can be supposed that each detector is equipped with a set of external fiducial marks that are accurately known with respect to the detecting elements within the detector.

In fact, it should not be imagined that movements within the detector would go unnoticed. In parallel with the alignment schemes discussed in this paper, a number of alignment schemes using charged particle tracks are being devised. These methods have
proved very useful at LEP. On the other hand, to rely solely on software solutions would be very hazardous.

2.2.3 Effect of measurement accuracy on momentum resolution

A charged particle of mass \( m \), charge \( e \) and velocity \( v \) moving in a magnetic field \( B \) will move in a circular orbit of radius of curvature \( r \) where

\[
eBv = m \frac{v^2}{r}
\]

(1)

This expression simply equates the centripedal acceleration to the Lorenz force. This is more conventionally expressed as

\[
r = \frac{p}{eB}
\]

(2)

where the radius \( r \) is in metres, the momentum, \( p \), in eV/c, the magnetic flux density \( B \) in T and the velocity of light \( c \) in metres per second. In the CMS experiment, with its 4 T field, the track of a proton with momentum of 1 GeV/c will have a radius of curvature of 0.83 m.

In the large LHC detectors the trackers consist of several layers of detecting elements. Typically these detecting elements contain sensitive strips that record the passage of the through-passing charged particle as a “hit” on one of these strips. Thus the record of the passage of the particle through the tracker will be a set of numbers, where each number corresponds to the index number of the hit strip in the detecting element. (It should be emphasised that this is a radically simplified description of the action of a tracker; in practice there are a multitude of particles passing through the detecting element and disentangling them is a major challenge to the track reconstruction experts. Another simplification that we have made is to assume that the detecting element only measures a single co-ordinate: in reality the position is often obtained by interpolating between several neighbouring strips that have been hit.)

Consider a detector that consists of \( N \) equally spaced detecting elements uniformly distributed between 0 and \( X \). The detecting elements measure the \( y \) co-ordinates, \( y_i \), of a passing charged particle track with a standard deviation of \( \sigma \). If the radius of curvature of the track’s orbit is very large then we can assume that this takes the form of a parabola:

\[
y = a + bx + \frac{c}{2} x^2
\]

where \( a \) is the intercept with the \( y \) axis when \( x = 0 \), \( b \) is the slope at \( x = 0 \) and \( c \) is the curvature of the trajectory. The quantities \( a, b, c \) can be calculated by making a least squares fit and their errors can be expressed as a function of \( N \) and \( \sigma \). In particular, it can be shown that the proportional error on the momentum measurement is:

\[
\frac{\Delta p}{p} = \frac{\sigma}{X^2} \sqrt{\frac{720}{N + 4} \frac{p}{cB}}
\]

(3)

It can be seen that the measurement error increases with increasing particle momentum and inversely as the product \( X^2 B \). In the CMS central barrel tracker, which, in fact is made up from several different detecting elements, there are 13 \( (N) \) roughly equally spaced detectors spanning about 1.2 m \( (X) \). The particles are bent in a 4 T field \( (B) \). To reach CMS’s required momentum resolution in \( \Delta p/p \) of \( \sim 0.1 \) p\( \% \), for a maximum \( p_T \) of 1 TeV/c, we find, from Eq. (3) that \( \sigma \approx 27 \) \( \mu \)m. This is, in fact, the approximate value of the weighted precision of the detecting elements. In reality there are factors, other than measurement errors, that limit the performance of the tracker; in ATLAS, for example the multiple scattering error is of order 1% at normal incidence. Again, in a real tracker the resolution of the inner-most detecting elements is greater than that of the outer ones since they are required also to give information on the topology of the interaction. However, the
important conclusion is that the positions of the detecting elements have to be known relative to each other to distances of the better of a few tens of micrometers to really exploit the high energies produced at the LHC. In addition, the relative positions of the individual detectors has to be known to this accuracy.

3. TOOLS

3.1 Straightness and distance methods

In this and following discussions we shall follow the normal practice in collider experiments and adopt a co-ordinate system with the $x$ axis in the horizontal plane and the $z$ axis along the direction of the intersecting particles. For barrel detectors, i.e. those in which the detecting strips are parallel to the $z$-axis, the $z$ co-ordinate of the measurement point is not used in calculating the transverse momentum of the particle. On the other hand, knowledge of the $z$ co-ordinate is very useful in distinguishing between the very many charged particle tracks that are present. For the forward detectors, which have radial detecting strips, the value of the $z$ co-ordinate is used in calculating the momentum. However, especially for very high momentum particles, the precision on the measurement of $z$ does not need to be any where as high as that of the $\phi$, azimuthal, co-ordinate.

This fact has led to two essentially different approaches to detector alignment. In the straightness method a set of straight lines pass close to the detectors. The distances from known points on the detectors to these lines allows the $x$ and $y$ co-ordinates of the origin of the detector, as well as its three angular rotations, to be measured in the global co-ordinate system. In this method the $z$ co-ordinate of the detector’s origin is measured by some independent means. In the distance method, as its name implies, distances between known points on the detectors are measured. The six degrees of freedom of the detectors are obtained by minimising the sum of squares of the differences between a large number of measured and calculated distances. In this method the $z$ co-ordinates of the detectors is obtained as a part of the fit. In practice a combination of these two methods is often used to determine detector position.

3.2 Rasnik

The Rasnik device [1] was developed at Nikhef laboratories in Amsterdam. The device is being considered for use in the ATLAS muon system and has been used in the Chorus experiment [2]. This device should not be confused with a simpler device used in the L3 experiment.

The basic idea behind the device is to create an image of a coded mask, illuminated by an infrared LED, on a CCD sensor by means of a lens. Deviations of the three elements from the ideal straight line, formed by the mask, the lens centre and the CCD, can be made by comparing the image of the coded mask with a reference image. In this way both the lateral movement of one of the elements can be measured as well as the relative rotation of the mask of sensor. The system is illustrated in Fig. 1.
By calculating the actual image size and comparing this with the mask size, the position of the lens along the Z-axis can be calculated. The coded screen has a unique pattern to allow for arbitrarily large displacements to be measured. A collimating lens between the LED and the mask increases the amount of light that reaches the projection lens. A diffuser is used to minimise effects of imperfections of the light-source. The standard video signal from the CCD sensor is digitised by means of a frame grabber and the signals are analysed by special reconstruction software (recent improvements in the mask and the reconstruction software are described in Ref. [3]).

The device has a quoted transverse resolution: of about 1 μm at 5.5 m. The longitudinal resolution is of the order of 30 μm.

3.3 Semi-transparent optical position sensors

These devices [4] exploit the fact that silicon is transparent in the infra-red. They will be used in the alignment of the ATLAS muon system and a modified version is also being developed for use in the ATLAS Inner Detector.

The device itself consists a thin film of hydrogenated amorphous silicon deposited between two layers of indium tin oxide (ITO) on a glass substrate. The top and bottom ITO electrodes are segmented by photolithographic methods into a pair of orthogonal rows of strips forming double-sided silicon-strip photodiodes. The strip pitch in the device described is 321 μm with an inter strip-gap of 10 μm (see Fig. 2)
A beam of infra-red light from a laser source passes through the detectors and its position is found using a centre-of-gravity method. Two different test devices yielded transmissions, at a wavelength of 690 nm, of better than 80% and 90%. Several detectors can be arranged in series as shown in Fig. 2. The typical combined noise level from an individual strip and the amplification electronics is of the order of 0.5%. Because of this excellent noise performance the local position resolution of the device is a fraction of a micrometre. Over the $20 \times 20$ mm$^2$ area of the tested sensors, position resolutions of the order of a few micrometres have been obtained.

With a suitable choice of glass substrate, the degradation of transmission due to irradiation can be avoided. In order to avoid deflection of the laser beam the glass used has to be of very high quality.

### 3.4 Frequency Scan Interferometry (FSI)

To measure the distance, to high accuracy, between a fixed point and a remote reflecting point, one needs to know the fringe order number (the integer giving the whole number of wavelengths in the path), the excess fraction (the additional fraction of a wavelength) and the wavelength. Unfortunately, the first of these quantities is unobtainable from conventional interferometry. The problem was first tackled by Benoit [5], using several spectral lines with the method of excess fractions. He succeeded in measuring optical path differences as large as 10 cm to obtain accuracies within a fraction of a wavelength of light. Since that time several other techniques have been developed to obtain the fringe order number by using two or more different wavelengths. See Ref. [6] for a description of some other methods.

The FSI method [7] uses tuneable lasers to obtain a continuous range of wavelengths. It is under study for use in the ATLAS Inner Detector. Laser light is sent through a delivery fibre from a station outside the experiment. The fibre ends on a known point within a detector and is beamed towards a corner cube reflector on a neighbouring known point. (A corner cube reflector has three mutually perpendicular reflecting faces: a light ray incident on one of the faces, reflects off the other two and returns along a path parallel to the incident one.) This reflected light goes along the return fibre. As illustrated in Fig. 3, the light emerging from the fibre encounters a beam splitter. At the emitting station this light combined with that from the corner cube reflector to produce an interference pattern.
Let the intensity from the corner cube reflector be $I_1$, and that from the beam splitter be $I_2$. The interference will be of the form

$$I(r) = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos(\phi_1(r) - \phi_2(r)) \quad (4)$$

Here $\phi_1$ and $\phi_2$ are the phases of the light at the spatial point $r$. The phase is defined by the quantity

$$\phi = \frac{2\pi L}{\lambda} \quad (5)$$

where $L$ is the optical path length and $\lambda$ is the wavelength of the light. Thus the argument of the cosine term in Eq. (4) contains information about the optical path difference between the two beams but tells us nothing about the fringe order number: taking the cosine of only gives us the optical path difference modulo the wavelength of the light. The trick in FSI is to vary the frequency of the light. Let $I(\nu)$ be the intensity of the entering light at a given point, $r$, when the frequency of the light from the laser is $\nu$ ($\nu = c/\lambda$) and $I(\nu + \delta\nu)$ when the frequency is slightly changed to $\nu + \delta\nu$. Subtracting the first from the second is equivalent to differentiating with respect to $\nu$

$$\frac{dI(r)}{d\nu} = -2\sqrt{I_1 I_2} \sin(\phi_1(r) - \phi_2(r)) \frac{d((\phi_1(r) - \phi_2(r))}{d\nu} \quad (6)$$

The $d/d\nu$ term on the right hand side is directly proportional to the optical path difference, and thus it is possible, in principle, to obtain this optical path difference by observing the variation of the interfering light as a function of frequency. When the errors in the system are considered, it is found that the frequency has, in fact, to be swept through a relatively large range. The number of fringes counted when the frequency is changed by $\Delta\nu$ is related the optical path difference $(D_1 - D_2)$ by:

$$\Delta N = \frac{\Delta \varphi}{2\pi} = \frac{\Delta \nu}{c} (D_1 - D_2) \quad (7)$$

In practice, the frequency scan is not totally continuous but takes the form of a series of sub-scans, each of around 30 GHz, separated by much larger gaps. These sub-scans are linked together over a total range of about one percent of the laser’s wavelength (approximately 800 nm). A laser linewidth of less that 1 MHz is required to meet the basic precision of better that 1 $\mu$m over 1 m given by the experimental requirements. In fact it is only the recent commercial availability of high performance tuneable lasers that has made FSI viable.

### 3.5 CCD based system

This technique, a straightness method, has been investigated at CERN for the alignment of the CMS tracker. A series of light sources is observed by means of a CCD camera. The images are treated to find the effective source location and the departure of the sources from an ideal straight line. Early tests [8] employed a lens producing out-of-focus images on the CCD; in focus images have to be avoided since their dimensions can be smaller than the pixel size. Preliminary tests [8] confirmed that a simple aperture can be as effective as a lens. More recent work has confirmed the advantages of this system.

#### 3.5.1 Source and aperture

In the present scheme light from a He-Ne laser is focused into a single-mode fibre (4 $\mu$m diameter). Light exits from the fibre in a narrow cone. It is observed by a camera consisting of a commercial CCD detector, with $512 \times 512$ pixels, situated some 50 mm behind a circular
aperture. Dust particles partially obscuring either the source or the aperture could give rise to errors. Thus, ideally, both the source and the aperture should have dimensions small compared to the system errors. At 4 μm diameter the source is indeed sufficiently small. On the other hand, there are difficulties in reducing the aperture diameter below a certain minimum. If the hole is too small then the diffraction pattern on the screen is correspondingly large and the amount of light collected by the CCD is too small. Another reason for not employing very small apertures is that, in the presence of nearby objects there can be variations of phase across the aperture. (This is seen in the classical case of diffraction at a straight edge.) In practice, aperture diameters of the order of 1 mm are adequate.

3.5.2 Modified centre of gravity method

A CCD camera, like all electronic devices, suffers from noise. The noise at a given pixel can be shown to have at least two components. If there is no light falling on the pixel there is still a signal, the so-called “dark current”. As the light level is increased the noise appears to be proportional to the square root of the signal height: a behaviour typical of Poisson statistics. The first type of noise is particularly dangerous since, if not corrected for, it can lead to an erroneous value of the centre of gravity.

Signals for the CCD are collected by means of a commercial frame grabber attached to an IBM compatible PC. The gain and offset of the device can be changed and the analogue signal is digitised using an 8-bit analogue to digital converter. In order to obtain the maximum precision, the gain and offset of the camera are adjusted to simultaneously give a minimum signal as small as possible but still greater than zero and a maximum signal as large as possible but less than 255. In this way the maximum dynamic range is obtained but the device is not saturated. In the algorithm for calculating the centre of gravity a maximum signal is searched for. This is taken as a first approximation to the centre of gravity and thereafter only signals in a window around this central point are considered. Signals within this window are now multiplied by a weighting factor of the form shown in Fig. 4 and the resulting centre of gravity is calculated.

![Fig. 4 Weighting function for correcting noise in the CCD signal.](image)

In the feasibility tests the function takes the form $e^{-\frac{(S_S)}{e^2}}$ for signal values less than $S_o$ and unity for larger signal values. This discriminates against low, noisy, signal values without significantly affecting the larger, information rich, signals. The question may be asked whether the multiplying function described is optimal. Preliminary work indicates that appreciable gains can be made by employing a multiplying function that is optimised to reduce the effects of the noise.

3.5.3 Resolution
Several series of tests have been carried out to test the validity of this method. In the first tests, a source situated at a given distance from the camera was moved in increments normal to the line joining the camera and source. The position of the centre of gravity of the image, as measured using the above technique, was then compared with the actual values. In this way, it was ascertained that the precision is of the order of 2-3% of the pixel dimension. In the CMS experiment, the object distance will be of the order of 3 m, the image distance is 50 mm and the pixel size is typically 10 μm. In this case, the error in the object space is of the order of 12 to 18 μm.

3.5.4 Radiation concerns

The radiation levels in the LHC experiments will be very high (of the order of several tens of MRad). Normally available CCD cameras are not designed to operate under these conditions, although it is possible that CCDs specially made for military purposes may be of use. Even if CCDs cannot be found to perform under such exacting conditions, it would be logical to use silicon particle detectors already widely used in the experiment and optimised to work in a radiation environment.

3.6 Mechanical methods

One of the traditional methods of measuring straightness is by comparison with a stretched wire. This method has been improved upon by the firm Fogale [9]. In this process, capacitative methods are used to measure the distance from a stretched wire. The wire itself is made either from uncoated carbon fibre or from carbon fibre enrobed in polyaramid fibre (Kevlar®). This method has been studied extensively by W. Coosemans, from CERN, who has tested the device for the alignment of accelerating elements in the proposed CLIC machine. According to Coosemans [10], the resolution achieved depends on the range required. For a range of some millimetres, precisions of one micrometre can be obtained. The method only gives an absolute measure of straightness for wires that are in a strictly vertical position since, in all other cases, there is an inevitable bending of the wire due to gravity. Note that even this error can be theoretically removed by applying two different stretching forces to the wire. Let the displacement (as measured against a fixed scale) of the wire at a given point be \( y_1 \) for a stretching force of \( F_1 \) and \( y_2 \) for a force of \( F_2 \). Since the sag of the wire is inversely proportional to the applied force, the displacement for an infinite force can be shown to be:

\[
y = y_1 - \frac{1}{1 - F_1/F_2} (y_1 - y_2)
\]

(8)

Although this method appears to be simple, there are many potential problems. For example, kinks in the wire can lead to errors as can any departure from a regular cross section.

4. SYSTEMS

4.1 The CMS inner tracker

The inner tracker of the CMS experiment contains sub-detectors using three different technologies. The inner-most sub-detector uses “pixels” (sensitive pads of silicon of the order of \( 100 \times 100 \mu m^2 \)). In the barrel region, there are three layers extending between mean radii of 4.33 cm and 11.0 cm. In the two end-cap regions, there are detectors at \( z = 32.5 \) cm and \( z = 46.5 \) cm. Further out, there are seven barrel layers of silicon micro-strip detectors (occupying radii between approximately 21 cm and 49 cm), and eleven end-cap silicon detector wheels on each side between \( z = 97.5 \) cm and \( z = 265 \) cm. The final, gas micro-strip detectors fill the rest of the tracker out to its nominal radius of 112 cm and total length of 550 cm. The three sub-detectors are arranged in three concentric cylinders such that there are complete lines of sight, over the full length of the tracker, between each of type of sub...
detector. Figure 5 shows a symbolic representation of one of the tracker “wheels” (it is recalled that there are three layers of such wheels, and several wheels along the length of the tracker).

So-called alignment rings are located at the extreme ends of the tracker. These alignment rings are constructed from low CTE material (e.g. carbon-carbon composite) and use a cross braced construction to obtain a highly stable structure. At the intersection of each of the struts in there is a CCD camera equipped with an aperture, as described above in Section 3.5. This aperture is surrounded by four quasi-point sources of light; the distance between adjacent sources is of the order of 30 mm. The positions of each of the apertures within the ring is accurately known (to within 5 μm). The rings are positioned mechanically such that the plane of each ring is perpendicular to the tracker axis to within a milliradian; this ensures that the error on the perpendicular distance from the axis to the centre of the apertures is less than a micrometre. In addition, the rotation of the ring about the tracker axis is monitored with a precision level. Each of the wheels is equipped with six sets of point sources that are angularly positioned to correspond to the positions of the cameras on the alignment rings. The quasi-points sources within a given set are arranged so that two of them face the camera on one ring whilst the other two face the corresponding camera on the other ring. The sources occupy a longitudinal distance of some 250 mm. The light sources on the detector wheels, although nominally on a straight line parallel to the axis, are, in practice, positioned with small offsets to prevent them from obscuring each other.

![Diagram of the CMS tracker alignment system](image)

**Fig. 5** Alignment system of the CMS tracker

A given camera on one of the alignment rings “sees” the light sources surrounding the corresponding camera on the opposite ring. Since the spatial positions of the aperture and its surrounding points are accurately known, we can materialise six reference lines passing through pairs of apertures and with known equations. Each of these lines originates and terminates at the centre of a camera aperture and, by extending the lines in both directions, we can find their intersections with the CCDs. The position of the image obtained by illuminating one of the cameras with light from one of the wheel-based sources of light can be compared with the intersection of the reference line with the CCD. In this way, assuming that the longitudinal position ($z$), of the wheel is known, the lateral position of the point source relative to the reference line can be calculated. (For example, if the co-ordinates of the image relative to the intersection of the reference line and the CCD are $Δx$, $Δy$ then the co-ordinates of the light source with respect to the reference line are $Δx \Delta d/\Delta d'$, $Δy \Delta d/\Delta d'$. Here, $d$ is the
distance from the aperture to the CCD and \( d' \) is the distance from the aperture to the source. Proceeding in this way for all six sets of point sources around a given wheel the two transverse co-ordinates (\( x \) and \( y \)) of the wheel’s origin, as well as its three angular displacements, can be found).

Simulations [11] have shown that the precision of wheel position determination using this method is more than adequate for the demands of the experiment.

### 4.2 The ATLAS SCT and pixel tracker

At the heart of the ATLAS experiment there is a pixel device made up of barrel and pixel detectors. The individual pixels measure 50 \( \mu \text{m} \) in the \( R \phi \) direction and 300 \( \mu \text{m} \) in \( z \). The pixel detector consists of three barrel layers, of radii between 4 and 14 cm and total length of about 80 cm. At each side there are four end-cap wheels situated between 40 and 110 cm. The detector is mounted on a single, light-weight, mechanical structure. The barrel is made up of detecting elements that are 6.24 cm long and 2.24 cm wide. The forward detecting elements are similar. The semiconductor tracker (SCT) surrounds the pixel detector and contains four barrel layers (between 30 and 52 cm and length of 160 cm). In addition there are end-cap detectors consisting of nine forward disks, per side, between \( z = 80 \text{ cm} \) and \( z = 280 \text{ cm} \). The sensitive strips have an average width of 80\( \mu \text{m} \). As in the pixel detector the basic building block is a single silicon detecting elements measuring, in the barrel, \( 6.36 \times 6.40 \text{ cm}^2 \). The forward detecting elements are of similar size but the strips taper to point to the beam axis. In both the pixel detector and the SCT, the detecting elements are overlapped to provide a maximum of hermeticity.

For the alignment of the Inner Detector the ATLAS collaboration has developed a strategy that comprises several stages. This is described in detail in Ref.[12].

The first part of this strategy is to study the properties of the detector elements themselves. The ATLAS collaboration is actively studying the stability of individual elements of the tracker using Electronic Speckle Pattern Interferometry. The results of these studies are being applied to the design of the tracker.

In a second step, the positions of the detector strips in the SCT will be initially determined using an X-ray source. This source is capable of delivering short wavelength photons (10 - 50 pm) and allows production of very narrow beams without diffraction problems. The beam divergence can be kept low enough (200 \( \mu \text{rad} \)) using collimating slits. By scanning the beam across a given strip a Gaussian-like profile is obtained. By measuring the centre of this profile the accuracy of the measurement (with and analogue read-out) should be around 2 \( \mu \text{m} \).

The ATLAS collaboration considers that the inner detector may not be sufficiently stable over short periods of time to rely on alignment calculation based on tracks and have decided that automatic procedures are required. Consequently, as the last part of their alignment strategy, the positions of the individual detecting elements within the SCT and pixel trackers will be continuously monitored. This is to be achieved by making many measurements using the FSI and straightness monitoring techniques. These will be combined using a highly over-constrained, three-dimensional, geodetic network. The networks will be based on the one dimensional length measurements using the FSI. These will be supplemented by straightness measurements in the end-cap region. At each of the nodes of the network there are small units (aptly named “jewels”) that contain the fibres and retro-reflectors described in Section 3.4. The positions of the optical elements in each jewel is determined with respect to the support structure by an initial series of measurements. An example of the geodetic alignment network in the barrel SCT detector (adapted from a figure in Ref. [12]) is shown in Fig. 6.
It is estimated that, taking into account the positioning tolerances of the FSI jewels and the errors arising from the geodetic network, the $R\phi$ precision obtained from the geodetic network itself will be around 5 $\mu$m.

The last part of ATLAS’s alignment strategy is to use track alignment. In a first step, particles passing in the overlap region between two elements are used to find the position of adjacent elements, both in $R\phi$ and $z$. For this purpose low $p_T$ muons are used. Assuming that 100 tracks are required to fix the overlap between two neighbouring elements, this procedure should take about 24 hours. In the second step, high $p_T$ muons from W/Z decays are used to align layers with respect to each other and correct for distortions in the structure. To determine simple parameters such as the rotations of one barrel with respect to the others should not take more than a few minutes. To calculate more complicated distortions would require much more time and data.

REFERENCES


ALIGNMENT BY FEEDBACK

M. Ross
Stanford Linear Accelerator Center, Stanford, CA, 94309*

Abstract
New technology is required to increase the reach of future accelerators. In keeping with several trends in technology development, new accelerator technology leads toward miniaturization and active control using high level process control systems. Linear collider designs have emphasized both and planned machines include challenging component positioning tolerances and comprehensive control systems. This paper reviews some examples of this for existing and planned linear colliders.

1. INTRODUCTION

The last decade has seen the implementation of a fundamentally new high energy particle accelerator design, the linear collider. From its conception, linear collider developers felt they could avoid the energy scaling rules associated with circular electron/positron colliders by using low emittance beams and strong interaction point focusing [1]. The SLAC Linear Collider (SLC), while not able to perform at its design luminosity, has nevertheless shown that the idea of colliding beams of very small size is feasible [2]. The salient operational accelerator issues that impact collider operation are 1) generation of intense, low emittance beams, 2) preservation of the emittance throughout accelerator and beam delivery systems and 3) stabilization over all time scales, from sub-second to hours or days.

The topic of this paper is the stabilization system, in particular the part of it that addresses mechanical components; why it is needed and what it does and what it will do in the SLC and in future linear colliders. Of course, application of beam based feedback is not new nor is its future application limited to linear colliders. Machines such as 3rd and 4th generation synchrotron light sources benefit from the use of automatic beam-based-steering feedback and component alignment schemes.

In this paper we first review linear collider tolerances, justify them and outline some of the observable signals where their impact may be seen. Following that we describe the function, design and implementation of feedback systems at SLC. In section 4 we describe experience at SLAC with SLC and outline some of the benefits of the feedback systems in use there. In the final section we describe plans for the implementation of feedback and related procedures for the Next Linear Collider (NLC) [3]. Most of the material on the SLAC based NLC design was taken from the NLC Zeroth-Order Design Report, issued in July 1996.

Stabilization is required over all practical time scales. However, in a well designed system, the more cumbersome disturbances should occur at a relatively low rate. A good example of a problem that is cumbersome to compensate for is a poorly localized, global

* Work supported by the Department of Energy, contract DE-AC03-76SF00515
error. Typical disturbance sources, ranging from fast to longer time scales, are shown in the Table 1.

Most linear collider designs propose operation in the range of 100 to 2000 Hz. This paper describes approaches to controlling the effects of instabilities 2–4 in Table 1 which have time scales longer than the interpulse time.

Table 1

<table>
<thead>
<tr>
<th>Timescale</th>
<th>Typical source</th>
<th>Feedback sensor/corrector</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Pulse to pulse (collider repetition rate)</td>
<td>Beam dynamics, pulsed devices</td>
<td>Feedforward</td>
</tr>
<tr>
<td>2  Fast (few Hz)</td>
<td>Vibration, power converter</td>
<td>Position monitor/steering</td>
</tr>
<tr>
<td>3  Slow (minutes ~ hours)</td>
<td>Thermal</td>
<td>Position monitor analysis</td>
</tr>
<tr>
<td>4  Very slow (days ~ months)</td>
<td>Ground ‘settling’</td>
<td>Procedure-based optimization</td>
</tr>
</tbody>
</table>

2. LINEAR COLLIDER TOLERANCES

2.1 Design guidelines

Linear colliders are expected to provide the next level of e+/e- collisions with energies substantially beyond what can be achieved using storage rings, albeit at the cost of some complexity and loss of stability. Several linear collider designs are quite mature [4, 5] and vary somewhat but all have: 1) a low emittance source of e+/e-, usually including damping rings, 2) long linacs and 3) final focus systems on either side of a particle detector. A critical design consideration is the preservation of emittance throughout the transport from the source to the linac, the linac itself, and the final focus. Typical normalized emittances are $\gamma e_0 = 4 \times 10^8$ and $\gamma e_0 = 4 \times 10^8$ m-rad at the beginning of the linac. In most designs, the beams at the interaction point (IP) are so small that feedback of several sorts is required for optimum performance.

Linear collider tolerances are derived from considerations of the impact that a given error has on the luminosity. As design tolerances are tightened with respect to available technology, an engineering tradeoff decision is made that separates the expected mechanical or electronic system performance, in the absence of any beam pulses, and the system performance that results from the added use of beam-based optimization schemes. In general, since there are usually observable effects associated with a particular kind of error, the performance with the inclusion of a beam-based compensation scheme is usually better and such schemes must be devised. A notable exception to this paradigm is the correction of errors that cannot be cleanly localized.

The widespread implementation of beam based tuning and optimization processes comes at some cost and can have a negative impact on the performance of the accelerator. A procedure may, in general, take time and disturb the beams substantially. Each procedure must therefore be evaluated in order to estimate how often and to what accuracy the compensation scheme must be implemented. Instrumentation accuracy, control system
latency and procedure development therefore play a vital role in the ultimate analysis of machine performance [6]. The system designers must ensure that the more complex, more difficult to correct errors need attention less frequently. In general this means that system designers must play the role of operators long before construction, not to speak of operation, begins.

Some of the most difficult errors arise in the main linac and final focus collider sub-systems. In this paper we focus on examples from those areas in both SLC and NLC. In the linac and final focus, alignment and field magnitude errors can cause significant emittance dilution or effective spot size increase as well as simple trajectory distortions. In an ideal system, the beam size would be known throughout the machine on each pulse and identification of errors would be relatively simple. In practice, the distribution of beam size monitors and their performance is somewhat limited. Thus most beam based feedback uses beam position monitors (BPM’s). BPM’s are distributed in great numbers throughout the LC and the ultimate ability of the feedback system to suppress long term instabilities depends critically on them. As is discussed in section 5, one of the biggest challenges for NLC feedback is the BPM offset and calibration control.

Table 2 shows some typical beam sizes and emittances associated with high energy colliders. The sizes shown for SLC are the achieved sizes during routine, nominal, operation. Design sizes are somewhat smaller.

<table>
<thead>
<tr>
<th></th>
<th>x (µm)</th>
<th>y (µm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLC linac</td>
<td>100</td>
<td>50</td>
</tr>
<tr>
<td>SLC IP</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>NLC - linac entrance</td>
<td>30</td>
<td>5</td>
</tr>
<tr>
<td>NLC - linac exit</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>NLC IP</td>
<td>0.270</td>
<td>0.005</td>
</tr>
<tr>
<td>LEP [7]</td>
<td>135</td>
<td>5</td>
</tr>
</tbody>
</table>

### 2.2 Final focus

The goal of the final focus is to demagnify the beam size by a factor of 10 to 100. One of its most critical components is the chromatic correction system (CCS) that compensates for the chromaticity of the final lens system. Typical sextupole strengths are quite large in the CCS, with pole tip fields around 5 kG and $\beta$ functions with peaks close 200 km in the NLC design. The sextupoles can cause large geometric aberrations if not properly compensated.

Consider the position tolerances of the final focus quadrupoles near the interaction point (IP). On a pulse to pulse basis, the beams must remain in collision, so trajectory errors caused by the displacement of the final magnets must not result in IP motion comparable to a small fraction of the beam size. Since the beam size is quite small, the sensitivity is great. The magnet motion $y$ required to move the beam by its size ($\sigma_y$) at the IP is:

$$
\Delta y = \frac{1}{k \sin \mu} \sqrt{\frac{\varepsilon}{\beta}}
$$

(1)
where $k$ is the inverse focal length of the magnet, $\mu$ is the phase advance to the IP, $\beta$ is the beta function inside the magnet in question and $\varepsilon$ is the beam emittance. At the SLC, $k \sim 1$ m$^{-1}$, $\varepsilon = 5 \times 10^{-10}$ m-rad, and $\beta$ is 10 km, giving a $\beta_y$ of roughly 200 nm. Movements much smaller than $\sigma$, have a significant impact on luminosity so the engineering tolerance on the support stability is tighter. While obtaining stability at this level is well within the state of the art of mechanical supports, it is not trivial and some effort must be expended. The error caused by a displacement is simply an offset and no further aberrations result. This tolerance therefore must be met for short time duration only, comparable to the time between pulses. Fast, but relatively simple, IP steering feedback can be used to correct it [8]. If the correction dipoles are close by, the feedback effectively aligns the quadrupole centers. In section 4.2 this feedback is discussed further.

The principle of operation of the CCS relies on the cancellation of geometric aberrations that result from sextupoles used in the chromatic correction. This cancellation is done using a focusing cell of exactly 180° $\beta$ phase advance between two places of equal dispersion. This cell is known as a ‘-I transform’.

The CCS is much more sensitive to errors generated within it than it is to incoming launch errors. With the ‘-I’ symmetry, incoming trajectory effects are canceled. Incoming energy errors are compensated by the sextupoles. This is after all the purpose of the CCS and by correcting the aberrations caused by the energy spread of a single pulse it can, if stable, also correct for the aberrations caused by energy differences between successive pulses.
Single sextupole positioning errors generate a residual quadrupole error that can destroy the -I transform inside the CCS. Trajectory errors generated within the CCS (by, for example, correction dipoles) can have the same impact and must also be carefully controlled. Quadrupole position errors have a similar effect through their steering. Typical positioning errors for NLC are shown in Table 3. The most significant symptom of the loss of the -I inside the CCS is residual dispersion at the IP. Waist motion follows closely behind that. Typical waist shift sensitivity at SLC is 5 mm of motion for 100 µm of sextupole displacement. Since each error is evaluated on a single magnet basis, the displacement that causes a 1% luminosity loss is listed. When all magnets are taken together, the impact is more significant.

<table>
<thead>
<tr>
<th>Magnet</th>
<th>*y (nm)</th>
<th>Leading Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCY quad</td>
<td>20</td>
<td>Dispersion</td>
</tr>
<tr>
<td>Final transformer quad</td>
<td>6</td>
<td>Dispersion</td>
</tr>
</tbody>
</table>

The examples given above show what some typical positioning errors can do to the performance of the final focus. In order to understand the response to systematic motion, for example, to seismic plane waves or a distortion of that nature, a formalism has been developed [9, 3] that uses the 2 dimensional power spectral density of ground motion, $P(\omega, k)$ and a lattice response function $G(k)$. The lattice response function is the normalized beam motion at the IP for quadrupole motion caused by seismic plane waves with wave number $k$. When combined with the spectral density of the ground motion ($P(\omega, k)$ or $P(k)$ for a given instant in time), the rms beam movement at the IP, $<\Delta y_\beta(t)>^2$, is given by the product integrated over seismic waves of all wave numbers:

$$<\Delta y_\beta(t)>^2 = \int_0^\infty P(k)G(k) \frac{dk}{2\pi}$$

$G(k)$ is expected to be large for wavenumbers that are consistent with typical sizes in the beamline layout. Figure 2 shows $G(k)$ for the NLC final focus. The extremes are of special interest. In the limit of very long waves, the whole system is moved or tilted together so the sensitivity $G(k)$ is quite small. As the waves become short enough to fit between major focusing elements the response is quite large and shows the amplification expected from the demagnifying system as illustrated in Eq. (1). At this end of the spectrum, the beamline components are moving more or less independently, in a manner quite similar to the movement caused by support vibration. Note that $G(k)$ is defined here only for perfect supports that are located directly under each magnet’s center. The analysis is concerned only with magnet motion associated with the plane waves. The up and down spikes in $G(k)$ arise from accidental coherence between magnets.

For given expected $\sigma_y$, positioning precision requirements may be evaluated using $P(\omega, k)$ and $G(k)$. Figure 3 shows $P(f)$ for a ‘quiet’ NLC final focus site. In the figure, $f$ is exchanged for $k$ using a dispersion relation measured at SLAC. The three curves show the
ground motion, $P(f)$, (long dash), the feedback driven damping, $F(f)$, (short dash, see Fig. 3) and the lattice response function, $G(f)$. Luckily, the effects of ground motion are strongly suppressed by the lattice response, there is little power at high frequencies. However, feedback is still required for actual physical sites, where cultural and engineering related noise sources inject significant amounts of high frequency motion. SLC experiences with such noise sources are described in section 4.

![Graph 1](image1.png) ![Graph 2](image2.png)

**Fig. 2** Lattice response function for NLC final focus

**Fig. 3** Seismic plane wave spectrum for quiet site, lattice response and the expected feedback suppression ratio.

### 2.3 Linac

In the linac, the most challenging goal is the preservation of the beam’s transverse emittance given reasonably achievable magnet and structure positioning tolerances. The NLC linac emittance growth is expected to be about a factor of 2.

The NLC linac is expected to cause about a factor of two growth in the vertical emittance. The most serious contributors to this are: 1) Transverse wakefields from beam interactions with the structure, 2) Dispersion and chromatic effects from the large energy spread beam and 3) Transverse pulse to pulse fluctuations (jitter) caused by quadrupole vibrations. Of these three, the first and last will concern us here because they will be to some extent addressed using high level automated tuning procedures and beam based feedback.

Tolerances for the linac can be derived from the same considerations of beam to beam targeting at the IP. Because of the longer, more periodic structure, $G(k)$ is larger. Alignment of the linac is critical because of the emittance growth that results from offset passage through the accelerating structures. Linac structure and quadrupole placement tolerances for NLC are shown in Figs. 4 and 5. The errors shown result in a 25% growth of $\varepsilon_y$. In the linac, because of its length, great care must be taken to avoid systematic misplacements of the quads and structures with characteristic length scales similar to the $\beta$ function. This is the reason for the tight tolerances with $\lambda \sim 100$ m.

In contrast, the SLC structure placement tolerances are about 20 times looser because $\varepsilon_y$ is 200 times larger. The NLC linac requires active quadrupole and structure movers for optimum performance.
3. FEEDBACK

3.1 Goal

The goal of feedback is to stabilize steering and other beam parameters such as energy, energy spread and phase space volume and orientation using information from beam monitors. It should do this at as high a rate as possible, without decreasing stability (increasing rms beam motion), up to the beam repetition rate. Feedback also should respond quickly to step changes. The latter goal can greatly reduce the moment to moment activities of control room operators.

Typically, a feedback loop will acquire data from a given set of instruments, process it in a local processor using a filter algorithm and apply corrections through a set of actuators. Some loops do not need the response time afforded by the local signal processing and can use a simpler, slower, workstation based system. In practice exception handling and other data checks and diagnostics dominate the effort required for the implementation of the loop.

Feedback loops that include beam derived information have been used in many applications in accelerators. Typical applications are used to stabilize microwave systems, control longitudinal and transverse coupled bunch instabilities or provide local steering for synchrotron light beamlines [10]. Microwave system feedback keeps the accelerating vector in a microwave cavity oriented properly through transients seen in a variety of conditions, such as injection. This type of feedback could be considered part of that subsystem. The last example of feedback is the servo-steering used in synchrotron light machines to stabilize the light using photon beam position monitors. This feedback is responding to thermal, mechanical and other instabilities and its goal is typical of the type of loop used in linear colliders. Synchrotron light source trajectory feedback typically has a much higher response bandwidth since it involves storage rings rather than pulsed linacs. Its function is similar to the steering stabilization feedback used in linear colliders. We will not discuss the first two types of feedback since they have little to do with component alignment.

3.2 Design

Feedback at linear colliders is applied as a control layer on top of cooling water, air temperature and power converter control loops that do not nominally include beam based information.
From the point of view of the high level controls, the feedback we will describe is a digital process control loop implemented to compensate for a particular instability. In former times an operator may have been able to perform this task, but now, because the number of such tasks and the complexity of the accelerator system has grown significantly, it is imperative to relegate the task to a process control machine and do the feedback cybernetically.

The internal design of the SLC feedback loops follows somewhat formal lines [11,12]. This was done because its application was anticipated in a wide variety of situations. The design follows the ‘state space’ formalism adopted by digital control engineers. The state space formalism complements classical digital control design techniques that use transforms. In practice the state space formalism is better suited to multi-input – multi-output control tasks, as most beam steering applications turn out to be. Results from application of the two techniques are identical.

The beam ‘state’ is conveniently defined to have some meaning in the abstract, and is not directly tied to the reading of an individual monitor. Examples of ‘states’ are the beam energy at the end of the linac and the angle and offset of the beam trajectory at an arbitrary location in the beam line. Data from BPM’s are processed through a matrix transformation and an overconstrained least squares linear fit to provide estimates of the states, which can almost never be measured directly.

The beam state information from the feedback process is used offline in accelerator modeling to interpret instabilities and other effects. It is thus a good way to connect the feedback to the bulk of the accelerator control.

The state space formalism breaks the job into two parts, the definition of a ‘control law’ and the evaluation of a ‘state estimator’. A generic control law that calculates actuator settings from a given input state and reference is:

$$u_{k+1} = K\hat{x}_{k+1} + Nr \quad (3)$$

where:

- $u$ is the vector of actuator settings to be used.
- $\hat{x}_k$ is the estimated state on pulse number $k$, ($x_k$ is the actual state).
- $r$ is the reference input. In accelerator examples it is often the difference between the nominal reference trajectory and the desired trajectory.
- $K$ is the gain matrix. It contains information about the response of the system and it also contains the results of an offline optimization of the response of the loop system to the beam noise conditions.
- $N$ is the translation from $r$ to actuators.

The system is managed using the knowledge of the evolution of the last known estimated state, $Φ\hat{x}_k$, the expected response from the motion of the actuators, $Γu$, (for example dipole correctors in many cases) and the filtered difference between the last state and the present measurement, $L(y - H\hat{x}_k)$. These terms are summed to give the new expected state, $\hat{x}_{k+1}$,

$$\hat{x}_{k+1} = Φ\hat{x}_k + Γu + L(y - H\hat{x}_k) \quad (4)$$
\( \Phi \) is computed from the expected time response function of the actuators so that \( \Phi \hat{x}_k \) represents what has happened due to the last correction, \( \Gamma \) is mapping from the actuators \( u \) to the beam state, \( L \) is the optimum filter function and depends on the time structure of the instability that is to be controlled. \( H \) is also derived from the accelerator optics, like \( \Gamma \).

Figure 6 shows the feedback processing schematic. The present actuator settings are used by a model of the machine to predict the new state, \( \hat{x}_k \). The difference between the predicted and observed measurements (\( y_k - \bar{y}_k \)) is used to generate the new estimated state, \( \hat{x}_{k+1} \), using Eq. (4). The result is then used in Eq. (3) to determine the next correction.

Feedback needs several sorts of calibrations. First, the time response function of the correctors must be measured. Figure 7 shows the response of some typical SLC correctors. In practice, the most serious source of corrector delay at SLC is the field penetration time through the accelerator structure (\( \sim 0.8 \text{cm copper} \)). Following that, the beam transfer functions must be measured using simple corrector - BPM response tests. Finally, the expected BPM noise must be included in \( L \) since this is the primary source of error in the estimated state. Errors in one or more of these calibrations will result in poorer than optimum performance of the feedback system.

![Feedback system schematic](image)

**Fig. 6 Feedback system schematic**

### 3.3 Non-linear feedback

The above examples of steering feedback are strictly linear and involve just the transfer matrices between beam line components. Non-linear feedback at linear colliders has the additional goal of automating optimization, a task usually performed by operators. Even with the application of trajectory control feedback, there remain local errors as well as the effects of accumulated global errors that require correction. An example of this is the correction of the trajectory in the linac. While the feedback provides a fixed trajectory downstream, in the region between one loop and the next upstream loop and within the region spanned by the feedback’s own hardware some trajectory error occurs and can cause some degradation in the beam emittance. This has to be addressed by a global emittance correction process.

Another example appears in the final focus where systematic changes in position can cause an IP waist shift or other aberrations. The waist shift or dispersion error is not
correctable with simple steering since, after correction by the IP steering feedback, the beam remains out of optimum focus. In general, the waist can move in either direction and, unless there is some external information, a test must be done by the feedback controller to determine which direction to move [13]. What is required is a trial excursion in one direction, in order to determine the sign of the derivative of the response. Synchronous detection of the excitation allows it to be small compared to nominal operating tolerances of the device. At SLC this is known as ‘dither feedback’ and is under development.

Dither feedback is intended to operate on the driven derivative of the beam parameters with respect to some excitation. Since most optimization curves are locally parabolic near the optimum, the derivative is expected to be linear.

Fig. 7 Beam position monitor (BPM) response to a step change in a correction dipole. The dipole is mounted over the linac accelerating structure. The data point spacing is 1/60 s giving a dB/dt ~ 30 Gm/sec.

4. EXPERIENCE AT SLC WITH FEEDBACK

4.1 Design

Feedback loops can develop as part of the system design, as in NLC, or they can be developed in response to an observed problem. There is a set of steps through which a given procedure evolves as it passes from development to routine to automated use. This is mirrored in manufacturing when a prototype is brought into mass production. It is typical of feedback’s application to the prototype collider, SLC.

The noise structure of the SLC beam (jitter) contains a large amount of broadband ‘white noise’, a significant component at low frequencies, corresponding to thermal time scales, and a few spikes at mechanical resonance frequencies in some cases. The feedback will easily suppress the low frequency part of this noise distribution, but its effects on the rest of the spectrum must be tuned.
Figure 8 shows the results of that tuning. The suppression is excellent for low frequencies, below about 0.5 Hz, and poor for higher frequencies. There is a region where some anti-damping is observed. Since a large part of the beam rms comes from the low end of the jitter spectrum, this design satisfies the two fold goal of reducing the jitter and providing excellent response to step changes. It is hard to reduce the anti-damping and it can lead to problems, especially if the beam motion is driven by support vibration in this frequency range.

The plot shows the rejection ratio as a function of excitation frequency. The disagreement between the model and the measured feedback response is due the variations in correction magnet slewing times. The figure shows the simulated response (smooth curve), data obtained from transforming step impulses (jagged line) and data obtained from transforming cyclic excitations at various frequencies (black dots). The step response data is more sensitive to BPM noise since only a handful of steps were used. The frequency domain data is much cleaner since more averaging could be done.

![Frequency Response Graph](image)

Fig. 8 Measured and simulated frequency response for an operating SLC feedback loop

### 4.2 Examples

Seventy percent of the feedback loops at SLC control local beam orbit steering. Typically, a group of 6 to 8 BPM’s, spread over 4 linac FODO focusing cells with ~70° β phase advance, are used in conjunction with a pair of dipole correctors in each plane. The dipole correctors, spaced by roughly 90° β phase advance, are usually upstream of the BPM’s so that their behavior is checked by the measurement vector y. Each such linac loop provides independent x, x’, y and y’ control for both e+ and e- beams. About 30 such loops are routinely used in SLC operation and have proven vital for operation, often in ways beyond simple stabilization of beam jitter and drift.
Fig. 9 Feedback installations at SLC. The figure shows a schematic of SLC with a code to indicate the function of the digital feedback loop at that location:
S - steering, E - energy, I - intensity, C - collision, M - minimization (dither feedback).

An example of a special purpose feedback loop at SLC is the IP collision steering feedback which uses a key effect seen at linear colliders, the beam-beam deflection [14]. The instrumentation used for the feedback are the two pairs of BPM’s for each beam (a total of 8) used to determine the incoming and outgoing path along the ±8 m near the IP. The BPM’s are located in the IP focusing triplet structure. Information from the incoming BPM’s is vital for two reasons: 1) typical beam pulse to pulse fluctuations are arbitrary in phase of origin and equivalent to about 20% of the $\sigma_y$, the spatial and angular size of the beam, and 2) $\sigma_y$, the size of the beam’s angular divergence at the IP is around 300 µrad, about the same order of magnitude as the observed deflection. Indeed, additional incoming trajectory information is needed to further constrain the fit and give more accurate results.

Figure 10 shows the beam-beam deflection at SLC. In this plot, the kink angle caused by the effect of one beam on the other is plotted vs the steering of one of the beams. To take the data, the steering is accomplished in a quick succession of pulses. The plot shows a fit to the expected deflection function. The fitted parameter $\Sigma_y$ shows the convolution of the two beam’s vertical sizes, 0.584 µm. The fit does not provide information about the individual $\sigma_y$, in this case it is estimated from measurements using more conventional size monitors.

Fig. 10 Beam-beam deflection at SLC

When the beams are right on top of each other, they are optimally in collision, and the deflection is zero. Near that point the deflection is linear as a function of offset and the slope
is a good indication of the width. Thus the feedback has a relatively simple job to do; simply apply the steering correction to one beam in proportion to the observed deflection. Since the deflection is a reflexive effect, it is not possible to determine which of the two beams has caused the error from it alone. Some care must be taken to avoid compensating the motion of one beam by moving the other.

Even at transverse separations of many times $\sigma_x$, a substantial kick is clearly visible at SLC. The figure shows this clearly, with a deflection many times the intrinsic error in the measurement still visible at the edge of the plot, approximately 20 $\sigma$ from the central region. This makes the loop quite robust and it often restore collisions quickly even after a significant time has passed or after beam steering changes have been made. This is one feature of the linear collider that compensates for some of its natural instabilities.

Table 4 summarizes the application of feedback and routine optimization in the SLC. The second column of the table indicates the type of feedback, restoration to a specific setpoint (F) or optimization, ‘best’ value tuning (O). Loops of types 1 and 3 have been the most successful and are described in the text. Energy feedback is quite similar to steering, and uses local spectrometer optics to estimate each pulse’s energy. This type of loop is also quite successful. Other rows refer mostly to optimization, done either manually or with the aid of semi-automatic data gathering and processing tools.

**Table 4**

Feedback and optimization in use or planned at SLC

<table>
<thead>
<tr>
<th>Parameter Controlled</th>
<th>F/O</th>
<th>Detection Instrument</th>
<th>Bandwidth max 120Hz</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Position and angle</td>
<td>F</td>
<td>BPM</td>
<td>20 Hz</td>
<td>provides diagnostic data</td>
</tr>
<tr>
<td>2 Energy</td>
<td>F</td>
<td>BPM</td>
<td>120 Hz</td>
<td></td>
</tr>
<tr>
<td>3 Collision</td>
<td>F</td>
<td>IP BPM’s (deflections)</td>
<td>120 Hz</td>
<td></td>
</tr>
<tr>
<td>4 Compressor optics</td>
<td>O</td>
<td>Wire scanners at linac launch</td>
<td>hours</td>
<td>Uses asymmetric gaussian fits</td>
</tr>
<tr>
<td>5 IP spot</td>
<td>O</td>
<td>Deflections and luminosity mon</td>
<td>minutes</td>
<td>Can use dither</td>
</tr>
<tr>
<td>6 Linac emittance</td>
<td>F</td>
<td>Wire scanners in linac</td>
<td>minutes</td>
<td>Uses asymmetric fits and skew moment propagation</td>
</tr>
<tr>
<td>7 Beam phase (linac energy spread)</td>
<td>F</td>
<td>BPM’s using dither phase synchronous excitation</td>
<td>minutes</td>
<td>All pulses must be dithered to achieve needed accuracy</td>
</tr>
<tr>
<td>8 Positron capture phase</td>
<td>F</td>
<td>Beam power integrator</td>
<td>120 Hz</td>
<td>uses estimated temperature</td>
</tr>
<tr>
<td>9 Kicker timing</td>
<td>O</td>
<td>Linac BPM’s</td>
<td>minutes</td>
<td>Correlates beam with kicker thyatron timing</td>
</tr>
<tr>
<td>10 Arc tuning</td>
<td>M</td>
<td>Arc BPM’s</td>
<td>days</td>
<td>Highly specialized; expert based optimization</td>
</tr>
</tbody>
</table>
4.3 Results

Perhaps the best way to illustrate the impact of feedback on machine performance and understanding is through illustrations of ‘history’ or time record data from relatively long time periods. In the figures, we show examples of the kind of hints that feedback can provide that are subsequently used to guide improvement efforts.

Through history records, feedback can indicate alignment degradation and component drifts over day and week time scales. At the other end of the frequency spectrum, at very short intervals, feedback and its related data acquisition can provide data relating to mechanical vibration, another form of poor mechanical subsystem performance.

Prior to the feedback era, the operator could examine the change in trajectory with respect to the saved reference trajectory and make hand corrections to null the difference. Since the changes are relatively slow, with multi-hour time scales, this is not an unreasonable way of responding to this slow drift. However, as the number of such locations grows and as better quality is required of the nulling process, it becomes unreasonable to expect the operator to correct the trajectory in each location. The primary goal of feedback at the SLC is exactly this. Through the records kept by the feedback process, a diagnosis of the source of the drift can begin. Without feedback, with each operator correcting the drift using a slightly different technique, the clear unfolding of the underlying causes is more difficult.

Figure 11 shows a record of the beam trajectory at the exit of the SLC damping ring for a period of several days. The two ‘history’ records in the figure show the performance of the feedback at the exit of the electron damping ring (top) and the beam intensity (bottom) during the 10 day interval May 5, 1996 through May 15, 1996. There are about 10 data points per day with a point to point spacing of 2 hours. In the figure the points are connected. The dips shown in the bottom plot are the intervals when the beam was absent from the ring. During that time, the ring cools substantially and the components in it move a little, maybe 20 to 40 µm; the exact amount is not known. In the top plot, the correction required to keep the vertical trajectory fixed to within a 5 to 10 µm rms in the extraction transport is shown. This is actually $F'\mathbf{u}$ from Eq. (4).

An interesting aspect of this plot is the movement required by the correctors to stabilize the orbit after the beam is restored. Typical decay constants are ~1 hour. One of the events is noteworthy, early in the afternoon of May 7 the response is more severe due to the de-excitation of the magnet power converters. In each of the other events, only the beam was off. To further reinforce this, note the two large corrector events late on the 13th and during the afternoon of the 14th. In these episodes the beam was present, but at a much lower duty cycle, about 15% of nominal. Thus the power dissipated in the ring by higher order mode losses in the ring’s internal structures and synchrotron radiation is reduced. These two taken together amount to 25 kW for nominal, full repetition rate, operation at $9 \times 10^{10}$ e- in two bunches (120 mA). The motions caused by the ‘beam off’ events are corrected by a 50 µrad kick, equivalent to a magnet strength of 1 to 2 Gauss-meters. The correction elements involved produce a 1 G-m kick with an excitation of 0.12A. Their tolerances are set at 10% of that.
Fig. 11 Ten-day history record of extraction from the electron damping ring showing correlation with beam intensity

Figure 12 illustrates another kind of thermal instability, this one associated with the ever present day night temperature changes of the California climate. The figure shows the temperature inside the SLC final focus (solid) during a 10 day interval in July 1996 superimposed on the vertical feedback correction command (dashed). The tunnel temperature changes are small (within 2.2 •F or 1.2 •C), but they are large enough to cause the beam trajectory to move. The outside temperature is shown above the plot in order to indicate the extent to which it contributes to the apparent motion. Note the phase shift between tunnel temperature and the feedback actuator. In this example, the temperature of the outside environment is causing slight misalignments of the machine components. This can be fixed with improved environmental control and by examining the local optics sensitivity.
Fig. 12 Ten-day history of initial beam launch conditions in the SLC final focus showing the correlation with temperature

Figure 13 shows another result from the SLC final focus. The data was not derived from the feedback process itself, as in the other examples, but from a similarly constructed monitor or ‘watchdog’ process. The watchdog in this location is intended to monitor the positions of the CCS sextupoles. It does this by keeping track of the average BPM readings in the ‘-I’ spaced sextupoles. The presence of an incoming and outgoing beam is an added feature of the final focus that adds redundancy to the monitor. The sextupole pair position illustrated in Fig. 13 exhibit roughly 100 µm peak to peak vertical day to night motion. A thermocouple mounted on the support of one of the magnets shows the temperature correlation, presumably indicating the cause of the problem.

The performance of the feedback depends on the incoming pulse fluctuation characteristics. For example, if the feedback is tuned for a broadband, smooth frequency pulse to pulse jitter spectrum with no particular single frequency lines, it will respond in a different way than if these lines were suddenly to appear with some strength. Causes for this may be associated with electronic or mechanical failures. The vertical motion spectrum of a pathological linac quadrupole is shown in figure 14. The quadrupole is pushed longitudinally at its support’s resonant frequency by the water cooling the nearby accelerating structure. Normally this would have no impact on the magnet’s vertical position but, in this case, the support is not directly under the magnet so the longitudinal motion is coupled into the vertical. Figure 15 shows the beam jitter spectrum in the SLC linac. Resonance lines associated with the mechanical support systems may be seen at 10 Hz. The beam size in this location is about 80µm. Using this analysis, we have been able to identify and correct various support related instability signatures in the 5 to 40 Hz range. Figures 14 and 15 show two such problems.
Fig. 13  15-day record from the SLC final focus sextupole positioning monitor showing the correlation with the sextupole support temperature. The units of the left side of plot are mm.

Fig. 14  SLC linac quadrupole vertical position motion spectrum

Fig. 15  SLC Linac beam jitter

In the pathological example shown in Fig. 14, the quadrupole has residual resonant behavior at 20–30 Hz. The top plot is the power spectrum from the commercial accelerometer. The bottom part of the figure shows the integral of the spectrum, starting from high frequencies. This plot shows the total amplitude of the motion in the frequency range 2–100 Hz. In this range, feedback performance is poor, but its data stream can be used to diagnose problems. The steps in the bottom plot, show by what fraction the total motion is reduced if the support is stiffened.

The BPM data shown in Fig. 15 come from a pair of digitizers per plane (x,y); the ratio of the numbers gives x or y. Since the numerator is close to zero, its bit noise is much more
significant and, in the top part of the plot, this digitization quantization is easily seen, with the least significant bit size of about 20 µm. However, the 10 Hz component of the beam motion is also clearly seen in the first section (0 - 1 s) of the top plot.

In the last example, Fig. 16 shows how feedback can be used to track jitter over long time scales. It shows the history of the pulse to pulse stability of the electrons at the entrance to the linac, integrated over about 10 seconds, for a 92 day period from May 1 to August 1. This case, also illustrating an anomalous event, the beam jitter grew in late July 1996 by about a factor of 2. The cause for this increase is not understood; typical angular beam sizes in that location are 50 µrad, so this jitter is quite small by comparison. The jitter ‘floor’ of about 30 nrad in the plot is caused in part by instrumental noise.

![Graph showing beam RMS horizontal motion over time](image)

**Fig. 16** Three-month history record of beam rms horizontal motion, or 'jitter', at the entrance of the linac (1.2 GeV)

## 5. APPLICATIONS – NEXT LINEAR COLLIDER

Widespread use of feedback systems forces a more careful look at instrumentation, particularly BPM, performance. In effect, by using BPM’s as the primary sensors for feedback systems, the BPM system becomes part of the power converter control system. It is therefore critical to avoid systematic errors in the instrumentation, such as thermally dependent gains and offsets since these will compete with similar offsets arising from thermal effects in the power converters and mechanical supports.

In the NLC linac an on-line BPM offset calibration scheme must be devised for generating corrections. If a monitor system with inherent offset stability or equivalently, an accurate offset calibration system were devised, then a beam based BPM calibration scheme would not be as critical. Most BPM systems do not have an offset calibration mechanism that is part of the BPM and calibrates the monitor, cable response and electronic offsets. Reference [15] is the only one found and it includes no long-term performance data. It is anticipated one will be required at NLC. Figure 17 shows how it might work.
Beam based BPM calibration schemes have been tested at LEP [16] and SLC. The schemes rely on either 1) sub-tolerance excitation and synchronous detection or 2) use multiple kinds of beams, such as positrons and electrons traveling in the same direction, or beams of different energies. In most of these cases a BPM offset, arising from within the instrumentation itself is not easily distinguished from the offsets of the beam line components themselves. Since the drifts in the instrumentation are equally as severe as movements of the components themselves, automated schemes are required for finding and correcting them. The schemes, both tried and proposed, fall into two categories, invasive and non-invasive. Non-invasive is somewhat of a misnomer since the procedure may still have an impact on the luminosity. We list here three schemes.

Reference [15] is a good example of the sub-tolerance excitation referred to in section 3.3. Individual LEP quadrupole magnet windings are excited using a sine function current source with $10^4$ strength of the nominal current. BPM zero offsets with respect to the quadrupole center are determined by watching the magnitude of the beam response to the excitation. If the offset is large, the response will be correspondingly large. Since the excitation is narrow band, in the range of 1 to 17 Hz, and the data acquisition is broadband and can provide data from each LEP turn, the signal/noise ratio is very good. Different frequencies can be used with different quadrupoles so that optimization of several can proceed at the same time. A beam pickup similar to those used for the tune measurement provides the line strength information. A local beam bump near the quadrupole in question can be varied until the transfer of the sine signal is minimized. An accuracy of about 100 µm has been achieved.

In the case of SLC linac [17] BPM’s offsets can be estimated using the trajectories of the e+/e- beams in a short, three BPM section. The simple geometry is illustrated in Fig. 18. The position monitor measurements, $\tilde{y}_m$, can be related to the incoming position and angle of both beams and the offsets of the intermediate magnet and BPM using:

$$\tilde{y}_m = M\tilde{v} + \tilde{c}$$

(5)

where
\[
\vec{v} = \begin{bmatrix}
y_0 \\
y_1 \\
\theta_0 \\
\theta_1 \\
y_n \\
y_Q
\end{bmatrix}
\]  

(6)

is the vector of initial conditions and the central monitor’s offsets, \( \vec{c} \) contains adjustments associated with the small correction dipoles and \( \mathbf{M} \) is the transformation matrix between these vectors. The last two elements in \( \vec{v} \) are the offset of Q2 with respect to the line drawn between Q1 and Q3 (\( y_Q \)) and the offset of the BPM inside Q2 with respect to its magnetic center (\( y_n \)). The latter term has contributions from the mechanical BPM electrodes, the cables connecting the BPM to the electronics and the electronics itself.

![Fig. 18 SLC BPM alignment test schematic](image)

The analysis works well where the phase advance per cell is small compared to 90° and it assumes that \( y_n \) is the same for \( e^+ \) and \( e^- \). This is not always be appropriate since the electronics will respond differently for opposite polarity signals. It has been used to monitor offset stability.

The NLC linac will use a combination of these three calibration mechanisms. NLC linac quadrupoles are mounted on movers so that their offsets with respect to each other can be perturbed and therefore nulled. With the cam based mover mechanism [18] it is possible to move the magnets in increments small compared to the single component position tolerances. This may also work with synchronous detection techniques.

### 6. CONCLUSION

In the last decade, commercially available test instrumentation with integrated processing has greatly improved in performance. This is clearly illustrated in the improvements to alignment instrumentation. It is important to develop this technology and try to realize the benefit and cost saving that it affords in the next generation of accelerators. Linear colliders and free-electron laser accelerator systems are perhaps the first to require such accurate component positioning, but they probably won’t be the last.

### REFERENCES


[17] This analysis was done by Chris Adolphsen.

[17] This analysis was done by Chris Adolphsen
BIOLOGICAL EFFECTS OF MAGNETIC FIELDS

S. Zannella  
CISE SpA, Segrate (Milan), Italy

Abstract

The use of devices generating high magnetic fields in industrial processes, energy production and storage, medical diagnostics, new transport vehicles and large scale research facilities is expected to expand significantly in the near future. Scientific and public interest has focused recently on the biological effects and the potential health risks associated with the exposure to magnetic fields. Over the last twenty years several laboratory studies and epidemiological surveys have been carried out in this field but no definite conclusions concerning the related risks for personnel or general public have been drawn. The aim of this article is to provide an overview of these investigations together with an analysis of the well-established effects of the interaction between static or ELF magnetic fields and living matter. The international guidelines and standards for exposure to magnetic fields are also reported and discussed.

1. INTRODUCTION

Magnetic fields are produced by electrical appliances, power lines, electromagnets and everything that carries electric current. The level of magnetic fields at which humans may be exposed has considerably increased over this last century. The exposure level is normally limited for practical reasons. To work in places with appreciable magnetic fields would be nearly unrealisable nowadays: computer disks may be erased, monitors would be distorted, magnetic light objects may experience rotational or translational forces. A much more important question regards the possible health risks associated to static and ELF magnetic field exposure. The potential health effects of biomagnetic interactions have been under discussion over the last twenty years and the debate is still open.

This article is a review of the known biological effects caused by the interaction between magnetic fields and living matter together with a survey of the laboratory and epidemiological studies described in the scientific literature. We focus our attention on static and Extremely Low Frequency (below 3 kHz) magnetic fields. In the ELF range ionising effects (when the energy carried by the field is so large that it can damage internal organs and biologic molecules like DNA) do not play a role; usually also the thermal effects (due to induced electric currents in the tissues) are negligible if we consider the typical magnetic field strengths associated to common electrical equipment. Nevertheless, there are other ways in which magnetic fields may interact with biological matter to produce biological changes. Whether these changes can lead to health risks, in the long term, is still an open question. The initiatives to establish guidelines and official standards for occupational or public exposure to static and ELF magnetic fields will be described and discussed.
2. THE MAGNETIC FIELD

A magnetic field is a region of space that results from the motion of electric charges, it is always associated with everything that carries electric current. The field may be pictured as lines of force, also called flux lines: the direction of the field at any point is given by the direction of the line in that point and its magnitude is proportional to the density of lines near the point. Unlike electrostatic field lines, the flux lines are continuous without beginning and end (this means that isolated magnetic poles do not exist). The magnetic field is described by two vector quantities: the magnetic field strength $H$ and the magnetic flux density $B$. These two quantities are related by the relation $\vec{B} = \mu \cdot \vec{H}$. The constant of proportionality $\mu$, the magnetic permeability, depends on the medium and in the case of biological tissues is assumed to be equal to the value of the permeability of free space $\mu_0 = 4\pi \times 10^{-7}$ (T m/A).

The magnetic flux density $\vec{B}$ may be defined in terms of the Lorentz force $\vec{F}$ acting on a charge $q$ that moves in a magnetic field with a velocity $\vec{v}$ [1]:

$$\vec{F} = q \cdot (\vec{v} \times \vec{B})$$

The unit of the magnetic field, in the SI system, is the tesla (T) while in the cgs system it is the gauss (G) with $1 \text{G} = 10^{-4} \text{T}$. The strength of the magnetic field decreases with the distance from the source. In the approximation of a long wire carrying an electric current (valid for small distances $r$ compared to the straight portion of the wire) the magnitude of the field varies as:

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

In the case of a dipole approximation (valid for example for field calculation at large distances from coils carrying electric current) the field amplitude decays more rapidly as:

$$B (r, \theta) = \frac{\mu_0}{4\pi} \frac{m}{r^3} \left(1 + 3 \cos^2 \theta \right)^{\frac{1}{2}}$$

where $m$ is the dipole magnetic moment of the coil and $\theta$ the angle with respect to the dipole axis.

3. THE EARTH’S MAGNETIC FIELD

The geomagnetic field of the Earth is dipolar (the magnetic poles are not coincident with the geographic poles) and varies at the surface from 26 µT near the equator to about 60 µT near the poles. In the last centuries the dipole moment is continually decreasing and it is assumed that it reverses every $\approx 200,000$ years. The magnetic field is maintained by the so called geodynamo: the interaction of the already existing Earth’s field with the molten iron of the outer core, that flows around the solid inner core, induces an electric current just as in a metallic wire that moves across a magnet. Once the electric current is established it generates
a self-perpetuating magnetic field that sustains the Earth’s field. The forces driving the conducting fluid arise from both the rotation of the Earth and heat.

4. **ARTIFICIAL MAGNETIC FIELDS**

The Earth’s static magnetic field has roughly not exceeded 100 μT over the last 80 million years. The natural magnetic field consists also of time-varying components, associated mainly with solar activity and thunderstorms, whose intensities vary from about 0.1 μT to 0.1 fT in the ELF range. Prolonged exposure to higher static and ELF fields is nowadays an ordinary situation as shown in Table 1.

<table>
<thead>
<tr>
<th>Field source</th>
<th>Frequency (Hz)</th>
<th>Magnetic flux density</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Offices, homes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Background</td>
<td>50/60</td>
<td>0.05–0.4 μT</td>
</tr>
<tr>
<td>Household appliances</td>
<td>50/60</td>
<td>0.01–0.5 μT at 1 m</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1–30 μT at 0.3 m</td>
</tr>
<tr>
<td>Video displays</td>
<td>30–3,000</td>
<td>0.02–0.6 μT at 0.3 m</td>
</tr>
<tr>
<td><strong>Research facilities (personnel areas)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear accelerator</td>
<td>0</td>
<td>0.1–5 mT</td>
</tr>
<tr>
<td>Bubble chamber</td>
<td>0</td>
<td>&gt; 50 mT</td>
</tr>
<tr>
<td>MHD and fusion plants</td>
<td>0</td>
<td>1–50 mT</td>
</tr>
<tr>
<td>NMR</td>
<td>0</td>
<td>1–60 mT</td>
</tr>
<tr>
<td><strong>Industries</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electrolytic processes</td>
<td>0, 50/60</td>
<td>1–10 mT</td>
</tr>
<tr>
<td>Aluminium production</td>
<td>0</td>
<td>1–10 mT / 60 mT</td>
</tr>
<tr>
<td>Electric and induction furnaces</td>
<td>0, 1–10,000</td>
<td>1–50 mT</td>
</tr>
<tr>
<td>Welding machines</td>
<td>0, 50/60</td>
<td>0.2–10 μT</td>
</tr>
<tr>
<td>Security systems</td>
<td>0.1–10,000</td>
<td>up to 1 mT</td>
</tr>
<tr>
<td>Average exposure of workers</td>
<td>50/60</td>
<td>1 μT, electrical</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.17 μT, non electrical</td>
</tr>
<tr>
<td><strong>Power systems</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>380 kV transmission lines</td>
<td>50/60</td>
<td>1–20 μT</td>
</tr>
<tr>
<td>15 kV distribution lines</td>
<td>50/60</td>
<td>0.05–0.4 μT</td>
</tr>
<tr>
<td>20 MWh S/C Magn. energy storage (SMES)</td>
<td>0</td>
<td>0.5 T (max. accessible field)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10 mT at 300 m</td>
</tr>
<tr>
<td><strong>Transportation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnetically-levitated trains (MAGLEV)</td>
<td>0</td>
<td>2–6 mT (head level)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20–50 mT (floor level)</td>
</tr>
</tbody>
</table>
Static fields of significant intensity are encountered mainly in industrial processes and in large scale scientific facilities but may be experienced by the public in medical equipment for diagnostic or therapeutic purposes and in new emerging technologies like magnetically-levitated trains.

Exposure to ELF magnetic fields, essentially from 50/60 Hz sources, is an ordinary situation. The magnetic field background level in homes (away from appliances and averaged over time) ranges from 0.05 μT to 0.4 μT (based on a EPRI study of nearly 1000 homes) and higher localised magnetic fields are present near household appliances. People living in the proximity of power transmission lines, or workers in some industrial sites, may be exposed all the time to magnetic fields higher than 1 μT. Levels of tens of μT can occur for short periods in certain working situations.

In the case of both static and ELF fields, ionising and thermal effects are negligible but other mechanisms play a role. Scientists are investigating the effects of these magnetic environments on humans. The well-established effects of the interaction between static or ELF magnetic fields and living organisms may be divided into three main categories: electrodynamical, magnetomechanical and induction of electric currents. The last one is effective only for exposure to ELF magnetic fields or in the case of rapid motion in high static fields.

### 5. INTERACTION MECHANISMS BETWEEN MAGNETIC FIELDS AND BIOLOGIC SYSTEMS

#### 5.1 Electrodynamic effects

A well-recognised effect of the interaction of magnetic fields with the cardiovascular system is the change in electrocardiograms (ECG). Moving ionic charge carriers (electrolytes) in the blood, when exposed to a magnetic field, are subjected to the Lorentz force, reported in Eq. (1), that induces an electric potential \( \phi \) given by:

\[
\phi = v \cdot B \cdot d \cdot \sin \varphi
\]  

where \( v \) is the velocity, \( d \) is the diameter of the artery and \( \varphi \) is the angle between the direction of the blood flow and the magnetic field. This phenomenon is the basis of the Hall effect in solid-state materials and magnetohydrodynamic (MHD) power generation. For example, in a man with a blood flow rate of 0.6 m/s and an aortic diameter of 0.025 m, the expected induced potential is 15 mV/T [2]. These induced electric potentials have been observed by ECG on mammalians exposed to magnetic field [3]. Clear typical modifications in the ECG signal are visible in the T-wave region delimiting the opening and closing times of the aortic valve.

The experimental investigations confirmed that:
a) the magnetically induced alteration in the flow potential is generally well observed above 0.1 T – 0.3 T (depending on body size);
b) the amplitude of the T-wave increases linearly with the field;
c) no changes in the arterial pressure are observed;
d) the effect is completely reversible without adverse effects and disappears instantly at the end of the exposure.

The last consideration is important to underline that an observed change in a biologic system, during field exposure, is not necessarily an evidence for adverse human health effects.

5.2 Magnetomechanical effects

5.2.1 Magnetic orientation

In a uniform magnetic field, both diamagnetic and magnetic substances, are subjected to a torque that will tend to orientate them. In biological systems, there are several examples of orientation in strong static fields. Diamagnetic macromolecules undergo a magneto-orientation owing to the anisotropy in their magnetic susceptibility along the different axes of rotational symmetry. These molecules, generally with a rod-like shape, will tend to rotate in order to achieve a minimum energy configuration. The degree of alignment $\beta$ is usually very small; however for stacked assemblies of $N$ macromolecular with parallel rotational axes, $\beta$ is increased of a factor $N$, giving rise to large effects. Observed examples, by in-vitro studies, of nearly complete magnetic alignment in static fields of 0.5 T – 2 T, are the outer segments of retinal rods [4] and cells containing chloroplasts [5]. Also sickled red blood cells have been observed to orient perpendicular to magnetic fields of 0.2 T to 0.5 T [5, 6]. Both these effects happen at field strengths used in MR imaging systems and may be important for safety considerations. However, the magnitude of the response is small and probably does not result in any detectable clinical consequences in humans.

Also some gel-like tissues, such as the vitreous fluid of the eye and the synovial fluid of the skeletal joints, may be affected by exposure to magnetic fields. The gelation temperature of aqueous 1.4 % agarose solutions, similar to these biologic fluids, showed an increase as a function of the magnetic field strength [7].

There are some interesting cases of orientations of living organisms that synthesise organic chain structures, containing magnetite (Fe₃O₄) crystals with a net permanent magnetic moment, called magnetosomes. It was discovered that these magnetosomes influence the direction of motion of magnetotactic bacteria. They align themselves with the Earth's magnetic field lines and swim toward the north and downward (due to the vertical component of the geomagnetic field) in the northern hemisphere and to the south and downward in the southern hemisphere. This motion allows them to survive in the oxygen-poor mud of their aquatic environments.

There is also experimental evidence that the Earth's magnetic field influences the geomagnetic orientation and navigation of some migratory (such as some species of salmons) and elasmobranch (such as sharks, skates and rays) fish, migratory bird species, homing pigeons, monarch butterflies and honeybees (during their waggle dances) [8, 9].

5.2.2 Magnetic translation
Ferro magnetic and paramagnetic materials exposed to a magnetic field gradient are subjected to a magnetomechanical force (that tends to move them along the gradient direction) given by:

\[ F = V \cdot \frac{\lambda}{\mu_0} \cdot B \cdot \frac{dB}{dx} \]  

where \( V \) is the volume of the magnetic substance and \( \lambda \) the magnetic susceptibility. Owing to the limited amount of magnetic substances in most living beings, the influence of this effect on biological functions is negligible.

Important safety considerations concern the possible displacement of metal implants and prosthesis that may experience significant forces and torques in strong magnetic field gradients.

5.3 Electric currents induction

Time-varying magnetic fields induce electric currents in biological systems that may be evaluated by the Faraday law of induction. In the case of sinusoidal fields with amplitude \( B_o \) and frequency \( f \), the magnitude of the induced current density is given by:

\[ J = \pi \cdot r \cdot f \cdot \sigma \cdot B_o \]  

The proportionality of the induced currents on loop radius \( r \) and tissue electrical conductivity \( \sigma \) has important consequences for biological systems. A fixed time-varying field may induce notable currents at the macroscopic level but much smaller ones at the cellular level. These currents are usually smaller than those naturally produced by the brain, nerves and heart.

5.3.1 Magnetophosphenes

A well known biological effect of ELF magnetic field is the induction of visual sensations (flickering white light in the eyes), called magnetophosphenes, when exposed to fields having frequency in the range 10–100 Hz and amplitude above 10–100 mT [3]. Magnetophosphenes have been found to occur also in strong magnetic field during movement of the head and in transient fields during energising or deenergising of high-field magnets. This effect was first described by d'Arsonval in 1896 [10] and the possible explanation was reported by Lövsund [11]. The maximum sensitivity is at 20 Hz where the flashes are synchronised with the field.

Other biological effects of circulating currents in the body are: bone healing, nerve stimulation, electroshock anaesthesia (therapy) and heart fibrillation. They may be classified on the basis of threshold values of the induced current densities [12] as shown in Table 2.

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threshold values of ELF induced current densities for producing biological effects</td>
</tr>
</tbody>
</table>
Induced current density (mA/m²) | Effects
---|---
< 1 | Same order of naturally flowing biocurrents, no effects.
1–10 | Minor biological effects.
10–100 | Magnetophosphenes, bone fracture healing, possible nervous system effects.
100–1000 | Influence on neuron excitability; stimulation threshold for sensory receptors, nerve and muscle cells with possible health hazards.
> 1000 | Possibility of ventricular fibrillation, continuous muscle contraction; definite health hazards.

Owing to differences in biologic matter conductivities and unknown current loops, the calculation of induced currents is rather complicated. However, using cautious assumptions, an estimation of the threshold magnetic field values for the different effects may be made as reported in Ref. [12]. These estimated values give an idea of the ELF magnetic fields that should not produce biological effects but may not be used as safe limit values.

6. LABORATORY STUDIES

Several kinds of biological effects have been reported in studies of exposure to magnetic fields by animal experimentation and by work with cell cultures, trying also to find biological evidence of adverse health effects. It is not possible to report here on the extensive literature existing on this topic [13]. Some of the results (change in functions of cells and tissues, decrease in the hormone melatonin, alterations of immune system, accelerated tumour growth, changes in brain activity and heart rate) were obtained with field levels that are orders of magnitude larger than fields involved in ordinary cases. Some effects on cell cultures due to ELF low fields of a few μT and less than 100 μT were also reported [14].

Laboratory studies confirmed, as shown above, no biological effects for induced currents lower than 10 mA/m². In the case of exposure to static fields, the reported studies for fields lower than 2 T seem to indicate the absence of irreversible effects on the main biological functions.

These results should be treated, in any case, with great attention because the human organism has many compensating mechanisms that may modify the effects observed in cell cultures.

Some researchers also reported data on biodetection of high static magnetic fields or gradients. In one experiment [15], rats avoided systematically to enter into regions where the field strength was 4 T and the gradient was up to 13 T/m. These findings were not observed at 1.5 T.

7. EPIDEMIOLOGICAL STUDIES
Although animal experiments, cell culture studies and computer models provide useful data, most researchers agree that potentially adverse health effects of static or ELF magnetic fields may be provided by studies of human population that are ordinary exposed to magnetic fields in residential or working areas. These observational studies, called epidemiologic, may show associations that could point out an increased risk of disease associated with some environmental factors. They already allowed the important risk factors for cancer to be identified as cigarette smoke (relative risk of 10 for lung cancer) and benzene. In several cases, scientists cannot be sure whether the association is one of cause and effect or if the increased risk may be related to other factors.

This methodology requires carefully consideration because a positive association between a disease and exposure is not necessarily a definite proof especially when the risk is small. To judge if the increased incidence of risk is real other correlation criteria must be considered such as consistency with other studies involving different methods and population, dose-response relationships (increasing exposure levels should correspond to higher disease rates), plausible biological explanation supported by laboratory results, reliability of information.

Several epidemiologic surveys on the possible health risks associated with static and ELF magnetic field exposure have been carried out over the last 20 years, principally on 50/60 Hz fields, but none of them has definitively convinced the legislators.

Table 3 reports a review of epidemiological studies of occupational and residential exposure to static and ELF magnetic fields. These studies examined mainly electrical workers that are ordinarily more exposed to ELF magnetic (mean exposure of about 1 μT against 0.2 μT of workers in other job categories). The doses of field exposure in the first surveys were based only on job titles and not on actual measurements fields.

Some epidemiologic residential and occupational studies have suggested a weak relation with a few types of cancer in humans, particularly leukaemia in children as well as brain and breast cancer in adults, while others reported no consistent evidence of relations between magnetic field exposure and any type of cancer. Moreover, the studies reporting a positive association are not quite concordant and do not agree upon the forms of cancer. Considering three recent studies we may observe in fact some controversial conclusions. The first one, conducted by the Swedish National Institute of Working Life, one of the largest studies involving a broad range of industries and occupations [24], found an association with chronic lymphocytic leukaemia and an increased risk for brain cancer for men exposed to an average field of more than 0.2 μT. The second one, conducted by Canadian and French researchers on 223,292 electric utility workers, between 1979 and 1989, found a relative risk of 3 to contract acute myeloid leukaemia [17]. The last study, the most recent one, conducted in North Carolina, and involving 139,000 US utility workers found no association with both types of leukaemia but supported an association with brain cancer [23].

A noteworthy survey has been conducted in Sweden on people living near high-voltage transmission lines [25]. This study, highly exposed in the media and government circles, suggested for the first time a dose/response relationship, although it was based on a small number of cases. The risk to contract childlike leukaemia was found to be 2.7 times higher for magnetic fields exposure of 0.2 μT and 3.8 times for fields above 0.3 μT. There were no
### Table 3
Epidemiological studies of occupational and residential exposure to static and ELF magnetic fields

<table>
<thead>
<tr>
<th>Human population</th>
<th>Average exposure</th>
<th>Reported risks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Static magnetic fields</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soviet workers in permanent magnet production (645 exposed people)</td>
<td>2–5 mT at hands 0.3–0.5 mT at head</td>
<td>Subjective and minor physiological effects [16]</td>
</tr>
<tr>
<td>US workers in aluminium plants</td>
<td>No field values reported</td>
<td>Increased risk of all classes of leukaemia [17]</td>
</tr>
<tr>
<td>High-energy accelerator labs, bubble chambers, high-field magnet facilities (792 controls)</td>
<td>Up to 2 T</td>
<td>No increased risk for 19 common diseases including cancers [18]</td>
</tr>
<tr>
<td><strong>50/60 Hz magnetic fields</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>US workers 438,000 death records</td>
<td>Field exposures based on job title</td>
<td>Increased risk of acute myeloid leukaemia [19]</td>
</tr>
<tr>
<td>US workers death data from 16 States</td>
<td>Field exposures based on job title</td>
<td>Higher incidence of brain cancer, but not leukaemia [20]</td>
</tr>
<tr>
<td>US electric utility workers (36,000 people)</td>
<td>Field exposures estimated on measurements in the workplace</td>
<td>No detection of risks for any type of cancer [21]</td>
</tr>
<tr>
<td>Canadian and French electric utility workers (223,292 people)</td>
<td>Field exposures estimated on measurements in the workplace</td>
<td>Relative risk of acute myeloid leukaemia among workers with higher cumulative exposure [22]</td>
</tr>
<tr>
<td>US electric utility workers (138,000 people)</td>
<td>Field exposures estimated on measurements in the workplace</td>
<td>No association between occupational exposure and leukaemia but link to brain cancer [23]</td>
</tr>
<tr>
<td>1015 different workplaces in Sweden involving 169 occupations</td>
<td>Field exposures estimated on measurements in the workplace</td>
<td>Increased risk for chronic lymphocytic leukaemia. Increase risk for brain tumours for age under 40 and average field above 2 μT [24].</td>
</tr>
<tr>
<td>Swedish report on people living near high voltage transmission lines</td>
<td>Field exposures estimated on measurements in the residential areas.</td>
<td>Childlike leukaemia risk 2.7 times higher for exposure of 0.2 μT and 3.8 above 0.3 μT [25].</td>
</tr>
</tbody>
</table>

elevated risks for other types of cancer. The relative risk values are closer to the border line of statistical significance and the value of 0.2 μT distinguishing exposed and unexposed people may not provide, from current knowledge, a sufficient basis for setting threshold exposure limits of such low intensity.

On the contrary, a recent survey on 36,000 electric utility workers reported no strong consistent evidence of association between magnetic fields and any type of cancer [23].

As already mentioned above, different studies disagree in important ways both on the value of excess risk associated with magnetic exposure and on the type of disease.
So far there is no laboratory evidence for health effects at the field values considered in these studies. If some effects occur, they are likely to be so weak that the body is almost able to compensate, making them very hard to study.

8. STATIC AND ELF MAGNETIC FIELD EXPOSURE GUIDELINES AND PROTECTIVE MEASURES

Currently available information has not confirmed evident and reproducible adverse health effects in order to clearly indicate safe limit values and durations for magnetic field exposure. There are several initiatives to establish official standards for occupational and public exposure to static and ELF magnetic fields. Some governments, being unable to determine standards with the support of the scientific knowledge, defined limit values based on what is technologically achievable and not on medical or epidemiological studies.

Most of the guidelines reported in Table 4 have not been issued by authoritative laws but have been defined by international institutions as recommended limit values.

**Table 4**

Limit values for occupational and public exposure to static and ELF magnetic fields. 
(These reference values are not intended as a threshold to a dangerous level.)

<table>
<thead>
<tr>
<th>Static magnetic fields</th>
<th>SLAC</th>
<th>CERN</th>
<th>CENELEC TC 111</th>
<th>ACGIH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working day/whole body</td>
<td>20 mT</td>
<td>200 mT</td>
<td>200 mT</td>
<td>60 mT</td>
</tr>
<tr>
<td>Working day/limbs</td>
<td>200 mT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short/whole body</td>
<td>200 mT</td>
<td>2 T</td>
<td>2 T</td>
<td></td>
</tr>
<tr>
<td>Short/limbs</td>
<td>2 T</td>
<td>5 T</td>
<td>5 T</td>
<td>2 T</td>
</tr>
<tr>
<td>General public/whole body</td>
<td>10 mT</td>
<td>40 mT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General public/limbs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Persons with pace-maker (or large metal implants)</td>
<td>0.5 mT</td>
<td>0.5 mT</td>
<td>0.5 mT</td>
<td></td>
</tr>
<tr>
<td>Average Earth's field</td>
<td></td>
<td></td>
<td></td>
<td>50 µT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ELF magnetic fields</th>
<th>CENELEC TC 111</th>
<th>ACGIH</th>
<th>ICNIRP (50/60 Hz)</th>
<th>Regulations for transmission lines (50/60 Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day/whole body</td>
<td>80/f 1.6 mT at 50 Hz</td>
<td>60/f 1.2 mT at 50 Hz</td>
<td>0.5 mT 5 mT (short)</td>
<td>Italy 100 µT 1 mT (short)</td>
</tr>
<tr>
<td>Day/limbs</td>
<td>1250/f</td>
<td>300/f</td>
<td>25 mT</td>
<td>Florida 15 µT</td>
</tr>
</tbody>
</table>
In Europe, a prestandard was approved in late 1994 by the Technical Committee CENELEC TC111 \textit{i}Electromagnetic fields in the human environment\textit{t}. This prestandard, whose validity is limited initially to three years, is divided into two parts dedicated respectively to occupational and public exposure to low frequency (0–10 kHz) and high frequency (10 kHz – 300 Ghz) electromagnetic fields [26]. It was issued for provisional application and may be modified, before conversion to a standard, on the basis of new scientific findings and experience gained during its application.

Presently there are no federal standards in the United States but the Federal Energy Policy Act established in 1994 a five-year program, known as EMF RAPID program (Electric and Magnetic Fields Research and Public Information Dissemination), managed by DoE and the National Institute of Environmental Health Sciences. This program should explore the potential relevance of EMF exposure for possible health effects.

Among the most important international organisations that developed exposure guidelines should be mentioned: the International Commission on Non-Ionising Radiation Protection (ICNIRP) [27], the European Organisation for Nuclear Research (CERN) [28] and the American Conference of Governmental Industrial Hygienists (ACGIH) [29]. The first widely used guidelines for researchers working with high magnetic fields were formulated by the Stanford Linear Accelerator Laboratory (SLAC).

Some studies [24, 25] reported possible long-term effects on health associated with magnetic field strengths lower than those specified in Table 4. However, as already mentioned above, they do not prove indisputably that harmful risks exist and are not supported by evident scientific confirmations. Consequently there are no sufficient bases for setting threshold exposure limits of such low intensities.

The American Physical Society (APS) issued a bulletin in 1995 on this subject reporting that both the scientific literature and reviews by other panels show no consistent and significant links between cancer and ordinary ELF magnetic fields.

On the basis of extensive laboratory studies, various other significant biological processes do not seem to be influenced significantly by static magnetic fields up to 1–2 T. These processes include: cell growth and morphology, DNA structure, reproduction, physiological regulation and circadian rhythms [30].
8.1 Pacemakers and implanted metal objects

Low intensity magnetic fields may be considered to be safe from the biological point of view but may cause problems to people with pacemakers. The majority of pacemakers implanted today are synchronous and are activated only when the heart rate, continuously monitored, falls below a preset level. Incorporated in the pacemaker is usually a reed relay that activates fixed-frequency pulses and is helpful for the physician to test the correct working of the apparatus. It was found that certain types of cardiac pacemakers in the presence of static fields, above 0.5–1.5 mT, may reverse into this fixed-rate mode (called asynchronous mode). This circumstance is potentially hazardous and may lead to fibrillation owing to the competitive pacing between the natural heart rate and the rate stimulated by the pacemakers [31]. Also time-varying magnetic fields, in excess of 0.1–0.4 mT, may alter the pacemaker functioning, by inducing voltages that may be recognised as cardiac signals.

Since magnetic fields decrease as the distance from the source increases, the best protective measure, when the magnetic field is higher than these limit values, is a separation distance. In any case warning signs for people with cardiac pacemakers or metal implants and prosthesis (that may experience significant forces and torques) must be displayed in accessible places where magnetic fields are above 0.5 mT.

9. CONCLUSIONS

There is controversy on the possible health effects of static and ELF magnetic fields on humans [32]. It is difficult to prove indisputably whether harmful risks exist or not. Although the potentially exposed population is large, the risks, if any, appear to be small and limited to specific situations. But, in any case, it is better to keep the public exposure well below the limit values. A valuable support may come from the new technologies that, if duly stimulated, may solve problems in ways that are compatible with the environment. There are already examples, in other sectors, pointing out how high-technology companies are able to find alternatives in a short time to comply with new regulations.

Most of the public attention is now focused on ELF fields, associated to high voltage transmission lines but studies on the effects of static fields are also expanding after the large diffusion of magnetic resonance imaging (MRI) systems employing static magnetic fields up to 2 T. Static fields are expected to attract much more attention when high magnetic field technologies, based mostly on superconductors, like superconducting magnetic energy storage plants (SMES), thermonuclear fusion reactors and magnetically levitated trains (MAGLEV), begin to spread.

REFERENCES


[26] European Prestandard Human Exposure To Electromagnetics Fields – Low Frequency (0–10 kHz), ENV 50166-1, and High Frequency (10 kHz – 300 Ghz), ENV 50166-2, CENELEC 1995.


[29] Threshold Limit Values for Chemical Substances and Physical Agents - Biological Exposure Indices, American Conference of Governmental Industrial Hygienists (ACGIH), pp. 113-120, 1996.


SUPERCONDUCTING MAGNET FABRICATION AND QUALITY CONTROL

R. Penco
Ansaldo Energia, Genova, Italy

ABSTRACT
A general review of the problems and production risk related to the industrial fabrication of prototypes and/or short series of superconducting magnets is presented. In particular, past and present experience on the manufacture and testing of s.c. devices is described. Cost implications of the design choices and/or requested guaranteed technical specifications are discussed. The types and number of possible factory quality control tests during and after the fabrication are presented.

1. INTRODUCTION
In the last forty years, solid-state quantum mechanical effects have been applied to industrial products. The macroscopic effects of this microscopic phenomena are well known by most people and are part of our everyday life: everybody is familiar with modern electronics! Anyway it seems that Heisenberg's uncertainty principle affects in some way the production of these industrial components based on the application of quantum mechanics. The probability that a quantum component fully meets the quality control tests is in general lower than with classical components.

The number of scrapped or declassified pieces in electronics is usually one order of magnitude higher than in mechanics. On the other hand the cost of making electronic components is much lower than for mechanical or electrical ones, and therefore the total production costs of the latter components are more related to the R&D and quality control costs than to the pure manufacturing process.

If the difficulty related to the use of superconductivity is added to that of producing large components, there is a fearful mixing of the most relevant problems present in the electronics industry (e.g. R&D and quality control cost, rejected pieces) and the ones related to the high cost of classically mechanical component production.

2. PRODUCTION RISKS
In the quality control system, a special section is normally devoted to “Special Processes” defined as the manufacturing technique that can be completely verified only by destructive tests: welding is a typical example. Usually, in this case, it is necessary to:
- Qualify the process
- Qualify the operator
- Qualify the non-destructive controls.

The production of a large superconducting device involves several special processes some of which cannot be fully checked before the final test of the complete device. In the following sections I will try to list the main quality control problems related to superconductivity applications and which trouble the life of the designers. These problems have to be added to
the standard ones of insulation, mechanical tolerances, welding, vacuum and cryogenics, etc. that will not be treated here.

2.1 The cable

Before the winding operation and the final test of a s.c. magnet only short sample tests can be made on the cable, and these only from the ends of the produced lengths. This situation is similar to what happens in welding which is sometimes extended, then cut to perform the mechanical test. The results given by this kind of test are significant only if you make the assumption that the production parameters are constant and reproducible in all phases of the production. The more the process is established, the more this method of testing can be safely adopted. Today, the production of s.c. single strands well satisfies these criteria: tests made by cutting long lengths of strand into several pieces then measuring the critical current of each piece never give significant differences. Furthermore, line control using eddy currents tests is normally added to detect local defects.

A different approach has to be taken for the production of cabled conductor and even more so for co-extruded conductor. The cabling in general produces slight degradation of the critical current of the single strand (this has to be added to a change in the transport characteristic due to the self-field effects). This degradation, if present, is often due to micro-rupture or damage of the internal filament of the strand or to the permanent deformation of the strand during the cabling process. That this degradation is constant along the whole length of a cable (usually some km.) is more questionable.

In the case of aluminium co-extrusion which requires a high temperature we have to add to the possible degradation related to the cabling, the possibility of a temperature degradation: a NbTi wire can withstand a temperature of up to 300°C without changing its critical current characteristic. If this limit is exceeded the degradation is a function of the time of exposure to the high temperature. At 350°C the effect occurs after several minutes, while at 530°C it is reduced to a few seconds.

In general the cabling process is designed to avoid any degradation or to minimise it. Normally the temperature of the environment is kept higher than that of the cable and the maximum temperature that the cable can reach is a function of the external temperature and the cable speed. It is clear that in the case of a stop in the process the cable can reach a temperature well above the designed one. Clearly this event will not affect the results of the short samples taken from the ends of the cable and so the eventual defects will not be detected before the final tests on the magnet. The only possibility to guarantee the results is to constantly monitor and record the process parameters during the fabrication of the cable. The risk increases the higher is the maximum temperature involved in the process: Aluminium co-extruded cable is the more critical but some problems can also occur during the fabrication of tin-soldered or varnish-insulated strand.

Clearly other important controls are more simple to make: for example from the point of view of the quality of the field the dimensions of the cable can be critical; the filament diameter can also have an influence on the persistent-current effect on the field harmonics. Table 1 gives an outline of the Quality Control Plan required to control production of the co-extruded Al cable for the FINUDA detector magnet.

2.2 The cable joints

Single lengths of s.c. cable must be joined in a way which guarantees very low heat dissipation during operation. Joule effect in the joint has to be limited to some milliWatts and for current in the order of kA demands an electrical resistance of the joint in the order of \(10^{-9}\) ohm. For persistent MRI or NMR magnets the specification reaches \(10^{-14}\) ohm or better. Clearly it is not possible to verify this value during magnet construction as it is necessary to
have the cable in the s.c. state to be able to measure this level of joint resistance. We have to treat the joint in the same way as the welding of a pressure vessel: we must qualify the process, the operators and the non-destructive controls to be performed. Qualification of the process can be made only by adding to the other possible controls a resistance measurement of samples at the nominal current in the nominal magnetic field. Sometimes a measurement of AC losses is also requested.

**Table 1**

Example of some of the controls required by the subcontractor for a co-extruded Al cable for detector magnets

<table>
<thead>
<tr>
<th>Examination test</th>
<th>Value</th>
<th>Test specimen or sample</th>
<th>Examination or test frequency</th>
<th>Test method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al chemical analysis</td>
<td>T.B.D.</td>
<td>Billet before extrusion</td>
<td>One sample for each casting</td>
<td>Supplier spec.</td>
</tr>
<tr>
<td>Al RRR</td>
<td>T.B.D.</td>
<td>Billet before extrusion</td>
<td>One sample for each casting</td>
<td>Supplier spec.</td>
</tr>
<tr>
<td>Wire integrity</td>
<td>No more than 5%</td>
<td>Extracted strand</td>
<td>One sample from each end of each continuous conductor length</td>
<td>Visual and eddy current inspection</td>
</tr>
<tr>
<td>Critical current min. at 1.5 T, 4.4 K (or the equivalent from the extracted strands)</td>
<td>&gt; 8000 A</td>
<td>Cable</td>
<td>One sample for each end of each continuous conductor length</td>
<td>ASTM B 714-82</td>
</tr>
<tr>
<td></td>
<td>(or the equivalent from the extracted strands)</td>
<td>Extracted strand</td>
<td>One sample for each end of each continuous conductor length</td>
<td></td>
</tr>
<tr>
<td>Transfer resistivity between matrix and Rutherford</td>
<td>@ 1.5 T &lt; 2 x 10^{-10} •m</td>
<td>Final conductor</td>
<td>One sample from each end of each continuous conductor length</td>
<td>App. 6</td>
</tr>
<tr>
<td>Bond shear test</td>
<td>&gt; 20 MPa</td>
<td>Final conductor</td>
<td>One sample from each end of each continuous conductor length</td>
<td>Supplier spec.</td>
</tr>
<tr>
<td>Surface conditions</td>
<td>No surface defects, burs, sharp edges, slivers, foils, laminations, dirt, inclusions or oxide</td>
<td>Final conductor</td>
<td>100%</td>
<td>Visual</td>
</tr>
<tr>
<td>“n” index at 1.5 T, 4.5 K</td>
<td>&gt; 40</td>
<td>Final conductor</td>
<td>Each end of every continuous length</td>
<td>Amplified voltage signal recording</td>
</tr>
<tr>
<td>Dimensional check</td>
<td>Cable A: (4.13 x 14.40) ± 0.01 mm</td>
<td>Final conductor</td>
<td>Each end of every continuous length</td>
<td>Dimensional</td>
</tr>
<tr>
<td></td>
<td>Cable B: (6.03 x 14.40) ± 0.01 mm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al RRR=R295/R10 Al Cu</td>
<td>&gt; 1100</td>
<td>Final conductor stabiliser and strand</td>
<td>One sample from each end of each continuous conductor length</td>
<td>App. 3</td>
</tr>
<tr>
<td></td>
<td>&gt; 100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cu/NbTi ratio</td>
<td>Cable A: &gt; 1.1:1</td>
<td>Extracted strand</td>
<td>One sample from each end of each continuous conductor length</td>
<td>Dimensional</td>
</tr>
<tr>
<td></td>
<td>Cable B: &gt; 1.1:1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edge curvature radius</td>
<td>&gt; 0.5 mm (no sharp edge)</td>
<td>Final conductor</td>
<td>Spot measurement every 1000 m of each continuous conductor length</td>
<td>Dimensional</td>
</tr>
</tbody>
</table>

This kind of test is clearly costly, both in money and time, in proportion to the field, the current and the physical dimensions of the sample. Furthermore the joint geometry and technique strictly depend on the type of cable and magnet. In bath-cooled multipole magnets for accelerators, usually Rutherford cable is simply tin soldered and the only problem could be to find the proper flux and brazing material. The Al co-extruded cables for the large detector magnets are TIG welded (with attention to possible temperature degradation) or tin brazed after (or without) copper plating of the aluminium.

Joining high-current CIC (cable in conduit) is more difficult because the problem of helium circulation and tightness is added to the electrical low resistance. A further problem occurs with Nb3Sn cable that has to be treated at 650°C after winding. In this case, due to the fragility of the reacted cable, the junction has to be made in different steps: position and press after winding, close and tin solder (if necessary) after the heat treatment but before vacuum impregnation.

In Figs. 1 and 2 are shown two examples of samples made for joint qualification for fusion magnets. Both use Nb3Sn CIC cable: some (circular shape) were tested up to 8 T with a 10 kA current (2 nO hm) [2], the other (straight) was tested to 40 kA in a 12 T field (0.2 nOhm) [3].

![Fig. 1 NET 40 kA short sample cable joints](image1)

![Fig. 2 12 T CIC cable joint sample](image2)

**2.3 The training or degradation of magnets**

Briefly, these effects are due to a mechanical movement of the coil (or in the coil) which releases energy and locally heats the superconductor which then quenches to the normal state. We speak of training if this event produces a crack in the resin insulation or a permanent deformation. In this case each subsequent time that his value of stress (or, if you like, this value of current and field) is exceeded there is no further energy release and the magnet can be so trained to reach the nominal working condition. But if this release of energy is elastic a quench will happen every time at the same value of stress, so this level of field never can be overcome and the degraded magnet will never reach its nominal condition.

We can act on two parameters to avoid this effect: increase the cable stability or decrease the possible energy realise.

**2.3.1 Stability**
Cable stability is a well-defined parameter: it is well known how to calculate and how to measure it. Different effects (flux jump, self field, dynamic, cryogenic stability) are used to define the wire or cable stability.

One way to increase the stability is by increasing the minimum propagating energy (MPE) (the energy release that the cable can withstand before starting quench propagation). This is related to the margin between the working temperature and the quench propagation temperature of the cable in the operating condition as well as the thermal capacity of the cable. The quench propagation temperature is related to the current sharing temperature (when the current starts to pass in the matrix because the superconductor is saturated) and to the electrical resistance of the matrix (that determines the joule dissipation). So, to increase the MPE, it is possible to use better s.c. material, more s.c., or more or better matrix material.

In general increasing the stability implies decreasing the overall current density with a consequent increase in the dimensions of a magnet and its cost, and a decrease in its general performance. If the overall current density is fixed the only degree of freedom is the choice of the s.c. and matrix materials and the ratio between them. Optimising this choice is one of the more delicate stages in the design of a s.c. magnet.

It is also possible to use the large heat capacity of helium to increase the stability of the cable. This was achieved in the past by allowing the coolant to wet the bare conductor. The technique is still used today in dipoles where the cable is wrapped with Kapton and in CIC magnets [3]. Unfortunately the “void fraction” used for the helium decreases the overall current density. Ratios between conductor and void area are usually in the range of 1:1 to 1:30.

2.3.2 Mechanical energy release

This problem is more difficult to treat as it is related not only to the design but to the whole fabrication process. From this point of view, we can, in practice, consider the total winding as a Special Process.

There are two possibilities to decrease the energy release:

– Avoid or reduce the coil (or turns) movement
– Let the coil (or turns) move freely with little or no energy release peak.

These two philosophies are quite different and it is dangerous to mix them in the same design.

In a s.c. magnet, the forces act mainly on the cable (while in the conventional magnet they mainly act on the iron): for this reason it is not possible to avoid stress and strain on the superconductor. To allow the turns to move freely is usually impossible since the forces acting on the different layers are different, so producing relative movement and friction between the layers.

It is sometimes possible to allow the whole winding to move freely if it is able to self sustain the electromagnetic force: this can happen in solenoids but not in other magnets which will always need an external mechanical structure to withstand the forces.

To avoid the relative movements, one can use:

- Bonding
- Pre stress

Both are normally used for fixing the turns between themselves: winding is carried out under tension, turns are compressed one against the others and finally bonded using pre-preg, wet-winding or vacuum-impregnation techniques. Usually local resin cracks can occur during first powering of the magnet but most magnets are stable enough to withstand them. Bonding the coils against the mechanical structure could be dangerous due to their different thermal and mechanical behaviour which could induce a large shear stress on the bonding itself. When a
magnet is designed to avoid relative movements between the coils and structure, the latter must be fixed extremely well: if the bonding or the pre-stress is not enough to withstand the shear, the resulting peak release of energy will usually be enough to induce a magnet quench and training or degradation will occur.

Now the problem is:

How can we forecast the value of the energy release peak for glass-epoxy cracks or friction movements?

How can we be sure about the bonding and pre stress in the magnet?

The number of reports on mechanical energy release in magnets is perhaps two orders of magnitude less than the number of papers about the stability of s.c. cables, whereas equal knowledge of both phenomena is necessary in designing a magnet!

The value normally used for resin-crack energy-density release at liquid helium temperature is about $10^5$ J/m$^3$ (100 mJ/ mm$^3$). As simple calculation on the work per volume unit involved in the movement of a wire in an external magnetic field gives: $E/V = B* J * s$; For $B = 5$ T and $J = 10^5$ A/mm$^2$, a movement of 1 micron give an energy of $5*10^3$ J/mm$^3$, fifty times that of the resin crack.

The choice of the stability margin to be used against possible energy release is still a matter of discussion by the designers of new magnets and is also related to the confidence they have on the capacity of the workshop to build the device in the proper way.

2.4 Heat treatment of the winding

Some coils have to be heat treated after being wound. This treatment is usually for the winding and reaction technique used for Nb$_3$Sn magnets but sometimes for other reasons such as the ageing of the aluminium in co-extruded cable. During this process the temperature has to be controlled for homogeneity and for a given time. To cure large A15 superconductor magnets such as the ITER model coil (4 x 3 m) a large oven with controlled temperature and atmosphere is required. Temperature homogeneity of ± 3°C is necessary at 650°C for 220 hours. This implies a guaranteed power supply able to maintain this temperature in the case of an energy supply fault.

2.5 Winding of reacted Nb3Sn cable

Since some tape insulation is not able to withstand the high temperature necessary for curing of A15 superconductors, it may be necessary to apply it after reacting the cable. However, reacted wire is able to withstand a strain of only 0.2% without degradation of the critical current. This makes the winding or transfer operation incredibly critical because there is no way to control the cable condition after the winding but before the final test of the magnet.

3. PROTOTYPES

It is particularly difficult to make decisions on the prototypes to be manufactured before starting the real production (to make them or not, how many, full scale or reduced in dimension, ...). The reasons for this are related to the high costs for prototype construction as well as to a relatively high risk of failure. On the other hand, due to the complexity of the production process it is difficult to be sure about it without having first fully tested it. Interactions between different parts of the magnet may not have been considered during the design and become evident only after testing the complete full-scale component. Anyway, for the production of large amounts of similar magnets everybody agrees on the necessity to make some prototypes before starting series production (see Fig. 3).
Unfortunately people do not always look on prototype construction with the same point of view: some think that it is useful for testing the reliability of the design and the production method, others think that it is an occasion to improve the product performance or to reduce production costs. These two approaches are in conflict since in the first case you need to make several prototypes all the same as the production magnets, in the second one often changes the design in order to optimise it.

In any case changing the design of superconductor magnets is never trivial: sometimes it is even dangerous after having made several prototypes to change during the production a minor item that was considered insignificant but turns out not to be.

The full-scale LHC prototypes made so far and tested at CERN used basically the same design but some technical details were different and gave different ways to reach the same goals. Nevertheless their tests gave different results.

It is important that a company large enough to construct a series of production magnets be asked to participate in the R&D programme and in any prototype construction right from the beginning in order to decrease the risk, cost and time for the final work [4].

![Fig. 3 Winding of LHC 15-m full-scale prototype](image)

4. **QUALITY CONTROL DURING THE PRODUCTION**

The quality system has normally to control: material, process and performance of the product.

*Control of material* usually starts at the subcontractor’s factory to monitor their production process and continues during magnet production by recording the components used for each magnet. This was done for the more critical materials where it was requested not only to control if the parameters were within the specified values but also to measure the actual value of that parameter (the s.c. cable is the clearest example where it is necessary to record which particular length was used for the winding of each particular magnet).

*During fabrication* it is important to control all the parameters related to the special processes: in particular, the cable tension, cleanliness integrity and compression of insulation, vacuum level during impregnation, cable temperature during the joint welding and so on.
Dimensional control is also very important as it is related to the coil pre-stress and to the final field quality. The mechanical tolerances requested are usually very tight and sometimes not even possible to reach in a single step construction. In this case it is necessary to use the technique of “measure and correct” or “measure and pair”. The winding of the Tore Supra magnet where the tolerance of 0.3 mm for the 2.75 m external diameter of the winding was achieved only by a constant feedback between the cable tension and the actual winding diameter. The HERA dipole coils were measured and then paired to decrease the relative difference.

These examples give an impression of the impact on the production cost of the quality control work. Sometimes the designers start to think about the problems related to production quality control only after the design and final drawings are completely finished. A typical example is the multipole magnet collars and yoke laminations where all dimensions, with very small tolerances, were given with respect to the centre of the beam that is outside the piece itself. This kind of drawing makes it more difficult to control the dimensions.

Electrical testing during fabrication is usually restricted to an insulation check of turns-to-turns and turns-to-ground. The number and types of these tests should be decided by the constructor who has to guarantee the final result. However, the client sometimes asks for electrical tests that could be dangerous for the coil if they are made before the coil is completed.
The final test of a s.c. device can be made only at low temperature (until room temperature s.c. materials are available that is). This implies costly tooling and manpower to cool down and maintain cold the component during testing. For this reason and also because not every factory had its own refrigeration plant only the room temperature tests (insulation, vacuum, low current field quality) were made in the factory while the final tests were made at the client’s premises. Today the tendency is to always ask for delivery of a fully tested magnet, an example being the FINUDA magnet which was tested with its iron yoke (250 tons), including complete field mapping, in the Ansaldo test room (see Fig. 4). Other examples of factory testing are illustrated in Figs. 5 and 6.

Fig. 4 CERN field mapping device in the FINUDA magnet under test at ANSALDO [5]

Fig. 5 AGOR cyclotron under test at

Fig. 6 ZEUS compensator [6] under factory
5. ANALYSIS OF REAL SERIES PRODUCTION

Unfortunately the series production of s.c. magnets or other devices is relatively rare, so that a statistical analysis is related only to a few examples. In the following some of the results related to production at ANSALDO are analysed.

5.1 HERA dipoles

Ansaldo produced 232 s.c. dipoles for the DESY accelerator in the years 1984–1988. Final discussions about the guaranteed performance of these dipoles was made with DESY only after completion and testing of the first ten prototypes. Since, however, the results were excellent, ANSALDO decided to accept to fully guarantee the requested values for field intensity and quality.

Tables 2–4 and Fig. 7 summarise the results [7] on the warm field quality measurements made in the factory before putting the coils into their iron yokes since the DESY specification obviously includes the variation induced by the iron. If a magnet was out of tolerance, the collars were removed and the shims changed in order to reach the right value. Among the 29 dipoles out of tolerance at the first attempt 27 had problems with A2 component, 1 with A2 and A4, 1 with B3.

Table 2

Attempts to reach the right harmonic content

<table>
<thead>
<tr>
<th>Harmonic content in tolerance</th>
<th>No. of dipoles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st attempt</td>
<td>201 (87.4%)</td>
</tr>
<tr>
<td>2nd attempt</td>
<td>23 (10%)</td>
</tr>
<tr>
<td>3rd attempt</td>
<td>4 (1.7%)</td>
</tr>
<tr>
<td>4th attempt</td>
<td>1 (0.4%)</td>
</tr>
<tr>
<td>5th attempt</td>
<td>1 (0.4%)</td>
</tr>
<tr>
<td></td>
<td>230</td>
</tr>
</tbody>
</table>

Table 3

Harmonic components of collared coils (over 230 dipoles)

<table>
<thead>
<tr>
<th>Harmonic component</th>
<th>DESY request</th>
<th>Measured values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nominal value</td>
<td>Mean value</td>
</tr>
<tr>
<td>A2 (quadrupole)</td>
<td>0 ± 4</td>
<td>0.47</td>
</tr>
<tr>
<td>A3</td>
<td>0 ± 3</td>
<td>-0.49</td>
</tr>
<tr>
<td>A4</td>
<td>0 ± 6</td>
<td>-0.07</td>
</tr>
<tr>
<td>A5</td>
<td>0 ± 2</td>
<td>-0.08</td>
</tr>
<tr>
<td>A6</td>
<td>0 ± 2</td>
<td>-0.05</td>
</tr>
<tr>
<td>A7</td>
<td>0 ± 2</td>
<td>-0.06</td>
</tr>
<tr>
<td>B2</td>
<td>0 ± 4</td>
<td>0.13</td>
</tr>
<tr>
<td>B3 (sextupole)</td>
<td>-14.3 ± 10</td>
<td>-15.4</td>
</tr>
<tr>
<td>B4</td>
<td>0 ± 3</td>
<td>-0.06</td>
</tr>
<tr>
<td>B5</td>
<td>1.3 ± 4</td>
<td>1.98</td>
</tr>
<tr>
<td>B6</td>
<td>0 ± 2</td>
<td>-0.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td>B7</td>
<td>0.4 ± 2</td>
<td>0.0</td>
</tr>
</tbody>
</table>
It has to be pointed out that some of the problems encountered were related to the measurement device itself. After several months the friction in the supports of the centring system degraded and falsely indicated a component in the harmonics which we then corrected by shimming the coils!

**Table 4**
Main harmonic component mean values (and their standard deviations)
measured in different positions along the beam axes

<table>
<thead>
<tr>
<th>Position</th>
<th>A2</th>
<th>A4</th>
<th>B3</th>
<th>B5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (electrical exit end)</td>
<td>-0.05 ± 2.6</td>
<td>0.1 ± 1.3</td>
<td>-13.6 ± 3</td>
<td>0.7 ± 1</td>
</tr>
<tr>
<td>2</td>
<td>0.25 ± 2.4</td>
<td>-0.15 ± 1.3</td>
<td>-15.0 ± 3.2</td>
<td>3.0 ± 1.1</td>
</tr>
<tr>
<td>3</td>
<td>0.75 ± 2.3</td>
<td>-0.22 ± 1.3</td>
<td>-15.4 ± 3.0</td>
<td>3.0 ± 1.0</td>
</tr>
<tr>
<td>4 (opposite end)</td>
<td>1.0 ± 3.1</td>
<td>0.54 ± 1.3</td>
<td>-17.4 ± 3.1</td>
<td>1.0 ± 1.2</td>
</tr>
</tbody>
</table>

Fig. 7 Distribution of some harmonic components over the whole production of HERA dipoles
Fortunately we had an old magnet with which to repeat the test and so found the mistake. From this experience it is clear how important it is to have not only the measurement device but also the possibility to periodically re-calibrate it.
Another analysis was made of the field intensity. The data are related to the 29 collared coils tested in our factory (Fig. 8) over the whole production run of 232 magnets [8]. As the test involved magnets made with different cable characteristics but little difference in the helium temperature, the quench data were given for the temperature margin of the quench defined as the difference between the helium bath temperature and the critical temperature of that coil at the quench field and current calculated from the short-sample measurement of the cable. As stated before, this difference is related to the mechanical disturbance and energy release that induces a transition before the short-sample critical current is reached.

Taking into account that we used a numerically controlled winding machine to ensure the most reproducible winding process, that the curing and collaring process were constantly monitored and that the 27 magnets tested were exactly similar in design and used the same tooling, it is surprising how variable were the results: only five of them reached the guaranteed current without training, another seven only after five or six attempts, the remainder falling somewhere between these extremes.

Fig. 8  Collared HERA coil during insertion in the vertical cryostat for quench testing at ANSALDO
6. CONCLUSION

As stated in the introduction, the production cost of superconducting devices (also for series production) is due not only to the material and manpower for simply manufacturing them but is strongly related to the quality control cost and the percentage of scrapped pieces. This highlights the importance of having the possibility to use skilled personnel and to produce prototypes to test the processes. It is important to optimise the cost from the point of view of reliability of the process more than, for example, by simply decreasing the material cost.

If a firm is asked for a fixed price offer with two possible options where in the second option the cost of material is lower but the process is a little less reliable, the firm has to forecast the price increase for quality control work and the possible number of rejected pieces. This is normally one of the more difficult items to be decided and, due to prudence, is usually overestimated. Since the cost of rejected pieces includes not only the lost material but also the cost of wasted work, the total price for the second option may well be higher than for the first one.

REFERENCES


